Evolutionary trajectory of the enzyme activation-induced cytidine deaminase (AID) within the Gadiformes lineage

By Atefeh Ghorbani

A thesis submitted to the School of graduate studies in partial fulfillment of the requirements for the degree of Ph.D., **Division of Biomedical Sciences, Faculty of Medicine, Memorial University of Newfoundland**

May 2021

St. John's, Newfoundland and Labrador

Abstract

Activation-induced cytidine deaminase (AID) is a DNA-mutating enzyme that initiates secondary antibody diversification process upon immune stimulation. One outcome of this diversification is the generation of antibodies with higher affinity for the cognate antigen. In human and mouse models, functional deficiency of AID leads to hyper IgM syndrome type II, exhibiting lack of secondary antibody diversification. Despite the central role of AID in instigating this diversification process, its off-targeting activity has been attributed to the initiation and progression of various type of cancers. The emergence of AID and, therefore, secondary antibody diversification process have been dated back to the common ancestor of jawed vertebrates. However, several studies investigating the Atlantic cod (*Gadus morhua*) immune responses revealed lack of high affinity antibodies and robust humoral response in this species. Moreover, genomic sequence of several Gadiformes species, including Atlantic cod, uncovered the loss of histocompatibility complex class II (*mhc II*), cluster of differentiation 4 (*cd4*), and invariant chain (*Ii*) genes in their common ancestor. These genes are involved in B cell activation in the mammalian model of immune system. Since AID is responsible for generation of high affinity antibodies in other vertebrates, we sought to examine the genetics, expression, and function of Atlantic cod AID. We also investigated the evolutionary trajectory of AID within Gadiformes species to shed light on the extent of immune system remodeling in this lineage. In chapter two, we showed that the AID gene synteny and transcript expression were conserved in Atlantic cod in comparison with other studied vertebrates. Interestingly,

we identified two distinct AID transcripts, one of which encoded a full-length AID, whilst the other one lacked the first exon. In chapter three, we synthesized, expressed, and purified Atlantic cod AID (Gm-AID) and examined its biochemical properties. Our results showed that despite having a similar DNA binding ability, Gm-AID exhibited extremely low catalytic efficiency compared with other studied vertebrates. In chapter four, we synthesized, expressed, and purified 36 AID homologs within and outside of the Gadiformes lineage. Previous studies have shown a drastic re-modeling of the Gadigormes' immune system where the loss of genes involved in antibody responses has coincided with an expansion of innate and cell-mediated immune genes. Our biochemical analyses revealed a vast diversity in the enzymatic properties of AID homologs. Remarkably, two Gadifomes AID homologs examined here did not exhibit any cytidine deaminase activity. By predicting and resurrecting the ancestral AIDs within and outside of Gadiformes lineage, we showed that the functional impairment of AID most likely has happened in the ancestor of Gadidae group. Since Gadidae species have successfully populated their natural habitats, the functional impairment of their AID enzyme did not hamper their fitness. This is most likely duet to the compensatory mechanisms such as the expansion of innate and cell-mediated immune systems. Our findings of the first example of a vertebrate species with a dysfunctional AID and secondary antibody diversification challenge the long-standing immunological concept that the loss of AID activity leads to immunodeficiency.

Acknowledgements

During my Ph.D. study, I was fortunate to be surrounded by many kind and supportive people. Here, I would like to thank my family, especially my parents, whose encouragement made it possible to overcome the obstacles. I would also like to thank my supervisor, Dr. Larijani, whose guidance and advise shaped me to be a better scientist. I would like to thank my committee members, Drs. Grant and Paterno, and collaborators Drs. Jentoft and Rise for their continued support and guidance. Also, I would like to acknowledge my friends, especially S. J. Khataeipour, Y. Menesses, K. D. Joris, C. D. Collins, A. Bakhshi, and B. N. Bolt and my colleagues at Memorial University, Ocean Sciences Center, and University of Oslo without whom this journey would not have been joyful. Finally, I would like to thank Pashmak and Fesgheli, my cats, for distracting me while I was writing my thesis. Thank you all for being an essential part of my academic and personal life in Canada.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3′-UTR: Untranslated region at the 3′ end of RNA transcript

5-mC: 5-methylcytidine

5′-UTR: Untranslated region at the 3′ end of RNA transcript

A3A: APOBEC3A

A3G: APOBEC3G

Ab-Ag: Antibody-antigen complex

ADARs: Adenosine deaminases acting on RNA

ADCC: Antibody-dependent cell-mediated cytotoxicity

Ag-AID: *Arctogadus glacialis* AID

aicda: Activation induced cytidine deaminase gene

AID: Activation induced cytidine deaminase

AIM2: Absent in melanoma 2

ALR: Absent in melanoma 2-like receptor

AM: Antibody affinity maturation

APE: Apurinic/apyrimidinic endonuclease

APOBEC: Apolipoprotein B-mRNA editing enzyme catalytic polypeptide-like complex family of cytidine deaminases

ASAL: Formalin-killed typical *A. salmonicida*

ASR: Ancestral Sequence Reconstruction

atps: ATP synthase H+ transporting, mitochondrial Fo complex, subunit F2

BAFF: B cell activating factor

B-ALL: B cell acute lymphoblastic leukemia

Bb-AID: *Brosme brosme* AID

BCR: B cell receptor

BER: Base excision repair

Bm-AID: *Bathygadus melanobranchus* AID

bNABs: Broadly neutralizing antibodies

bp: Base pair

Bs-AID: *Boreogadus saida* AID

B-T zone: B and T cell zone boundary

C: Immunoglobulin constant domain

Ccap: C terminus of an α-helix

CD: Cluster of differentiation

CDR: Complementarity determining region of antibodies

cGAS: Cytosolic DNA sensor

C_H: Heavy chain constant gene

CLL: Chronic lymphoid leukemia

CLR: C-type lectin receptor

CML: Chronic myeloid leukemia

CR: Complement receptor

Cr-AID: *Cyttopsis roseus* AID

CSR: Class switch recombination

CTL: Cytotoxic T cell

D: Immunoglobulin diversity segment

DAMP: Damage-associated molecular patterns

DC: Dendritic cell

dC: Deoxycytidine

Del: deletion

DLBCL: Diffuse large B cell lymphomas

DNA: Deoxyribonucleic acid

DNP-KLH: 2,4-dinitrophenyl-keyhole limpet hemocyanin

DNTT: DNA nucleotidylexotransferase

DPF: Days post fertilization

Dr-AID: *Dani rerio* AID

ds: Double-stranded

DSBs: Double-stranded breaks in DNA

dT: Deoxythymidine

dU: Deoxyuridine

DZ: Dark zone

EBI2: Epstein-Barr virus-induced receptor 2

eEF1α: Translation elongation factor 1α

EMSA: Electrophoretic mobility shift assay

FasL: Fas ligand

FDCs: Follicular DCs

FITC-KLH: Fluorescein isothiocyanate (FITC) conjugated to keyhole-limpet hemocyanin

FL: Fetal liver

FRs: Antibody framework regions

G4: G-quadruplex

Ga-AID: *Gadiculus argenteus* AID

GC: Germinal center

Gd-ANC: Gadidae ancestral AID

Gds-ANC: Gadidae sister group ancestral AID

Gf-ANC: Gadiformes ancestral AID

Gg-AID: *Gallus gallus domesticus* AID

GIALT: Gill-associated lymphoid tissue

Gm-AID: *Gadus morhua* AID

GSP: Gene-specific primers

GST: Glutathione S-transferase

GTR: General time reversible model

GTRCAT: General time reversible model with the CAT model of rate heterogeneity

HIGM: Hyper-IgM syndrome

HIV: Human immunodeficiency virus

HPI: Hours post injection

HR: Homologous recombination

Hs-AID: *Homo sapiens* AID

HSCs: Hematopoietic stem cells

IFN: Interferon

Ig: Immunoglobulin

IGC: Immunoglobulin gene conversion

IgH: Immunoglobulin heavy chain

IgL: Immunoglobulin light chain

IgNAR: Immunoglobulin new antigen receptor

IgSF: Immunoglobulin superfamily

Ii: Invariant chain

IL: Interleukin

ILC: Innate lymphoid cell

IMC: Innate myeloid cell

Ip-AID: *Ictalurus punctatus* AID

iPS: Pluripotent stem cells

IPTG: Isopropyl β-d-1-thiogalactopyranoside

IRF: Interferon regulatory factor

IS: Isotype switching

ISP: Isoform-specific primers

J: Immunoglobulin joining segment

K_d: Dissociation constant

l: Loop

Lla-AID: *Laemonema laureysi* AID

Llo-AID: *Lota lota* AID

LPS: Lipopolysaccharides

LTi: Lymphoid tissue inducer cell

LZ: Light zone

Ma-AID: *Melanogrammus aeglefinus* AID

MALT: Mucosa-associated lymphoid tissues

MAPK: Mitogen-activated protein kinase

Mb-AID: *Macrourus berglax* AID

MHC: Histocompatibility complex

mIgM: Membrane-bound IgM

ML: Maximum likelihood

MM: Michaelis-Menten

Mma-AID: *Muraenolepis marmoratus* AID

MMC: Melano-macrophage cluster

Mm-AID: *Mus musculus* AID

Mmerla-AID: *Merlangius merlangus* AID

Mmerlu-AID: *Merluccius merluccius* AID

Mmol-AID: *Molva molva* AID

Mmor-AID: *Mora mora* AID

MMR: Mismatch repair

Mo-AID: *Malacocephalus occidentalis* AID

MQ: Macrophage

mRNA: messenger RNA

MSA: Multiple sequence alignment

MyD88: Myeloid differentiation factor 88

MZ: Marginal zone

Mz-AID: *Melanonus zugmayeri* AID

nABs: Natural antibodies

NBH: B cell helper neutrophils

Ncap: N terminus of an α-helix

NER: Nucleotide excision repair

NES: Nuclear export signal

NET: Neutrophil extracellular trap

NF-κB: Nuclear factor kappa-light-chain-enhancer of activated B cells

NHEJ: Non-homologous end-joining

NK: Natural killer cell

NKT: Natural killer T cell

NLR: Nucleotide oligomerization domain-like receptor

NLS: Nuclear localization sequence

NLS: Nuclear localization signal

NMR: Nuclear magnetic resonance

NOD: Nucleotide oligomerization domain

Ol-AID: *Oryzias latipes* AID

ORF: open reading frame

PAMP: Pathogen-associated molecular pattern

pAPC: Professional antigen presenting cell

Pb-AID: *Phycis blennoides* AID

PBS: Phosphate-buffered saline

PCR: Polymerase chain reaction

PDB: Protein databank

pIC: Polyinosinic:polycytidylic acid

Pj-AID: *Polymixia japonica* AID

PKA: Protein kinase A

Pm-CDA1: *Petromyzon marinus* cytidine deaminase 1

Pp-AID: *Phycis phycis* AID

PRR: Pattern recognition receptor

Pt-AID: *Percopsis transmontana* AID

Pv-AID: *Pollachius virens* AID

Pw-AID: *Pleurodeles waltl* AID

RACE: Rapid amplification of cDNA ends

RAG: Recombination-activating gene

RIG: Retinoic acid-inducible gene

RLR: Retinoic acid-inducible gene-I-like receptor

RNA: Ribonucleic acid

RPA: Replication protein A

rplp1: 60S acidic ribosomal protein P1

RSS: Recombination signal sequences

RT: Reverse transcriptase

RT-PCR: Reverse transcription polymerase chain reaction

S: Immunoglobulin switch region

Sc-AID: *Stylepnorus chordatus* AID

SCS: Subcapsular sinus

SHM: Somatic hypermutation

SLC: Surrogate light chain

ss: Single-stranded

Ss-AID-1: *Salmo salar* AID variant 1

Ss-AID-2: *Salmo salar* AID variant 2 (Ss-AID-1^{V41G})

SSBs: Single-stranded breaks in DNA

ssDNA: Single-stranded DNA

ssRNA: Single-stranded RNA

STING: Stimulator of interferon genes

SV: Chromosomal structural variation

T1 B cell: Transitional 1 B cell

T2 B cell: Transitional 2 B cell

TADs: tRNA deaminases

TCR: T cell receptor

TD: T cell-dependent

TDG: Thymidine DNA glycosylase

TdT: Terminal deoxynucleotidyl transferase

TFH: Follicular TH

TGF-β: Transforming growth factor-beta

T-Gm-AID: *Gadus morhua* AID truncated isoform

TI: T cell-independent

TI-1: T cell-independent antigens 1

TI-2: T cell-independent antigens 2

TLR: Toll-like receptor

Tmi-AID: *Trisopterus minutus* AID

Tmu-AID: *Trachyrincus murrayi* AID

TNF: Tumor necrosis factor

Tr-AID: *Takifugu rubripes* AID

TREG: Peripheral regulatory TH

TRIF: TIR domain-containing adaptor-inducing IFN-β factor

Tsc-AID: *Trachyrincus scabrus* AID

TSS: Transcription start site

Tsu-AID: *Typhlichthys subterraneus* AID

TS-WGD: Teleost-specific whole-genome duplication

UDG: Uracil-DNA glycosylase enzyme

UNG: Uracil-N-glycosylase

V: Immunoglobulin variable domain

V_H: V region heavy chain

V_L: V region light chain

Xl-AID: *Xenopus laevis* AID

Zf-AID: *Zeus faber* AID

Zf-ANC: Zeiogadaria ancestral AID

α: α-helix

Ψ: Pseudogenes

List of Appendices

Author contributions

Chapter 2: I designed the research proposal, performed the experiments, analyzed the data, and wrote the manuscript. I conducted these experiments at Dr. Rise's lab at Ocean Sciences Center, Memorial University, NL, Canada. K. Eslamloo assisted with fish dissection, RNA extraction, and qPCR analyses. M. Rise provided guidance in qPCR analyses and edited the manuscript. M. Larijani was the principal investigator.

Chapter 3: I designed the research proposal, performed the experiments, analyzed the data, and wrote the manuscript. I conducted these experiments at Dr. Larijani's lab at Health Sciences Center (HSC), Memorial University, NL, Canada. D. N. Hubert, K. X. N. Hernandez, and I performed computational modeling and DNA/protein docking. M. H. Solbakken characterized the Atlantic cod *Ig* loci at the Center for Ecological and Evolutionary Synthesis (CEES), University of Oslo, Oslo, Norway. I conducted the gene synteny analyses. S. J. Khataeipour and I performed WRC analyses. S. Jentoft provided guidance for characterizing the Atlantic cod *Ig* loci and edited the manuscript. M. Larijani was the principal investigator.

Chapter 4: I designed the research proposal, performed the experiments, analyzed the data, and wrote the manuscript. I conducted the biochemical characterization of AID homologs at Dr. Larijani's lab at HSC. I performed the AID gene identification and ancestral sequence reconstruction analyses at CEES. S. J. Kataeipour and I performed the machine learning analyses. S. J. Khataeipour contributed to the writing of the manuscript.

C. D. Collins assisted in protein purification and biochemical analyses. M. Larijani was the principal investigator.

Peer-reviewed publications

Branton S. A., **Ghorbani A**., Bolt B. N., Fifield H., Berghuis L. M., and Larijani M., "Activation-induced cytidine deaminase can target multiple topologies of double-stranded DNA in a transcription-independent manner". The FASEB Journal, 34 (7), 9245-9268 (2020)

Eslamloo K., **Ghorbani A.**, Xue X., Inkpen S. M., Larijani M., and Rise M. L., "Characterisation and expression analyses of Atlantic cod *viperin*". *Frontiers in Immunology*, 10:311 (2019)

Abdouni H.S., King J.J., **Ghorbani A.**, Fifield H., Berghuis L., and Larijani M., "DNA/RNA hybrid substrates modulate the catalytic activity of purified AID". *Molecular Immunology*, 93, 94-106 (2018)

Daliri K., Aref-Eshghi E., Taranejoo S., Modarresi S., **Ghorbani A.**, Nariman A., Savaie M., Falasiri S. M. M., Akhondi-Kharangh F., and Askari H., "Emerging cytokines in allergic airway inflammation: A genetic update". *Current Immunology Reviews*, 12(1), 4-9 (2016)

Naghavi F. S., Hanachi P., Soudi M. R., Saboora A., and **Ghorbani A.**, "Evaluation of the relationship between the incubation time, and carotenoid production in *Rhodotorula slooffiae* and *R. mucilaginosa* isolated from leather tanning wastewater". *Iranian Journal of Basic Medical Sciences*, 16(10), 1114–1118 (2013)

Ghorbani A., Solbakken M. H., Huebert D. N., Eslamloo K., Berghuis L. M., Jentoft S., Rise M. L., and Larijani M., "Evolutionary trajectory of activation-induced

cytidine deaminase (AID) in the extant and ancestral Gadiformes species". This paper contains the results of chapter 2, and most of chapter 3, and 4. *Manuscript in preparation*.

Ghorbani A.[*], Quinlan E. M.[*], and Larijani M., "Evolutionary comparative analyses of DNA-editing enzymes of the immune system: 5-dimensional structures, immunological insights, and applications to protein engineering". This review paper contains materials from chapter 1 and 5. *Manuscript under revision in the journal of Frontiers in Immunology*.

[*]Denotes joint-first authorship.

Ghorbani A., King J. J., and Larijani M., "DNA-binding residues proximal to its catalytic pocket regulate pH sensitivity of activation-induced cytidine deaminase (AID)". This paper contains some of the findings from chapter 3. *Manuscript in preparation.*

Chapter 1:

Introduction

1.1 Overview

The immune system, which is the species' defense mechanism against pathogens and abnormal self, is comprised of two major arms that recognize a wide variety of antigens: innate and adaptive immunity. Innate immunity is the first line of defense that reacts quickly but non-specifically to a wide variety of pathogens. Adaptive immunity is highly antigen-specific but requires a longer time (days to weeks) to develop to its full measure and effectiveness. The innate immune system consists of physical and chemical barriers, such as epithelial layers, stomach acid, lysozyme, *etc.*, and cellular component that includes pattern recognition receptors (PRRs) and innate immune cells. In jawed vertebrates, B and T lymphocytes are the evolutionarily conserved major cell types involved in adaptive immunity, mediating antibody (humoral) and cell-mediated immunity, respectively (Owen, 2019). The innate and adaptive immune systems protect the host against pathogens by working individually and in collaboration with each other.

1.2 Innate immune system

Upon exposure to a pathogen, innate immunity components are effective immediately or rapidly induced. The physical and chemical barriers of the innate immune system are the body's first defensive structures. Physical barriers include the epithelial layer that isolates the body's interior from the outside world. Chemical barriers consist of any substances that exert antibacterial activity, such as stomach acid, fatty acids in sebum, mucus, proteins, peptides, *etc.* Lysozyme, lactoferrin, surfactant proteins, S100 proteins, defensins, cathelicidin, histatins, and dermcidin are some examples of innate antimicrobial proteins and peptides (Owen, 2019).

If the pathogen manages to overcome these physical and chemical barriers, then the cellular component of the innate immune system is quickly induced. Myeloid cells, such as macrophages and granulocytes, and the innate lymphoid cells (ILCs), such as natural killer cells (NKs), are the main cell types of the innate cellular response. The cellular components of the innate immune response are triggered when PRRs, expressed by the immune cells, interact with pathogen-associated molecular patterns (PAMPs), or damage-associated molecular patterns (DAMPs) released by the host's damaged or dying cells (Owen, 2019; Roh & Sohn, 2018). The activation of cellular modules results in the release of antimicrobial molecules, cytokines, and chemokines, as well as the induction of the phagocytic and killing activity of immune cells (Owen, 2019).[a] Innate cellular responses constitute the host's second line of defense against pathogen invasion.

1.2.1 Overview of innate cellular immunity

The essential step in initiating an innate cellular response is the recognition of PAMPs and DAMPs by PRRs. PAMPs are the unique antigenic structures specific to a group of pathogens (Owen, 2019; Rajaee et al., 2018). Lipopolysaccharides (LPS), lipoproteins, peptidoglycans, lipoteichoic acid, flagellin, unmethylated CpG dinucleotides, and rRNA are some examples of the bacterial PAMPs (Kumar et al., 2011; S. Kumar et al., 2013; Owen, 2019; Rajaee et al., 2018). Some of the known viral PAMPs are single-stranded (ss) and double-stranded (ds) RNA and coat proteins, such as the fusion protein of respiratory syncytial virus and the G glycoprotein of vesicular stomatitis virus. Zymosan

[a] Another important part of the innate immune response is the activation of the complement system through PAMP recognition by lectins.

(β-glucan) and mannans are widely considered the predominant fungal PAMPs (Goyal et al., 2018; Kumar et al., 2011; Owen, 2019). Glycosylphosphatidylinositol-anchored mucin-like glycoproteins and hemozoin are examples of the parasite-related PAMPs (Kawai & Akira, 2011). DAMPs are the host's endogenous danger signals that originate either from the extracellular matrix or intracellular compartments as the result of cell and tissue injury. Biglycan, tenascin C, and fibrinogen are derived from the extracellular matrix, while S100 proteins, heat shock proteins, F-actin, ATP, histones, and mitochondrial DNA are some of the DAMPs formed from intracellular content (Roh & Sohn, 2018). Since PAMPs and DAMPs are not specific to individual pathogens, the specificity of the innate immune response is limited (Owen, 2019).

1.2.2 Pattern recognition receptors

PRRs are membrane-bound or cytosolic proteins which include six protein families: (1) Toll-like receptors (TLRs), (2) C-type lectin receptors (CLRs), (3) nucleotide oligomerization domain (NOD)-like receptors (NLRs), (4) absent in melanoma 2 (AIM2)-like receptors (ALRs), (5) retinoic acid-inducible gene (RIG)-I-like receptors (RLRs), and (6) cytosolic DNA sensors (*i.e.*, cyclic GMP-AMP synthase [cGAS], and stimulator of interferon genes [STING]) (Owen, 2019; Thompson et al., 2011). PRR engagement with PAMPs and DAMPs commences a series of complex signaling pathways that activate the innate immune cells to induce the proper effector mechanisms (Jain & Pasare, 2017).[a] It

[a] Although the innate cell type impacts the exact consequence of PRR crosstalk, the general prompted effector mechanisms depend on the detected pathogen. For example, interferons (IFNs) are expressed against viral infection, while activated phagocytes target extracellular bacteria, and programmed cell death is induced in infected cells

must be emphasized that although individual PRRs are strong immunomodulators, the simultaneous engagement of multiple PRRs is required to mount a robust adaptive immune response (Jain & Pasare, 2017). In essence, the diversity in the host's arsenal of PRRs is an evolutionary strategy to tailor the immune response to a specific group of pathogens.

1.2.2.1 Toll-like receptors

Thus far, the best characterized PRRs are TLRs. TLRs are type I transmembrane proteins found either on the plasma membrane, which recognize PAMPs on the outside of pathogens, or on the endosomal/lysosomal membrane that detect the released PAMPs during endosomal/lysosomal degradation of pathogens (Owen, 2019; Thompson et al., 2011). The mammalian TLR family consists of 13 members of which ten members are found in humans (TLR1 to 10), and 12 of which have been discovered in mice (TLR1 to 9 and TLR 11 to 13) (Takeda & Akira, 2015). Table 1-1 compares some of the TLRs' characteristics, ligands, and expression pattern across various immune cell types (Doan, 2013; Koblansky et al., 2013; Owen, 2019; Shi et al., 2011; Takeda & Akira, 2015).[a] Interaction with its cognate ligand facilitates TLR dimerization into either a heterodimer or a homodimer. Two key adaptor proteins which associate with the cytoplasmic domain of dimerized TLRs are MyD88 (myeloid differentiation factor 88) and TRIF (TIR domain-containing adaptor-inducing IFN-β factor) (Takeda & Akira, 2015).[b] A shared component of all TLR signaling pathways is the activation of the NF-κB (nuclear factor kappa-light-

[a] The TLRs extracellular ligand-binding domain is made up of leucine-rich repeats (LRRs), and their intracellular domain is called the Toll/IL-1 receptor (TIR) domain due to shared structural similarity with the interleukin 1 (IL-1) receptor family.
[b] MyD88 interacts with all TLRs except TLR3, while TRIF only associates with TLR3 and endosomal TLR4.

chain-enhancer of activated B cells) transcription factor (Owen, 2019). TLR-induced signaling pathways result in the secretion of cytokines, chemokines, and antimicrobial proteins.

The MyD88-dependent pathway generally activates three transcription factors: NF-κB, activator protein 1 (AP-1, through mitogen-activated protein kinase [MAPK] pathway), and IRF7 (interferon regulatory factor 7, only in the case of TLR7 to 9). TLR 7 to 9 are lysosomal and bind microbial nucleic acids. IRF7 activation guarantees the expression of potent antiviral interferon genes downstream of TLRs that detect viral components. TRIF-dependent signaling, however, triggers the expression of type I interferons (*i.e.,* IFN-α and IFN-β) through the activation of IRF3 and induces delayed activation of NF-κB (Owen, 2019; Takeda & Akira, 2015). The abovementioned variations in the TLR signaling pathway enable tailoring of the innate cellular immune response to the specific group of pathogens detected by TLRs.

Table 1-1: Characteristics of mammalian Toll-like receptors (TLRs)

	Ligands	Microbes	Dimerization	Location	Organism found in	Expressed on
TLR1	Triacyl lipopeptides	Mycobacteria Gram-negative bacteria	TLR2/1	Plasma membrane	Human Mouse	Monocytes/macrophages Dendritic cell subset B lymphocytes
TLR2	Peptidoglycans lipoteichoic acid Lipomannan, lipoproteins GPI-anchored proteins Zymosan Phosphatidylserine	Gram-positive bacteria Gram-positive bacteria Other bacteria Mycobacteria Trypanosomes Yeast and other fungi Schistosomes	TLR2/1 and TLR2/6	Plasma membrane	Human Mouse	Monocytes/macrophages Subset of dendritic cells Mast cells
TLR3	Double-stranded RNA	Viruses	Homodimer	Endosomal membrane	Human Mouse	Dendritic cell B lymphocytes
TLR4	LPS F protein Envelope glycoprotein G glycoprotein Mannans Heat shock proteins Extra domain A (EDA) Hyaluronic acid	Gram-negative bacteria Respiratory syncytial virus Mouse mammary tumor virus Vesicular stomatitis virus Fungi	Homodimer	Plasma membrane Endosomal membrane	Human Mouse	Monocytes/macrophages Dendritic cell subset Mast cells Intestinal epithelium
TLR5	Flagellin	Bacteria	Homodimer	Plasma membrane	Human Mouse	Monocytes/macrophages Dendritic cell subset Intestinal epithelium
TLR6	Peptidoglycans Diacyl lipopeptides	Gram-positive bacteria Gram-negative bacteria	TLR2/6	Plasma membrane	Human Mouse	Monocytes/macrophages Mast cells

	Ligands	Microbes	Dimerization	Location	Organism found in	Expressed on
	Zymosan	Mycobacteria Yeast and other fungi				B lymphocytes
TLR7	G-/U-rich ss RNA Imidazoquinoline	Viruses	Homodimer	Endosomal membrane	Human Mouse	Monocytes/macrophages Dendritic cell subset B lymphocytes
TLR8	ssRNA Imidazoquinoline	Viruses	Homodimer	Endosomal membrane	Human Mouse	Monocytes/macrophages Dendritic cell subset Mast cells
TLR9	CpG unmethylated dinucleotides Dinucleotides Herpesvirus components Hemozoin	Bacterial DNA Some herpesviruses Malaria parasite heme by-product	Homodimer	Endosomal membrane	Human Mouse	Monocytes/macrophages Dendritic cell subset B lymphocytes
TLR10	Unknown	Unknown	Homodimer	Plasma membrane	Human	Monocytes/macrophages B lymphocytes
TLR11	Unknown Profilin Flagellin	Uropathogenic bacteria Toxoplasma gondii Salmonella typhimurium	Homodimer and TLR11/12	Plasma membrane	Human (non-functional) Mouse	Macrophages Liver epithelial cells Kidney epithelial cells Bladder epithelial cells
TLR12	Profilin	Toxoplasma gondii	Homodimer and TLR11/12	Plasma membrane	Mouse	Macrophages Dendritic cell subset
TLR13	rRNA Unknown	Bacteria Vesicular stomatitis virus	Homodimer	Plasma membrane	Mouse	Macrophages Dendritic cell subset

1.2.2.2 Other types of pattern recognition receptors

The myeloid C-type lectin receptors (CLRs) are also involved in PAMPs recognition by the immune system. CLRs are type I or type II transmembrane proteins, characterized by exoplasmic space-located N-terminal or C-terminal domains, respectively (Brown et al., 2018; Cao, 2018; Mayer et al., 2017). The myeloid CLRs, expressed on the surface of monocytes, macrophages (MQs), dendritic cells (DCs), and neutrophils, engage with carbohydrates PAMPs, on the surface of extracellular pathogens, in a Ca^{2+}-dependent manner (Brown et al., 2018; Mayer et al., 2017; Owen, 2019).[a] CLRs engagement with the corresponding ligand triggers various antimicrobial effector mechanisms such as the respiratory burst and the formation of neutrophil extracellular traps (NETs). It also stimulates the expression of different cytokines, chemokines, and immunomodulatory lipids (*e.g.*, eicosanoids) (Brown et al., 2018).

Unlike TLRs and CLRs, other PRRs are cytosolic receptors. NLRs respond to various PAMPs and DAMPs, and their functions can be divided into inflammasome formation, signaling transduction, transcription activation, and autophagy (Kim et al., 2016; Yang et al., 2019). Association of cytosolic bacterial and viral DNA by ALRs leads to inflammasome formation and subsequent cytokine maturation and pyroptotic cell death (Wang et al., 2019). An antiviral response[b] is mounted when the RLRs detect the presence of viral RNA in the cytoplasm and begin signaling pathways resulting in NF-κB and IRF

[a] Dectin-1, dectin-2, macrophage inducible cytotoxic T lymphocyte (Mincle), and dendritic cell-specific intercellular adhesion molecule3-grabbing non-integrin (DC-SIGN) are some examples of CLRs.
[b] *i.e.*, expression of interferons and inflammatory cytokines.

(mainly IRF3 and IRF7) activation (Barik, 2016). The non-self DNA and dinucleotides can be sensed through cGAS and STING. This association prompts the activation of NF-κB and IRF3 initiating type I IFN and cytokine synthesis (Owen, 2019).

1.2.3 The innate immune cells

The innate immune cells are the major players in the cellular response of innate immunity. These cells divide into two main groups: the innate myeloid cells (IMCs) and the innate lymphoid cells (ILCs). IMCs include granulocytes (*i.e.*, neutrophils, eosinophils, basophils, and mast cells), monocytes, MQs, and DCs (Palgen et al., 2018). IMCs are the first cells to respond to pathogen invasion. Their activation, through PRR engagement with PAMPs and/or DAMPs, triggers the secretion of antimicrobial molecules, cytokines, and chemokines and stimulates the phagocytosis of the pathogens or the infected cells (only in the case of phagocytic cells) (Owen, 2019). Moreover, monocytes, MQs, and DCs are also considered professional antigen-presenting cells (pAPCs). These pAPCs present exogenous antigens in the context of major histocompatibility complex type II (MHC II) to helper T (T_H) cells to activate the adaptive immune system (Owen, 2019; Palgen et al., 2018).

Amongst APCs, DCs are unique in that they are activated (also referred to as licensed) when their PRRs interact with PAMPs or DAMPs, which causes DCs to uptake pathogens and then break them down into short peptide fragments to load onto their MHC I or II (depending on the nature of the pathogen) molecules (Owen, 2019; Reis e Sousa, 2004). Alternatively, the licensing process happens when an activated $CD40L^+$ helper T cell (*e.g.*, T_H1) engages with the MHC II-peptide complex found on a DC's surface (Owen,

2019). Consequently, they begin to express costimulatory receptors and secrete cytokines that are essential for the activation of naïve T cells. Due to these activation events, DCs are the only APCs capable of activating naïve T cells (Owen, 2019; Reis e Sousa, 2004). Therefore, DC activation is an essential event that interlinks the innate and adaptive immune systems.

Generally, ILCs are functionally parallel to the T cell grouping and secret similar cytokine profiles. However, they lack antigen-specific receptors and mainly reside in peripheral tissues, specifically at barrier surfaces, except for the NK cells that circulate in the bloodstream (Kotas & Locksley, 2018; Mjosberg & Spits, 2016; Vivier et al., 2018). ILCs are activated through cytokines and stress signals generated from mucosal stromal and myeloid cells (Mjosberg & Spits, 2016). The International Union of Immunological Societies (IUIS) has divided ILCs based on their secreting cytokine profile, required transcription factors, and development into five subsets: NK cells, ILC1s, ILC2s, ILC3s, and lymphoid tissue inducer (LTi) cells (Colonna, 2018; Vivier et al., 2018).

NK cells are efficient cell killers and the most studied ILCs. They possess various types of receptors that are either inhibitory or activating in nature (Mandal & Viswanathan, 2015). For example, they express inhibitory receptors for self MHC I. When infected cells downregulate the surface expression of MHC I, they are targeted by NK cells. NK cells also express surface receptors for the F_c fragment of antibodies (*i.e.*, F_c receptor). Through these receptors, they can pick up antibodies and attack infected cells or bind antibodies attached to infected cells in an antigen-specific manner. This mechanism is called antibody-dependent cell cytotoxicity (ADCC) (Owen, 2019). Remarkably, the NK cell population

exhibits a significant heterogeneity due to the ability of different NK cells to express various combinations of activating and inhibitory receptors (Mandal & Viswanathan, 2015).

ILC1s, ILC2s, and ILC3s are functionally reminiscent of $CD4^+$ T helper type 1 cells (T_H1), T_H2, and $T_H17/22$ subsets of adaptive immune cells, respectively (Kotas & Locksley, 2018; Mjosberg & Spits, 2016). ILC1s react to viruses and tumors, while ILC2s defend against large extracellular parasites and allergens, and ILC3s fight extracellular microbes (Vivier et al., 2018). During fetal development, LTi cells play a central role in lymphoid organogenesis (Mjosberg & Spits, 2016; Vivier et al., 2018; Zhong et al., 2018). Although our knowledge of ILCs is still growing, the important role of innate immune cells in protecting our body is apparent.

1.3 Adaptive immune system

In jawed vertebrates, innate immunity plays another important role beyond initiating immediate defense; it activates the more efficient long-standing adaptive immune response. T cell and B cells are the main adaptive lymphoid cells. T cells are divided into two cell types based on the CD4 or CD8 surface expression. $CD4^+$ T cells are called T helper cells (T_H) that "help" the activation and regulation of the other effector immune cells. $CD8^+$ T cells, natural killer T cells (NKT), and NKs are cytotoxic effector cells that establish cell-mediated immunity. They induce cell death by triggering apoptosis in their target cells (Owen, 2019). On the other hand, activated B cells, known as plasma cells, mount the humoral immune response by secreting antibodies and are the only non-myeloid

pAPC (Owen, 2019). Thus, T/NK/B cell-mediated cellular and humoral immunity constitute the adaptive immune system in all studied jawed vertebrates.

1.3.1 Overview of T cells

T cells are one of the main adaptive immune cells. Their development occurs in the thymus where they rearrange one of their T cell receptor (TCR) subtypes (*i.e.*, αβ or γδ), undergo positive and negative selections to acquire MHC restriction and self-tolerance, respectively, and commit to either $CD4^+$ helper or $CD8^+$ cytotoxic lineages. The cells that complete their development successfully enter the bloodstream as naïve T cells (Owen, 2019). The $CD4^+$ helper T cells engage with the MHC II-peptide complex, presented only by APCs, while the $CD8^+$ cytotoxic T cells recognize peptides in the context of MHC I molecules, expressed by all nucleated cells (Owen, 2019). Upon activation, naïve T cells give rise to effector and memory T cells. $CD8^+$ effector T cells are called cytotoxic T lymphocytes (CTLs). However, the naïve $CD4^+$ T_H cells can differentiate into different effector subtypes such as T_H1, T_H2, T_H9, T_H17, T_H22, follicular T_H (T_{FH}), and peripheral regulatory T_H (T_{REG}) cells (Kmiec et al., 2017; Owen, 2019; Takeuchi & Saito, 2017).[a]

1.3.1.1 T cell activation

An adaptive immune response is initiated when naïve T cells are activated through three signals, a process known as the three-signal paradigm (Jain & Pasare, 2017). The first signal is the formation of an immune synapse between a naïve T cell and a licensed DC.

[a] Note that the distinction between T_H and CTLs is not absolute. Some CTLs may play a T_H cell-like role by secreting a verity of cytokines to impact other cell types. Also, $CD4^+$ T cells that secrete granzyme B and perforin can exert CTL-like cytotoxic activity.

The immune synapse is established between the T cell receptor (TCR) on the naïve T cell and the MHC-peptide complex presented on the licensed DC. The concurrent interaction of T cell CD8 or CD4 coreceptors with DC MHC I or II, respectively, strengthens the immune synapse. Following this, the interaction between the constitutively expressed CD28 costimulatory receptor on the naïve T cell and its ligands on the DC (*i.e.*, CD80 and CD86, expressed only upon activation through PRR/PAMPs engagement) provides the second signal. The local cytokines deliver the third signal when they bind to their receptors located on naïve T cells (Jain & Pasare, 2017; Owen, 2019).

Amongst the local cytokines, IL-2 is essential for optimal activation and proliferation of all T cell subtypes. Therefore, upon receiving the first two signals, the T cell begins to secrete IL-2 and express the high-affinity IL-2 receptors on its surface (Owen, 2019). On the other hand, a subset of local cytokines known as polarizing or priming cytokines dictates the specific fate of the activated T cell (Jain & Pasare, 2017; Owen, 2019). Although non-immune cells can contribute to the production of some cytokines such as IL-1 related cytokines, polarizing cytokines are exclusively secreted by immune cells (Jain & Pasare, 2017). Thus, the abovementioned three signals initiate a network of signaling pathways that culminate in cell survival, proliferation, and differentiation of naïve T cells into specific effector T cell subsets.

1.3.1.2 Cytotoxic T lymphocytes

The naïve CTLs, also referred to as CTL precursors, are a subset of the T lymphocytes that express CD8 receptor and, thus, recognize foreign antigens in the context of MHC I molecules (Owen, 2019; Williams & Bevan, 2007). Since all nucleated cells

express MHC I on their surface, CTLs identify and eliminate altered self-cells (*e.g.*, a virus-infected or a cancerous cell) by mounting a cytotoxic reaction and lysing the target cell (Owen, 2019). Activated CTLs form conjugates with the target cells followed by membrane attack, CTL dissociation, and target cell destruction (Owen, 2019). Secretion of perforin and granzymes from CTLs initiates the target cell destruction through pore formation in the cell membrane and genomic DNA fragmentation, respectively. Some CTLs that lack perforin and granzymes express Fas ligand (FasL) on their membrane and deliver a death signal to the target cell through Fas/FasL interaction. Both granzymes and Fas/FasL interaction activate an initiator caspase, initiating death pathways within the target cell (Owen, 2019).

1.3.1.3 Helper T cell

The secreted IL-12 and IFN-γ, in response to intracellular pathogens, induce differentiation of naïve $CD4^+$ helper T cells into T_H1 cells, which subsequently secrete IFN-γ and tumor necrosis factor (TNF) (Kmiec et al., 2017). T_H1 IFN-γ enhances the APC activity of MQs, antibody class switching to IgG classes in B cells,[a] and promotes CTL differentiation and activation (Owen, 2019). Hence, T_H1 promotes cell-mediated immunity. In response to extracellular microbes, IL-4 is secreted, driving the differentiation of T_H2 cells (Kmiec et al., 2017). Effector T_H2 cells secrete various cytokines, including IL-4, IL-5, and IL-13. By activating B cells, eosinophils, and MQs,

[a] IgG classes enhance phagocytosis and fixation of complement

inducing IgE antibody class switching, and inhibiting T_H1 development, T_H2 cells protect against parasitic worms (Kmiec et al., 2017; Owen, 2019).

T_H9 development requires a combination of IL-4 and transforming growth factor-beta (TGF-β). T_H9 cells produce IL-9 that contributes to protection against worm infections and possibly cancer (Kmiec et al., 2017; Owen, 2019). T_H17 subtype is divided into anti-inflammatory and pro-inflammatory subsets. Anti-inflammatory (also known as non-pathogenic) T_H17 cells are developed in response to TGF-β and IL-6, secrete immunosuppressive cytokine IL-10, and inflammatory cytokine IL-17 and IL-21. Pro-inflammatory (also known as pathogenic) T_H17 cells are established in the presence of TGF-β, IL-6, and IL-23 and produce only inflammatory cytokines IL-17, IL-21, and IL-22 (Wu et al., 2018). Although pro-inflammatory T_H17 enhances protection against bacterial and fungal infections at mucosal barriers, it can also result in chronic inflammatory and autoimmune diseases (Lee, 2018; Owen, 2019; Wu et al., 2018). IL-6, TNF, and IL-23 induce T_H22 differentiation. These cells secrete IL-13 and IL-22, contribute to wound repair, and protect against infections at epithelial surfaces (Owen, 2019; Wu et al., 2018).

T_{FH} cells are produced in response to IL-6 and IL-21, which are secreted by activated APCs. They induce differentiation of B cells into plasma cells and are vital for the germinal center (GC) formation and antibody affinity maturation (Kmiec et al., 2017; Owen, 2019). T_{FH} cells are unique in that they require signals from both coreceptor CD28 and ICOS to fully develop. These cells secrete IL-4 and IL-21 and express high levels of surface CD40L, all of which are necessary for B cell activation. They also express the

chemokine CXCR5 that enables them to move to B cell follicles to establish germinal centers (Kmiec et al., 2017; Owen, 2019). Thus, T_{FH} cells promote humoral immunity.

T_{REG} cells play an important negative regulatory role in preventing autoimmune and chronic inflammatory diseases by negatively regulating immune responses and maintaining peripheral tolerance. In the periphery, TGF-β drives T_{REG} differentiation in the absence of proinflammatory cytokines (Kmiec et al., 2017; Lee, 2018). Beside these periphery-derived T_{REG} cells (pT_{REG}), T_{REG} cells can also arise during thymic development (tT_{REG}) when the developing T cell receives a strong TCR stimulation by self-antigen-MHC complexes presented on thymic APCs (Kmiec et al., 2017; Lee, 2018; Owen et al., 2019). These cells secrete potent anti-inflammatory cytokines of IL-10 and TGF-β, which suppress the activity of immune cells (Kmiec et al., 2017; Lee, 2018; Owen, 2019). Differentiation of naïve T_H cells into distinct subsets of effector T_H cells is one of the host's adaptive immune system's main regulatory mechanisms.

1.3.2 Overview of B cells

Humoral immunity fights off a wide range of pathogens, and its activation is the basis of most vaccines (Dickinson et al., 2015). B cells mediate humoral immunity by secreting antibodies that neutralize and/or opsonize the pathogens and/or toxins. There are four distinct subsets of mature B cells[a] that differ in terms of development, location, and function: marginal zone (MZ), B-1a, B-1b, and conventional B-2 cells (Dickinson et al.,

[a] Other minor B cell subtypes with innate-like functions have been detected such as innate response activator B cells, T-bet$^+$ B cells, natural killer-like B cells, IL-17-producing B cells, and human self-reactive V_H4-34-expressing B cells. For more information refer to Tsay and Zouali, 2018.

2015; Ghosn et al., 2019; Haas, 2015; Montecino-Rodriguez et al., 2006).[a] In general, B-1 and MZ B cells are involved in T cell-independent humoral immunity, while B-2 cells mediate the T cell-dependent antibody responses (Dickinson et al., 2015; Owen, 2019). Regardless of their differences, all B cells undergo DNA recombination events to create their B cell receptor (BCR; *i.e.*, membrane-bound antibody), and they all secrete antibodies to fight pathogens. Despite the developmental, distributional, and functional differences amongst B cell subtypes, they all play important roles in the host's humoral immunity.

1.3.2.1 B-1 and marginal zone B cells

In mice, B-1a cells reside in the spleen and the pleural and peritoneal cavities, where the pre-existing B-1a cells divide to regenerate new ones (Ghosn et al., 2019; Haas, 2015; Wong et al., 2019).[b] B-1a cells predominantly rearrange some specific heavy and light chain gene fragments and minimally express the enzyme terminal deoxynucleotidyl transferase (TdT) that generates junctional diversity during *Ig* gene recombination. Consequently, the B-1a cell antibody repertoire has limited diversity, especially in their heavy chain complementarity determining region 3 (CDR3) (Owen, 2019; Wong et al., 2019). Additionally, their antibodies (mainly recognize carbohydrate and lipid antigens) are polyreactive, cross-react between self and microbial antigens, and are secreted spontaneously in the absence of the cognate antigen and T_H assistance (Haas, 2015; Owen,

[a] In mice, B-1 cells consist of two subtypes of $CD5^+$ B-1a and $CD5^-$ B-1b cells.
[b] In the mouse fetal liver (FL), the pro-B cells experience a lower level of IL7R/pSTAT signaling that causes concurrent rearrangement of *Ig* heavy and light chains. This phenomenon results in bypassing the conventional step of pairing the heavy chain with the surrogate light chain (SLC). Poor binding of SLC to autoreactive heavy chains contributes to the elimination of autoreactive B cells. Consequently, this alternative B cell development in the FL promotes the formation of autoreactive B cells, which give rise to B1-a cell progenitors (Wong *et al.*, 2019).

2019; Palma et al., 2018).[a] Therefore, B-1a cells are considered innate-like cells. B-1a antibodies, known as natural antibodies (nABs),[b] support the immune system by providing fast protection against bacterial, viral, and parasitic infections (Haas, 2015; Hillion et al., 2019; Palma et al., 2018; Wong et al., 2019).[c] Moreover, the self-reactive properties of nABs assist in tissue homeostasis by clearing the dead cells and debris (Palma et al., 2018).[d] The natural IgM may also contribute to immune system homeostasis by removing autoreactive B cells during B cell development in the bone marrow (Nguyen et al., 2015).[e]

Although B-1b cells inhabit similar anatomical sites as B-1a cells, they are developmentally and functionally distinct (Baumgarth, 2011; Dickinson et al., 2015; Ghosn et al., 2019). In adult mice, their cell pool is maintained by self-renewal of pre-existing B-1b cells (Baumgarth, 2011).[f] Unlike B-1a cells, B1-b cells undergo clonal selection and secrete antibodies with a diversity comparable to that of B-2 cell antibodies (Dickinson et al., 2015). Only upon pathogen exposure, but without T_H assistance, B-1b cells are activated and secrete antibodies (Dickinson et al., 2015). B-1b antibodies detect

---

[a] However, the presence of the T_H cells can enhance their antibody secretion and mediate some degree of secondary antibody diversification. Secondary antibody diversification includes two processes: class switch recombination (CSR) and antibody affinity maturation (AM) to change the effector function of antibody and to increase the affinity of the antibody for cognate antigen, respectively. These events will be further discussed in the following sections. The T cell independent activation of non-conventional B cells results in the lack of or minimal secondary antibody diversification. This is a safeguard to avoid generation of high-affinity autoreactive antibodies.

[b] nABs are defined as antibodies that are secreted in normal conditions without antigen presence. These antibodies are highly cross-reactive and bind a wide range of antigens with low affinity. B-1 and MZ B cells mainly secrete these antibodies as an IgM isotype with a small fraction of IgG, IgA, and IgE. nABs contribute to pathogen resolution, dead cell clearance, control of inflammation and autoimmune responses, and the regulation of B cell development and activation. The autoreactive nABs may participate in the pathogenic response of autoimmune disorders such as rheumatoid arthritis.

[c] Such as *Streptococcus pneumoniae*, *Francisella tularensis*, and influenza virus.

[d] B1-a cell antibodies recognize conserved self antigens such as phosphatidylcholine.

[e] The process of removing autoreactive B cells in the bone marrow is called central tolerance.

[f] A precursor cell in fetal liver and adult bone marrow can give rise to pre-existing B-1b cells.

bacterial proteins, polysaccharides, and synthetic hapten. For example, B-1b cells mount specific antibody responses towards capsular polysaccharide antigens (also known as type-2 T cell independent [TI-2] antigens) of *Streptococcus pneumoniae*, *Salmonella enterica*, and *Enterobacter cloacae* (Dickinson et al., 2015). B-1b cells also produce IgA in response to mucosal pathogens (Tsay & Zouali, 2018). More importantly, activated B-1b cells can form memory cells and generate a lasting antibody response against pathogens such as *Borrelia hermsii* and *S. pneumoniae* (Haas, 2015). Thus, B-1a cells play a valuable role in early response against pathogens, while B-1b cells participate in long-lasting protective humoral immunity.

MZ B cells inhabit the splenic marginal zone, but they self-renew in the periphery (Owen, 2019). MZ B cells not only contribute to the first line of defense against bloodborne pathogens[a] by responding rapidly and efficiently to their antigens, but they also establish the primary antibody response towards TI-2 antigens (Owen, 2019; Zandvoort & Timens, 2002). These B cells respond to protein and carbohydrate antigens with or without T cell help (Dickinson et al., 2015). MZ B cells are generated from transitional 2 (T2) B cells when their BCR binds to a self-antigen with strong affinity and also receives signaling through the Notch2 pathway (Owen, 2019). MZ B cells have formed memory cells against some pathogens such as *Ehrlichia muris* (Zandvoort & Timens, 2002). Similar to B-1a cells, they express natural antibodies that are both self-reactive and polyreactive

[a] Some characteristics of MZ B cells and their environment contribute to their ability to respond rapidly to the bloodborne pathogens. These characteristics include low blood flow in their microenvironment, low activation threshold, high surface expression of complement receptor 2 (CR2, also known as CD21), and polyreactive BCR.

(Zandvoort & Timens, 2002). MZ B cell activation through commensal TI antigens may result in the initiation of the secondary antibody diversification and consequent class switching into IgG and IgA independent of GCs. Alternatively, T cell-dependent activation of these B cells may induce their migration into the B cell follicles, where they possibly contribute to GC formation and undergo secondary antibody diversification (Grasseau et al., 2019). Although B-1 and MZ B cells are considered innate-like B cells due to the property of their antibodies and their independence from T cell assistance, their vital contribution to humoral immunity is indisputable.

1.3.2.2 B-2 cells

B-2 cells, the prevalent B cell subtype in blood, arise from hematopoietic stem cells (HSCs) in the bone marrow and recirculate between blood and lymphoid organs (Outters et al., 2015; Owen, 2019; Yam-Puc et al., 2018). These cells are responsible for generating high-affinity, antigen-specific, antibody responses towards protein antigens with the T_{FH} cell assistance (Owen, 2019). Their development begins in the bone marrow, where HSCs give rise to immature B cells by enduring stepwise processes of immunoglobulin heavy and light chain (*IgH* and *IgL*) recombination, allelic exclusion,[a] and central tolerance (Outters et al., 2015; Owen, 2019; Yam-Puc et al., 2018).[b] Following these steps, the immature B cells (also known as transitional 1 [T1] B cells) migrate to the spleen, where

[a] Allelic exclusion happens after successful rearrangement of the *Ig* gene to ensure the expression of only one BCR per B cell. For more information regarding the regulation of this process, refer to Otters *et al.*, 2015.
[b] The elimination of self-reactive B cells in the bone marrow is called central tolerance. Upon successful expression of BCR, self-reactive B cells are eliminated through (1) BCR-induced apoptosis, (2) reactivation of the enzymatic machinery to edit their antibody light chain, or (3) development into anergic (unresponsive) B cells.

they undergo negative and positive selections to become mature conventional B-2 cells (Outters et al., 2015; Owen, 2019).[a] Since conventional B-2 cells are the major B cell subtype in humans and mice, they will be referred to as B cells hereafter. These mature, but naïve, B cells join the bloodstream and recirculate between blood and the lymphoid organs, where they enter B cell follicles to search for the cognate antigen (Owen, 2019).[b] Only 1-10% of the newly formed B cells survive and join the peripheral B cell pool (Granato et al., 2015). Antigen detection through their BCR activates these cells and culminates in their differentiation into either IgM secreting plasma cells or GC precursor cells to start GC formation (Outters et al., 2015; Owen, 2019; Yam-Puc et al., 2018). The B-2 cell activation initiates the adaptive humoral response against the cognate antigen.

1.3.2.3 Immunoglobulin protein structure and gene organization

Immunoglobulins are a member of the immunoglobulin superfamily (IgSF) characterized by their immunoglobulin (Ig) domain. This domain consists of two amphipathic β sheets (made from three to six β strands), hydrophobic sides of which are held together by hydrophobic forces (Owen, 2019; Schroeder & Cavacini, 2010). An intrachain disulfide bond connecting the two β sheets stabilizes this β-sheet-sandwich structure. Two heavy and two light chains create an antibody molecule with two distinct domains: the variable (V) domain that binds the antigen and the constant (C) domain that

[a] In the T zone of spleen, self-reactive T1 B cells are eliminated through negative selection. The B cells that survive the negative selection migrate into the follicular zone as T2 B cells. In the follicular zone, the tonic stimulation through BCR triggers B cell activating factor (BAFF) receptor expression. T2 B cells that succeed in expressing the BAFF receptor on their surface receive the required survival signal, hence are positively selected, and join the bloodstream as mature B cells. The negative selection in the spleen is essential for B cell peripheral tolerance.

[b] The conventional B-2 cells are also known as the follicular B cells.

determines the class of antibody and, therefore, its effector function (Schroeder & Cavacini, 2010). Within each immunoglobulin domain, loops, the loosely folded polypeptide chains, link each β strand with the adjacent ones (Owen, 2019; Schroeder & Cavacini, 2010). In the V domain, loops form the antigen-binding region (*i.e.*, CDR),[a] while the β-sheet-sandwich structure creates a stable framework. Three functionally distinctive hypervariable regions make up CDR: CDR1, 2, and 3. The V region of both light and heavy chains (V_L and V_H) contributes to each CDR conformation (Mix et al., 2006; Owen, 2019; Schroeder & Cavacini, 2010).

In humans and mice, the immunoglobulin gene family consists of one heavy and two light chain loci (*i.e.*, kappa [κ] and lambda [λ]) located on separate chromosomes (Owen, 2019; Schroeder & Cavacini, 2010; Tomlinson, 1998). At the DNA level, separate gene fragments of the variable (V), diversity (D, only in the heavy chain), and joining (J) segments are joined to create the antibody V region. However, one constant exon encodes the antibody's constant region. In humans, the heavy chain locus contains three separate clusters of approximately 45, 23 and 6 V, D, and J segments, respectively, followed by nine constant (C) exons (*i.e.*, C_μ, C_δ, $C_\gamma 3$, $C_\gamma 1$, $C_\alpha 1$, $C_\gamma 2$, $C_\gamma 4$, C_ε, $C_\alpha 2$) (Owen, 2019; Schroeder & Cavacini, 2010). The genomic structure of κ light-chain consists of a V cluster that roughly contains 41 functional V fragments and a J cluster that consists of five segments followed by one constant exon (*i.e.*, C_κ). However, the gene organization within

[a] Many proteins with recognition and cell adhesion function contain the Ig domain. The flexibility of the loops ensures considerable amino acid adaptability. This phenomenon ensures the accommodation of a substantial variety of structures and sequences without disruption of the overall structure.

the λ gene locus is slightly different, where a cluster of nearly 33 V gene segments precedes multiple paired J and C segments (about five pairs) (Owen, 2019; Schroeder & Cavacini, 2010). During B cell development, both heavy and light chain loci undergo DNA recombination events, known as V(D)J recombination, to randomly join their gene fragments to create a functional V region. After the *Ig* gene transcription, the mRNA splicing process connects the functional V region to the C region.

1.3.2.4 V(D)J recombination

V(D)J recombination is a lymphocyte-specific DNA rearrangement event during which V, D (only in the case of the heavy chain), and J segments are assembled to form the variable region of the antibody. The main enzyme complex of the lymphoid-specific recombination-activating gene 1 and 2 (RAG1/2) is responsible for V(D)J recombination. First, this enzyme complex recognizes the recombination signal sequences (RSS) that flank each V, D, and J segment. RSS consists of a conserved heptamer and nonamer fragments separated by 12 or 23 nucleotides. The recombination occurs only when the RAG1/2 endonuclease complex binds to two RSSs with different lengths (12/23 role). This role will ensure the attachment of V(D)J fragments in the correct order.[a] Following binding to RSSs, the RAG1/2 complex introduces double-strand breaks (DSBs) at the RSS sites. The repair of these breaks with the help of the non-homologous end-joining (NHEJ) DNA repair system results in the ligation of the two gene segments (*e.g.,* V_L and J_L; Figure 1-1) (Johnson et al., 2009; Malu et al., 2012; Musat et al., 2019; Owen, 2019; Roth, 2000; Roth, 2014; Schroeder & Cavacini, 2010; Seifert et al., 2019).

[a] The D_H fragment is flanked by the 12-base pair (bp) spacer while the V_H and J_H fragments contain 23-bp ones. The V_κ and J_λ contain 12-bp spacers whereas V_λ and J_κ have 23-bp ones.

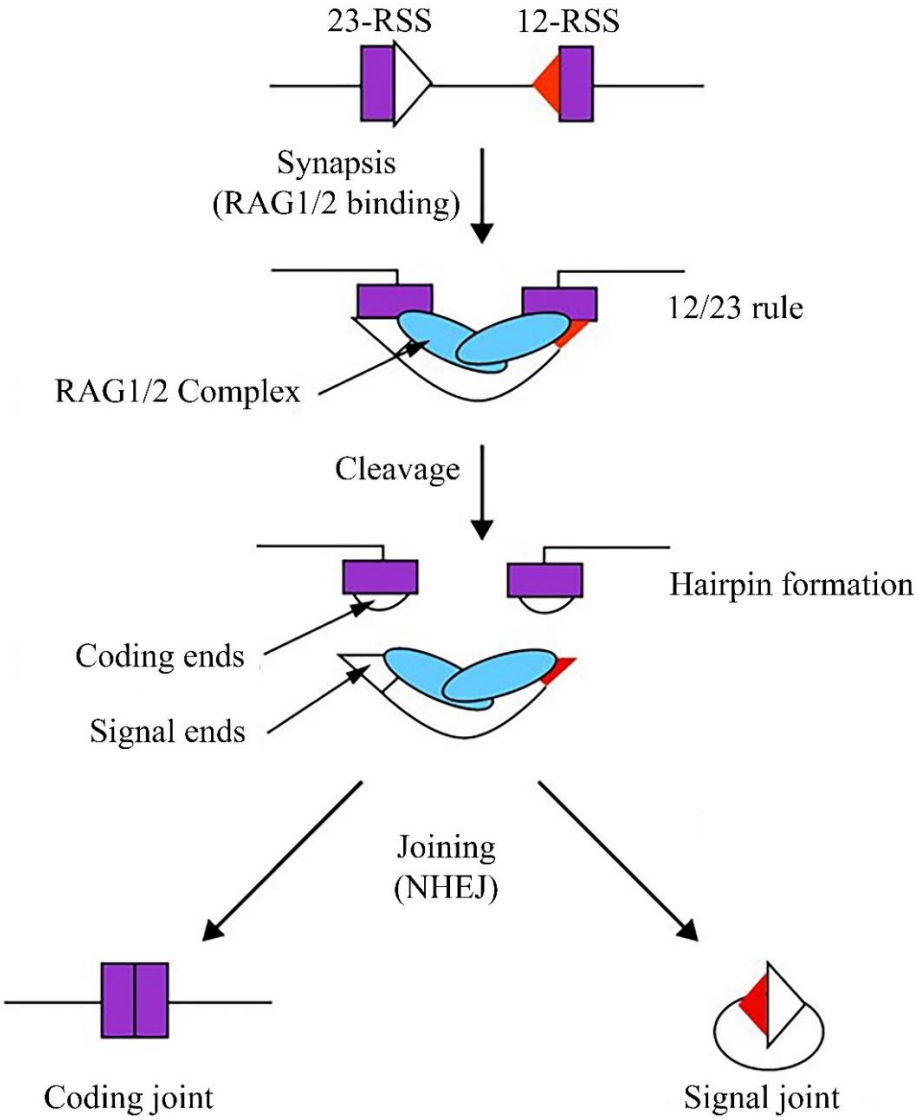


Figure 1-1: Schematic representation of V(D)J recombination. In the first step, RAG1/2 complex binds the RSS of different sizes (12/23 rule). Then, RAG1/2 complex introduces DSBs at the RSSs. In the last step, NHEJ repairs the DSBs, joining the coding and the signal ends together. Modified from Roth D. B., 2000. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited (https://www.biomedcentral.com/about/policies/reprints-and-permissions).

B cells begin with the joining of D and J_H fragments, followed by recombination between V_H and the D-J_H segment. The B cell then expresses its *IgH* gene in combination with the surrogate light chain to test the functionality of the newly recombined heavy chain (Malu et al., 2012; Owen, 2019; Winkler & Martensson, 2018).[a] If the heavy chain rearrangement is successful, the B cell initiates light chain recombination at *Igλ* or *Igκ* loci. Following the light chain recombination, the complete BCR is expressed[b] and examined for its reaction with self-antigens (Schroeder & Cavacini, 2010). The elimination of these self-reactive B cells contributes to the B cell central tolerance.[c] In humans, both λ and κ light chains contribute to the circulating B cell pool, where 60% of mature B cells carry a rearranged λ light chain (Owen, 2019).

1.3.2.5 Immunoglobulin isotypes

In humans and mice, IgM, IgD, IgG, IgE, and IgA are the five main classes of antibodies based on their constant regions of Cμ, Cδ, Cγ, Cε, and Cα, respectively.[d] The membrane-bound form of IgM (mIgM) is the first BCR expressed during B cell development (Aziz & Iheanacho, 2019). Fully matured B cells express high levels of mIgD and low levels of mIgM. However, antigen-stimulated mature B cells cease mIgD expression and secrete IgM during the primary antibody response, where it mainly

[a] This is the first checkpoint in B cell development. If the rearranged heavy chain-surrogate light chain complex generates a productive pre-BCR, the DNA rearrangement at the second *IgH* allele is permanently shut down, and the light chain recombination event is initiated.
[b] This is the second checkpoint in B cell development when the combination of the newly rearranged light chain with the previously rearranged heavy chain produces a functional BCR.
[c] This is the third checkpoint in B cell development.
[d] There are four subclasses of IgG: IgG1, IgG2, IgG3, and IgG4. Similarly, IgA is further divided into two subtypes: IgA1 and IgA2.

functions as an opsonin (coating agent) to enhance phagocytosis (Schroeder & Cavacini, 2010). IgG is the prevalent class of antibody in the serum during the secondary humoral response (Aziz & Iheanacho, 2019). Generally, IgGs mediate complement fixation, toxin neutralization, and bacterial opsonization (Aziz & Iheanacho, 2019; Schroeder & Cavacini, 2010). Different subclasses of IgG function differently; therefore, they are produced in response to different antigens.[a] Protein antigens stimulate IgG1 and IgG3 production while polysaccharide antigens trigger class switching to IgG2 and IgG4 (Schroeder & Cavacini, 2010).

IgA is the main antibody at the mucosal surfaces and in secretions where it either directly neutralizes toxins, viruses, and bacteria, or prevents their binding to the body surfaces (Aleyd et al., 2015; Aleyd et al., 2014; Heineke & van Egmond, 2017).[b] IgA1 structurally, and consequently functionally, differs from IgA2. IgA2 is more resistant to bacterial proteases and dominates the mucosal secretions, while IgA1 is mainly present in the serum (Schroeder & Cavacini, 2010). Since IgE is a very potent antibody due to its ability to activate granulocytes and Langerhans cells (Schroeder & Cavacini, 2010), it has the lowest serum concentration.[c] This antibody is involved in allergic reactions and protection against worm infections (Schroeder & Cavacini, 2010). Attributable to the abovementioned functional differences amongst classes of antibodies, the host immune

[a] Various IgG subclasses differ regarding serum levels, flexibility, functional affinity, and ability to fix complement.
[b] At the mucosal barriers, neutrophils expressing IgA receptors can clear IgA-coated pathogens through inducing proinflammatory functions, such as ADCC, degranulation, production of reactive oxygen species, release of NETs, and cytokine and chemokine secretion.
[c] These cells express the high-affinity IgE receptor, also known as FcεRI. Moreover, circulating IgE upregulates the expression of this receptor. Thus, IgE is a very potent antibody.

system must activate the expression of the proper antibody isotype during immune responses.

1.3.2.6 B cell activation

Except for the nABs, all B cell responses require the presence of antigens. These B cell responses are triggered by B cell activation upon antigen contact and result in antibody secretion. The B cell activation is achieved with or without the T_H cell assistance, referred to as T cell-dependent (TD) and T cell-independent (TI) B cell activation, respectively. Generally, B-2 cell activation occurs through TD pathway in response to protein antigens, while the TI pathway mostly activates non-conventional B cells in response to multivalent or highly polymerized antigens (Eibel et al., 2014; Owen, 2019; Pieper et al., 2013). T_H participation in B cell activation improves B cell proliferation, enhances memory cell development, and induces the secondary antibody diversification process (Eibel et al., 2014; Pieper et al., 2013). Consequently, B-2 cells are the primary source of the high-affinity antigen-specific humoral responses while other B cells constitute the early innate-like, low-affinity polyreactive antibody responses.

1.3.2.6.1 T cell-independent B cell activation

In the early stages of an infection, a rapid low-affinity IgM response is mounted by non-conventional B cells (*i.e.*, B-1 and MZ B cells) towards antigens that contain polyvalent or repeating determinants and are shared among microbial groups (Owen, 2019; Pieper et al., 2013). The antigen neutralization and opsonization with these early antibodies not only inhibit initial pathogen replication but also boost the ensuing antigen-specific B-2 cell response by enhancing follicular DCs (FDCs) antigen retention (Baumgarth, 2011;

Kranich & Krautler, 2016).[a] The antigens capable of eliciting TI-B cell responses are referred to as TI antigens. The TI antigens are further divided into two groups: TI-1 and TI-2 antigens (Owen, 2019).

TI-1 antigens, such as LPS, are multivalent and mitogenic to all B cells including immature and B-2 cells (Mond et al., 1995; Owen, 2019). These antigens bind to the innate immune receptors (*e.g.*, PRRs) on all B cells. High levels of TI-1 antigens cause receptor cross-linking, leading to B cell activation and subsequent antigen secretion (Owen, 2019). Since B cell stimulation through TI-1 antigens occurs through innate immune receptors and independent of BCR specificity, high levels of TI-1 antigens mount polyclonal antibody responses. However, at a lower level, B cell activation occurs when the antigen binds PRR and BCR simultaneously, resulting in PRR/BCR cross-linking and subsequent B cell activation (Coutinho et al., 1974; Owen, 2019). In this case, only B cells bearing BCR capable of detecting the antigen are induced. Therefore, the TI-1 antibody reaction in response to lower levels of antigen is oligoclonal.

TI-2 antigens, such as bacterial capsular polysaccharides and flagellin, are highly polymerized and mainly stimulate B cells through their BCR. The repetitive nature of these antigens facilitates BCR cross-linking and subsequent B cell activation. Opsonized TI-2 antigens by complement system (*i.e.*, C3d and C3dg fragments) also interact with complement receptor 2 (CR2, also known as CD21) on B cells. MZ B cells express high levels of CD21 on their surface and constitute the main TI-2 responding B cell subtype

[a] FDCs, located in primary and secondary B cell follicles, play an important role in retaining and presenting native antigens to the B cells by capturing the antibody-antigen complexes through their Fc receptors.

(Mond et al., 1995; Owen, 2019; Zandvoort & Timens, 2002). Full TI-2-mediated B cell activation requires help from other cells, such as monocytes, neutrophils, MQs, and DCs (Hendricks et al., 2018). These cells secrete BAFF, which induces B cell survival, maturation, and antibody secretion (Dickinson et al., 2015; Owen, 2019).

Humoral responses towards TI antigens mainly consist of low-affinity IgM. However, TI-activated B cells may also undergo a limited degree of secondary antibody diversification process and form long-lasting memory cells (Haas, 2015; Zandvoort & Timens, 2002). For example, it was shown that neutrophil assistance resulted in production of higher affinity IgG or IgA responses by TI-2-activated MZ B cells (Hendricks et al., 2018).[a]

1.3.2.6.2 T cell-dependent B cell activation

The TD B cell responses constitute the adaptive humoral immunity, which is mediated exclusively by B-2 cells (Owen, 2019). When the B-2 cells complete their development in the spleen, they join the bloodstream and recirculate between blood and lymphoid organs (Owen, 2019; Pieper et al., 2013). These naïve cells enter the B cell follicles of the secondary lymphoid organs in search of their cognate antigen (Owen, 2019; Pieper et al., 2013).[b] Inside these follicles, naïve B-2 cells browse the antigen pool using

[a] These neutrophils are referred to as B cell helper neutrophils (N_{BH}), which colonize the splenic MZ and differ from circulating neutrophils.

[b] MQs, DCs, and B cells can pick up the soluble antigens from the subcapsular sinus (SCS) region of the lymph node by extending their process into the system of conduit emanating from the SCS. Antigen-transporting cells, such as MQs, DCs, and non-cognate B cells, can carry antigen-immune complexes (antigens associated with complement or antibodies) to the B cell follicles through their complement or Fc receptors.

their BCR (Yam-Puc et al., 2018).[a] Antigen-activated B-2 cells briefly spread their membrane over the antigen bearing cell to obtain most antigens. The ensuing B-2 cell membrane contraction results in BCR clustering, also known as antigen-induced BCR oligomerization, during which the BCR complex moves transiently into the lipid rafts (Harwood & Batista, 2010; Owen, 2019; Varshney et al., 2016).[b] Through these changes in the membrane, the B-2 cell forms an immunological synapse with the antigen presenting cell (Harwood & Batista, 2010).

In addition to BCR, coreceptors, cytokine, and BAFF receptors are also involved in the immunological synapse (Pieper et al., 2013). Within B-2 cells, the formation of this synapse stimulates BCR-mediated signaling, antigen uptake, and antigen presentation on MHC II (Owen, 2019).[c] The BCR-mediated signaling induces two vital changes in the B-2 cell. First, it alters the B-2 cell chemokine receptor profile causing the B-2 cell migration to the boundary of the B- and T-cell zones (B-T zone) within the follicle.[d] Second, it upregulates the B-2 cell surface costimulatory CD80, CD86, and CD40 molecules, which significantly enhance the subsequent interaction between the B-2 cell and the cognate T_H cell (Owen, 2019; Seifert et al., 2019; Yam-Puc et al., 2018). Hence, the instigated

---

[a] Inside follicles, FDCs are the main antigen presenting cell type. In the spleen, MZ B cells participate in antigen presentation.

[b] Lipid rafts are small, heterogeneous, dynamic, and highly ordered domains in the cell membrane that are enriched in cholesterol and sphingolipids. Since lipid rafts incorporate receptors and signaling proteins, they act as a signal transduction platform.

[c] B-2 cells acquire the cognate antigen in two different ways. 1) B-2 cells cleave the antigen from the antigen bearing cell by directly releasing lysosomal proteases into the synaptic junction. The cleaved antigen is either directly loaded onto the MHC II molecule or internalized with BCR into the endosomal pathway and enters the subsequent exogenous antigen presentation route. 2) If the BCR affinity for the cognate antigen is significantly high, the B-2 cell tugs the antigen from the antigen bearing cell and internalizes it with BCR.

[d] The chemokine receptors include CCR7, CXCR5, and Epstein-Barr virus-induced receptor 2 (EBI2), which interact with CCL19 and 21, CXCL13, and 7α,25-dihydroxycholesterol (7α,25-OHC), respectively.

signaling pathways through the immunological synapse provide the first signal in B cell activation.[a]

In the B-T zone, the antigen-stimulated B-2 cells engage with their conjugate T_H cells through their peptide-MHC II complex and coreceptors (*i.e.*, CD40, CD80, and CD86) (Mesin et al., 2016).[b] These interactions form a synapse between the B cell and the cognate T_H cell, which constitutes the second signal required for TD B cell activation. The formation of this synapse stimulates the T_H cell to secrete cytokines such as IL-4 and IL21 into the synaptic cleft to help the B cell differentiation to proceed (Mesin et al., 2016; Owen, 2019). These cytokines provide the third signal for B cell activation. The B cell, in return, increases its surface expression of receptors for these cytokines (Mesin et al., 2016; Owen, 2019).

Following this stage, the activated B cells remodel their surface chemokine receptor profile and follow one of the two paths (Gars et al., 2019; Owen, 2019; Yam-Puc et al., 2018). B cells that undergo a strong initial interaction with the antigen enter the splenic red pulp or the lymph node medullary cords. These cells form primary foci and differentiate into plasmablasts that secrete unmutated, mainly IgM antibodies, and form the early TD humoral responses (Gars et al., 2019; Yam-Puc et al., 2018).[c] These low-affinity antibodies are efficient opsonins but cannot effectively neutralize pathogens and toxins (Zhang et al.,

---

[a] Note that several factors strongly impact the outcome of BCR-antigen engagement: 1) the maturation status of the activated B cell, 2) the magnitude and duration of immunological synapses, and 3) the involvement of coreceptors (such as CD21 and CD40), cytokine receptors (such as IL-4-R and IL-21R), and survival factor receptors (such as BAFF-R).

[b] This engagement may last for a few minutes to several hours.

[c] During the early stages of primary humoral responses, some memory cells are also generated that express unmutated IgM.

2016). B cells with higher affinity BCR enjoy a longer B-T interaction and are more likely to become pre-GC B cells (Yam-Puc et al., 2018).[a] These cells return to the interior regions of the B cell follicle, where they contribute to GC formation and differentiate into GC B cells (Owen, 2019; Yam-Puc et al., 2018). These B cells undergo the secondary antibody diversification process and are responsible for the production of high affinity antibodies (mainly IgG) later in the immune response and generation of memory B cells (Gars et al., 2019; Owen, 2019; Pieper et al., 2013; Yam-Puc et al., 2018).

1.4 Diversification of the antibody repertoire

The diversification of the antibody repertoire is a vital step in the arms race between the host's antibody response and pathogens. This diversification happens in two steps. The primary diversification, which gives rise to the naïve BCR repertoire, occurs during the B cell development. The naïve BCR repertoire is responsible for detecting antigens upon first exposure. The secondary antibody diversification is initiated when the B cell binds the cognate antigen (Maul & Gearhart, 2010; Owen, 2019). As a result of the secondary antibody diversification, the activated B cells, expressing low-affinity IgM, give rise to the B cells secreting high affinity IgG for the cognate antigen (Meffre et al., 2001). Thus, the primary and secondary antibody diversifications are essential in the initial recognition of an antigen by naïve B cells and the efficient neutralization of the cognate antigen by the activated B cells, respectively.

[a] The signaling through B cell CD40 and T_{FH} cell CD40L is crucial for GC formation.

1.4.1 Primary antibody diversification

The main primary antibody diversification occurs during V(D)J recombination of *Ig* genes (Briney & Crowe, 2013; Maul & Gearhart, 2010; Owen, 2019).[a] Several mechanisms contribute to this diversification. First, there are multiple gene segments in the V, D, and J clusters from which novel random combinations are selected to create the variable coding sequences. Second, since both V_H and V_L regions participate in the formation of antigen-binding domain, different combinations of heavy and light chain pairs provide further BCR diversity. Third, the enzyme activity of Artemis, terminal deoxynucleotidyl transferase (TdT),[b] and exonuclease increase junctional diversity by palindromic (p) and non-templated (n) nucleotide addition or deletion, respectively (Johnson et al., 2009; Malu et al., 2012; Owen, 2019; Patel et al., 2018; Schroeder & Cavacini, 2010; Thomson et al., 2020). The primary antibody diversification can create up to 5×10^{13} unique BCRs in humans and mice (Granato et al., 2015; Malu et al., 2012; Pieper et al., 2013).

Immunoglobulin gene conversion (IGC) is another approach to expand the naïve B cell repertoire in some avian and mammalian species, such as chicken, turkey, cattle, and rabbit (Haakenson et al., 2018; Lundqvist et al., 2006; Tang & Martin, 2007; Walther et al., 2016). Their *Ig* loci contain one or a very limited number of functional V, D, and J segments, hence the V(D)J rearrangement leads to a limited primary antibody repertoire

[a] In humans, while V(D)J recombination is the main mechanism of primary antibody diversification, a less frequent mechanism of V(DD)J recombination (*i.e.*, D-D fusion) may also contribute to the diversification of naïve BCR repertoire.
[b] TdT is also known as DNA nucleotidylexotransferase (DNTT).

(Leighton et al., 2018). These species mainly utilize the IGC as a mechanism to heighten the versatility of their primary antibody repertoire during B cell development (Bastianello & Arakawa, 2017; Leighton et al., 2018).[a] The first step in IGC involves introducing mutations, causing single-strand breaks (SSBs) in the pre-rearranged V fragment, which are then resolved by the homologous recombination (HR) system. The HR uses the upstream pseudogenes (ψ) as a template and incorporate their sequence into the functional, pre-rearranged V segment, which increases the diversity of the primary antibody repertoire (Frieder et al., 2006; Leighton et al., 2018).[b]

1.4.2 Secondary antibody diversification

After exposure to an antigen, the subsets of naïve B cells bearing a BCR that recognizes the antigen become activated. In the GC, the antigen-activated B cells undergo secondary antibody diversification (Maul & Gearhart, 2010; Owen, 2019). In humans and mice, the secondary antibody diversification includes two processes: antibody affinity maturation (AM) and isotype switching (IS). At the molecular level, AM and IS are achieved through somatic hypermutation (SHM) and class switch recombination (CSR), respectively (Briney & Crowe, 2013; Chi et al., 2020; Maul & Gearhart, 2010).[c] AM increases the affinity of the antibody for the cognate antigen, whilst IS changes the class of antibody from IgM into other isotypes (*i.e.,* IgG, IgA, or IgE).

---

[a] It is suggested that IGC may have initially evolved in the common ancestor of mammals and birds and was later lost in the evolutionary branches leading to humans and mice.
[b] The ψ genes do not possess any promoter or RSS and usually contains 5′ or 3′ stop codons.
[c] In humans, beside SHM, less frequent mechanisms of SHM-associated insertions and deletions, and affinity maturation and antigen contact by non-CDR regions of the antibody also contribute to affinity maturation of activated B cells.

Affinity maturation is achieved by introducing point mutations in the rearranged V(D)J fragment, particularly in regions directly contacting the antigen (*e.g.*, CDRs). The activated B cells that express mutated antibodies undergo clonal selection, resulting in a gradual rise in the affinity of the antibodies in the course of an immune response (Maul & Gearhart, 2010; Owen, 2019). During CSR, double-stranded breaks in the intronic regions, known as switch (S) regions, initiate an NHEJ event resulting in the replacement of $C_H\mu$ with other C_H isotypes that changes the antibody's function (Maul & Gearhart, 2010; Owen, 2019). S regions, which are highly repetitive and GC-rich, flank the heavy chain constant (C_H) genes and are considered as CSR sites (Yu & Lieber, 2019).

Although AM is mainly restricted to TD-activated GC B cells and happens later in the immune response, CSR occurs early before the GC formation and can also be stimulated by TI antigens with the help of other immune cells, such as APCs and neutrophils (Owen, 2019). Nevertheless, the outcome of conventional secondary antibody diversification is the production of different isotypes of antibodies with a higher affinity (as much as a 1000-fold increase) for the cognate antigen (Magor, 2015).

1.4.3 Cellular basis of antibody affinity maturation

The structure of the GC is an ideal place for the secondary antibody diversification process. GC consists of two histologically and functionally distinct regions. The rapidly dividing activated B cells (known as centroblasts) establish the dark zone (DZ), where they undergo SHM (~10^{-4} to 10^{-3} per base per generation compared to 10^{-9} genomic basal mutation frequency) in their *Ig* V region genes (mainly CDR3) (Gars et al., 2019; Melchers, 2015; Qiao et al., 2017; Wong & Germain, 2018). The GC light zone (LZ) is enriched with

FDCs and contains a limited but crucial pool of the cognate antigen-activated T_{FH} cells (Gars et al., 2019; Melchers, 2015). The clonal selection of the B cells expressing mutated antibodies with a higher affinity for the cognate antigen occurs within the LZ (Melchers, 2015).[a] The LZ B cells are referred to as centrocytes. The activated B cells constantly change their surface chemokine receptor profile to move back and forth between DZ and LZ, a model known as cyclic re-entry (Mesin et al., 2016; Yam-Puc et al., 2018). The affinity of the antibody pool for the cognate antigen gradually rises as a result of the iterative processes of SHM, proliferation, and clonal selection.

Following the introduction of SHM in the DZ, B cells enter the LZ where they go through an elegant selection process. In the LZ, FDCs present naïve antigens to the GC B cells primarily in the form of immune complexes (*i.e.*, antigen-antibody or antigen-complement complexes). When the B cells re-enter the LZ, they browse FDCs to test their mutated BCR. B cells are required to uptake antigen and present it on their MHC II molecules to receive "help" from T_{FH} cells (Maul & Gearhart, 2010; Mesin et al., 2016; Owen, 2019). The B cells that successfully received "help" from T_{FH} cells are programmed to suppress the MHC II expression and re-enter the DZ for further SHM.[b]

The affinity of the antibody in the immune complex is an indirect measurement for the affinity of the newly mutated BCR (Mesin et al., 2016). For the B cells to acquire enough antigen to present to the T_{FH}, the affinity of their BCR must be high enough to

---

[a] GC B ells are highly prone to apoptosis unless they receive survival signals from their environment. In the LZ, B cells compete to receive survival signals from T_{FH} cells. The limited number of T_{FH} cells guarantees that only B cells bearing BCR with the highest affinity for the cognate antigen receive survival signals.

[b] Only 10-30% of the B cells successfully get permission to re-enter the DZ. These cells are required to halt their MHC II expression to avoid carrying over any antigen to the next round of selection in the LZ.

break the interaction between the antigen and the antibody-complement in the immune complex. The cells that succeed in stripping the antigens from FDCs express the peptide-MHC II complexes on their surface and subsequently receive survival signals from T_{FH} (Maul & Gearhart, 2010; Mesin et al., 2016; Owen, 2019). The B cells with mutated BCR that can no longer bind the cognate antigen are eliminated by apoptosis. The higher the affinity of mutated BCR for the cognate antigen, the higher the densities of peptide-MHC II on the B cell. This increase in peptide-MHC II surface expression greatly improves the chances and the length of B cell interaction with the limited number of T_{FH} (Mesin et al., 2016; Owen, 2019). Therefore, the B cells expressing BCR with a higher affinity for the cognate antigen will outcompete the lower affinity B cells.

While some of the B cells return to the LZ for more rounds of mutations and selection, some B cells expressing high affinity BCR differentiate into plasma cells and begin to secrete antibodies. These secreted antibodies replace the old antibodies in the immune complexes of the FDCs. This replacement increases the selection threshold of the newly mutated antibodies by decreasing the B cell accessibility to the antigens. Subsequently, the overall affinity of serum antibodies gradually rises (Mesin et al., 2016; Wong & Germain, 2018).[a] It should be emphasized that some of the high affinity GC B cells differentiate into memory cells, which drive the faster and more robust humoral

[a] A chronic GC response could result in the generation of the potent broadly-neutralizing antibodies (bNABs) in response to viral infections, such as influenza and human immunodeficiency virus (HIV). Intriguingly, insertions and deletions are common aspects of these antibodies and they accumulate high levels of mutations in their CDRs as well as the framework regions (30-40 and >100 mutations in bNABs against influenza and HIV, respectively).

immunity upon re-exposure to the same antigen (*i.e.*, the secondary antibody response) (Good-Jacobson, 2018; Kuraoka et al., 2009; Owen, 2019).

1.5 Activation-induced cytidine deaminase and antibody diversification

Activation-induced cytidine deaminase (AID) is the enzyme responsible for initiation of secondary antibody diversification process (Muramatsu et al., 2000; Revy et al., 2000). In the first step of IGC, SHM, and CSR, AID introduces a high frequency of mutations in the *Ig* genes. AID converts deoxycytidine (dC) into deoxyuridine (dU) on single-stranded DNA (ssDNA) in any sequence, with a several fold (2-6 fold) preference for deaminating dC in the context of WRC (W=A/T; R=A/G) motifs, known as AID "hotspots" (Bransteitter et al., 2003; Bransteitter et al., 2006; Frieder et al., 2006; Kolar et al., 2007; Larijani, Frieder, Basit, et al., 2005; Meffre et al., 2001; Muramatsu et al., 2000; Muto et al., 2000; Nagaoka et al., 2002). However, the recent crystal structure of AID revealed that in the S regions, the G-quadruplex (G4) substrates might override the AID specificity for WRC motifs (Qiao et al., 2017).[a] Nevertheless, if the DNA replication occurs before DNA repair or in the absence of the dU:dG mismatch sensors (see the fallowing paragraphs), the dU:dG mismatches cause G:C→T:A transversion mutations. Otherwise, the dU:dG mismatches recruit DNA repair systems where either base excision repair (BER) or mismatch repair (MMR) attempts to repair the DNA. However, a unique feature of SHM is the recruitment of the non-canonical (*i.e.*, error-prone) version of BER and MMR which results in introducing more mutations other than G:C→T:A transversion

[a] In the mammalian S regions, the abundant tandem G repeats interspersed by AGCT (AID hotspot) develop G4 structure on the non-template strand, which stabilizes the R-loops formed during transcription.

and creating double-stranded breaks (DSBs) required for CSR (Figure 1-2) (Chi et al., 2020; Di Noia & Neuberger, 2007; Maul & Gearhart, 2010; Saribasak et al., 2012; Wilson et al., 2005).

Specifically, when the BER is recruited, the enzyme uracil-N-glycosylase (UNG) acts as the dU:dG mismatch sensor and removes the dU generating an abasic site, which is successively cleaved by the apurinic/apyrimidinic endonuclease I (APE I). This nick is then processed and filled with error-prone DNA polymerases that introduce more mutations. Either this abasic site can serve as a non-informative template for DNA synthesis or initiate a version of long-path BER which generates a DNA gap that is filled with error-prone DNA polymerases (Methot & Di Noia, 2017). The specialized DNA polymerase REV1 can bypass the abasic site by inserting a dCMP across it, causing transversion mutations at C:G pairs, and the error-prone DNA Polη (and to lesser extent DNA Polζ, Polκ, and Polι) can fill out the generated gap, introducing mutations at A:T pairs (Di Noia & Neuberger, 2007; Faili et al., 2009; Maul & Gearhart, 2010; Maul et al., 2016; Methot & Di Noia, 2017; Saribasak et al., 2012; Wilson et al., 2005; Zanotti & Gearhart, 2016). In the case of MMR, Msh2 and Msh6 enzymes form the MutSα complex, which acts as the dU: dG mismatch sensor. An endonuclease then cleaves the dU-containing strand at 5′ end, creating the necessary nick for the 5′→3′ exonuclease enzyme to remove the damaged base and its surrounding bases. Similar to the BER pathway, an error-prone DNA polymerase (such as DNA Polη) then fills the gap and introduces more mutations (Figure 1-2) (Methot & Di Noia, 2017).

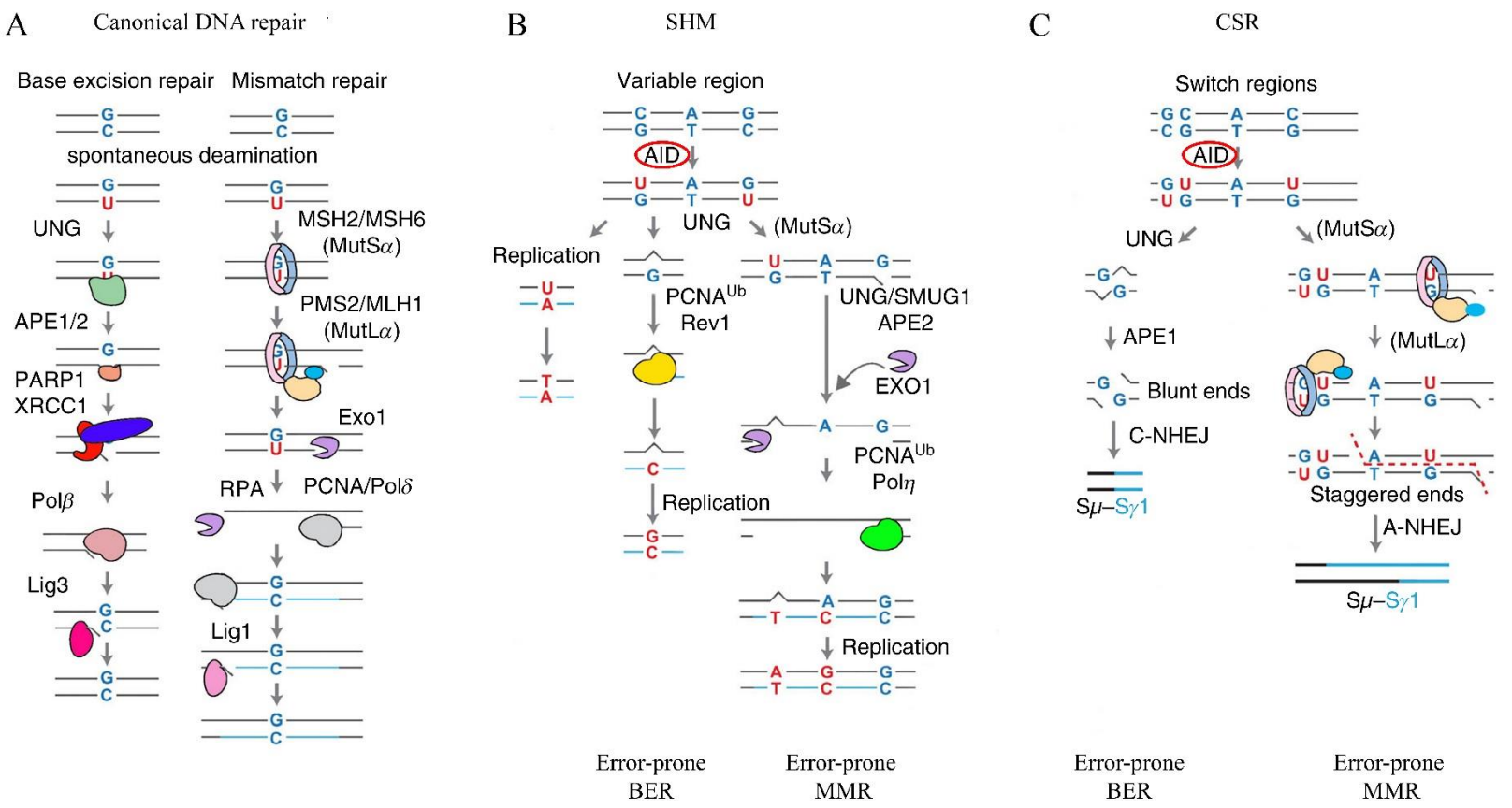


Figure 1-2: Overview of A) the canonical Base excision repair (BER) and mismatch repair (MMR), B) the error-prone BER and MMR during SHM, and C) the error-prone BER and MMR involved in CSR. For more details, refer to the text. Adapted from Chi X, et al., 2020 with permission.

At the molecular level, the transcriptomics of centroblasts differs significantly from that of centrocytes. In the DZ, the highly proliferative GC B cells express high levels of AID and the error-prone DNA polymerase eta (DNA polη), indicative of SHM occurrence (Mesin et al., 2016). Remarkably, the presence of dU in the V and S regions of *Ig* genes were detected as early as 24 hours after B cell stimulation (Maul et al., 2011). However, when the B cells enter the LZ, they suppress AID expression and display an activated phenotype characterized by the expression of activation markers, such as CD86, and genes involved in CD40 and BCR signaling pathways (Mesin et al., 2016).[a]

The introduction of the AID-mediated mutations also initiates CSR to change the effector function of antibodies.[b] Preceding each C_H exon (except for C_δ), there is a G-rich, repetitive stretch of DNA known as the switch (S) region (Owen, 2019; Schroeder & Cavacini, 2010; Wong & Germain, 2018). The S region is comprised of repetitive tandem sequences containing the AID hotspot, where mutations occur in close proximity.[c] Due to the proximity of these mutations, the BER or MMR attempts to resolve the lesions lead to the formation of DSBs in the donor and acceptor S regions (Maul & Gearhart, 2010; Wong & Germain, 2018). These DSBs are then resolved using classical or alternative NHEJ pathways, joining the two broken S regions through a loop-out deletion, which results in changing the effector function of the antibody. Local cytokines dictate the new antibody

[a] The activated phenotype is a result of two signals. The first signal is received through BCR when it binds the antigen, while the second signal is delivered by interacting with the cognate T_{FH}.
[b] Switching between IgM and IgD is the result of an alternative mRNA splicing event. However, switching to other antibody isotypes requires an irreversible DNA recombination event (*e.g.*, CSR).
[c] Sμ contains the highest repetitive number of AID hotspots (5′-AGCT-3′) amongst all S regions; therefore, Sμ is the most common target of AID.

isotypes synthesized by initiating the transcription of the donor and acceptor C regions, making the DNA accessible to AID (Owen, 2019; Pieper et al., 2013; Wong & Germain, 2018).[a]

The absolute requirement of AID for secondary antibody diversification is apparent in the case of hyper IgM syndrome type II (HIGM II) patients. In these patients, a deficiency in the AID gene leads to the abolition of AID-mediated mutations, and consequently, the lack of AM and IS (Minegishi et al., 2000; Revy et al., 2000). Therefore, AID is essential in generating a robust humoral immune response by increasing affinity and diversifying the functional specificities of antibodies.

AID is also involved in diversifying the primary antibody repertoire (*i.e.*, before the antigen contact) by introducing somatic mutations in pre-rearranged V regions. The first evidence of these mutations was observed in chickens where the deletion of the ψV genes, or disruption of genes involved in recombination repair pathway, caused a shift from IGC to AID-mediated somatic mutations at G/C bases (Buerstedde & Arakawa, 2006). Studies on cattle also revealed the contribution of AID-mediated somatic mutations in the formation of the primary antibody repertoire in these species (Haakenson et al., 2018; Zhao et al., 2006).[b] It has been suggested that the limited germline-encoded combinatorial

[a] IL-4 promotes switching to IgG1 and IgE, while IL-5 enhances IgA production. The presence of TGF-β stimulates IgA or IgG2b recombination, while IFN-γ triggers IgG2a and IgG3 production.

[b] Ten percent of cattle antibodies have a unique ultralong CDR3 loop, which form a "stalk and knob" structure and is responsible for antigen recognition. In humans, a typical CDR3 is 8-16 amino acid long, while the cow's ultralong CDR3 is 40 to 70 amino acids in length. Ig_HD8-2 gene segments encode the CDR3 of the bovine ultralong antibodies. An interesting feature of the ultralong antibodies is their structural diversity due to disulfide bonds. There are existing and potential cysteine codons in the Ig_HD8-2, which can form disulfide bonds within the CDR3. In Ig_HD8-2, 30 of the codons that can be converted to cysteine with a single nucleotide mutation (*i.e.*, potential cysteine codons) overlap with 19 AID hotspots. Thus, AID significantly contributes to structural diversification of the bovine ultralong antibodies.

diversity observed in the *Ig* loci of some species, such as sheep, prairie vole (*Microtus ochrogaster*), and the guinea pig (*Cavia porcellus*), might be an indication of IGC and/or AID-mediated somatic mutation involvement in the production of the initial B cell repertoire in these species (Guo et al., 2012; Qin, Liu, et al., 2015; Qin, Zhao, et al., 2015).

Interestingly, AID expression and some levels of somatic mutations were detected in the immature T1 B cells in patients with hyper IgM syndrome type I (HIGM I) (Kuraoka et al., 2009). A deficiency in CD40 ligand (CD40L; also known as CD154), typically found on the activated T cells, causes HIGM I, which is characterized by normal to elevated levels of serum IgM, lack or minimum levels of IgG, and the absence of GC, SHM, CSR, and memory cells (Hirbod-Mobarakeh et al., 2014). Unlike the conventional SHM that happens in GC, the observed mutations in HIGM I patients showed no evidence of antigen-driven selection.[a] This mechanism of antibody diversification in HIGM I patients is reminiscent of the primary antibody diversification in chicken and cattle (Buerstedde & Arakawa, 2006; Haakenson et al., 2018; Zhao et al., 2006). These findings lead to the hypothesis that most or even all vertebrates might share the observed AID expression and SHM during B cell development (Kuraoka et al., 2009).

1.5.1 AID structure

AID is a small, positively charged, globular protein displaying high binding affinity (~nM-range) for its negatively charged ss-DNA substrate (Larijani et al., 2007). Despite

[a] Due to the lack of antigen-driven selection, mutations observed in HIGM I patients revealed a low replacement/silent mutation ratio, were widely dispersed within V regions, and were found in antibodies with different V_H regions and potentially different specificities.

the extensive effort in the past two decades, AID's molecular structure is not fully understood yet. Due to highly charged surface, extensive non-specific protein-protein/DNA/RNA interactions, polydisperse oligomerization, and genotoxicity for the host cell, elucidation of native AID structure by means of X-ray crystallography and nuclear magnetic resonance (NMR) has been a challenge (King & Larijani, 2017; Pham et al., 2016). Hence, to enhance the solubility and/or crystallization of AID, substantial alterations including mutations, deletions, and truncations have been made to the only two available X-ray structures of AID (Pham et al., 2016; Qiao et al., 2017). Figure 1-3 is a representative computational model of human AID (Hs-AID). This model was generated through homology modeling of available partial X-ray structures of AID.

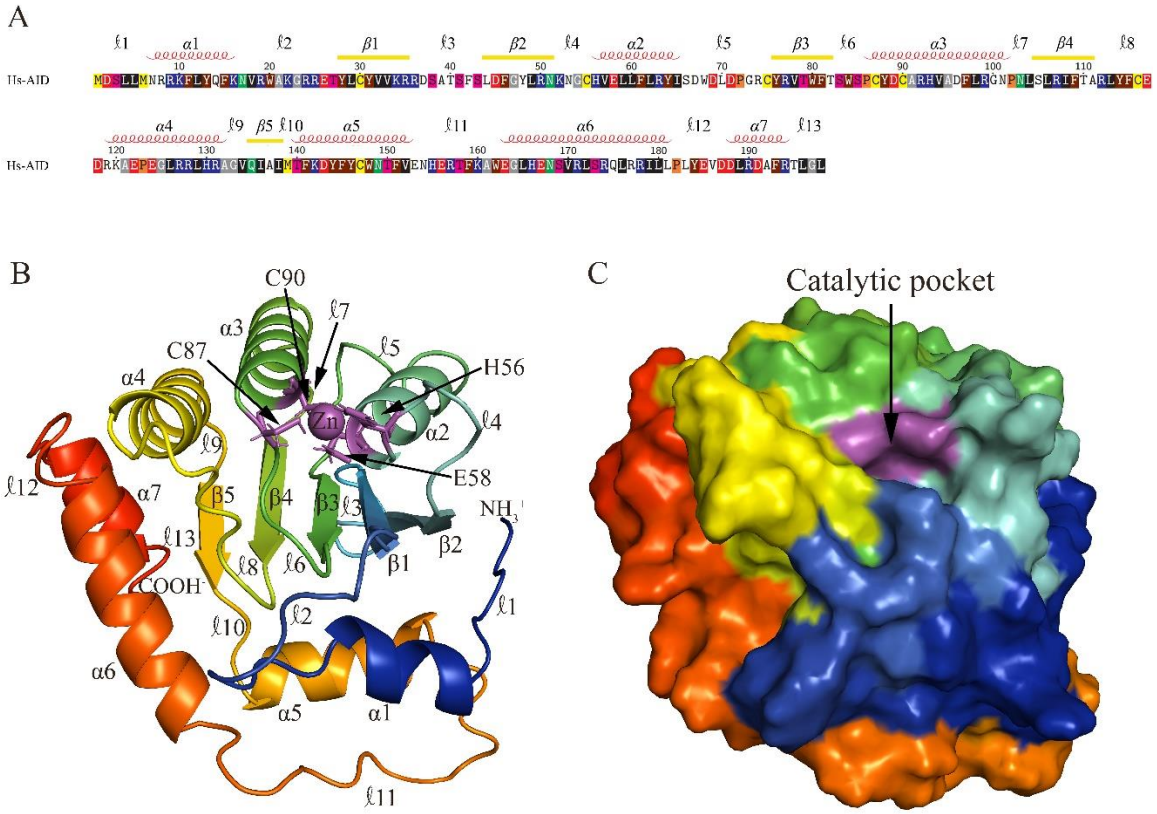


Figure 1-3: General structural features of human AID (Hs-AID). A) Sequence of Hs-AID. The approximate secondary structure of α-helical (α), β-strand (β), and loop (ℓ) regions are shown. Residues are colored according to chemical properties of the side chain. B) Representative ribbon model of predicated Hs-AID structure. In the model, blue to red color change indicates N to C terminus progression. The catalytic residues and zinc ion are shown in purple. Loops, β-strands, and α-helices are labeled. C) Predicted surface topology of Hs-AID. Catalytic pocket is shown in purple.

Before the availability of the X-ray structure of AID, homology modeling of wildtype AID based on the solved structures of the related APOBECs by X-ray or NMR revealed important aspects of AID structure-function relationship including AID's nuclear localization signal, substrate specificity loop, surface charge/topology, DNA binding grooves, secondary catalytic residues, and catalytic pocket dynamics (*i.e.,* Schrodinger's CATalytic pocket) (Abdouni et al., 2013; Carpenter et al., 2010; Dancyger et al., 2012; King & Larijani, 2017; King et al., 2015; Kohli et al., 2009; Larijani & Martin, 2012; Patenaude et al., 2009; Prochnow et al., 2007). The hallmark of this approach was the birth of the "computational-evolutionary-biochemical" method in which the computational models and biochemical analyses of various AID homologs were compared. Through this novel approach, AID's DNA binding groove 1 and 2, secondary catalytic restudies, and catalytic pocket opening and closure dynamic were discovered (King & Larijani, 2017; King et al., 2015), most of which were later confirmed through X-ray or NMR structure of AID and other APOBECs (Hou et al., 2019; Qiao et al., 2017; Shi et al., 2017). These findings are further described in the following sections.

In 2017, the most native X-ray crystal structure of AID was published (PDB: 5W1C, 5W0U, 5W0R, and 5W0Z) which included 10 point mutations and N- and C-terminal truncations (AID.crystal: Hs-AID$^{\Delta 5\text{-}N7D\text{-}R8P\text{-}R9A\text{-}K10T\text{-}L12T\text{-}F42E\text{-}H130A\text{-}R131E\text{-}F141Y\text{-}Y145E\text{-}181\Delta}$) (Qiao et al., 2017). This "near native X-ray crystal structure" of AID confirmed the presence of DNA binding groove 1 and also elucidated some aspects of its substrate specificity, such as its preference for the G4 structure and its lack of activity on dU and RNA. AID's preference for the G4 structure is due to its bifurcated substrate-binding

surface where the negatively charged amino acids of loop 8 (ℓ8) wedges the two positively charged substrate channel and assistant patch (*i.e.*, α6) (Qiao et al., 2017). Based on this model, one ssDNA overhang interacts with the substrate channel (*i.e.*, DNA binding groove 1) and the active site while the other one binds the assistant patch and improves the binding affinity (Qiao et al., 2017). Thus, the disruption of the assistant patch compromises the CSR without significant impact on SHM.

The AID crystal structure revealed that H56, W84, and Y114 hold cytidine in place while S85 and T27 hydrogen bond with the atom N4 and O2 of cytidine, respectively. The replacement of cytidine N4 with an O4 in uracil interrupts the formation of the stabilizing hydrogen bond with S85 (Qiao et al., 2017). Moreover, in the AID/dCMP crystal structure, R25 interacts with 5′ phosphate, N51 hydrogen bonds with the 3′-OH, and Y114 interacts with O5′. Replacing RNA with DNA creates steric clashes between R25 and the 2′-OH in ribose, AID therefore binds RNA but cannot act on it. This suggests that the proper interaction with 5′ phosphate is essential in placing the dC in the catalytic pocket for efficient AID activity (Qiao et al., 2017).[a]

1.5.1.1 Conserved structural features of AID

AID has several functional regions that appear to have conserved structure-function relationships within the vertebrate class (Barreto & Magor, 2011). Among these well-established functional domains are the catalytic domain, the secondary catalytic residues, the substrate-binding groove(s), the conformational classical nuclear localization signal

[a] The AID crystal structure also showed that the binding of the substrate might induce or stabilize the F115 side chain to flip.

(NLS), the nuclear export signal (NES), the cytoplasmic retention residues, and the putative phosphorylation sites (Barreto & Magor, 2011; King et al., 2015; Qiao et al., 2017). Collectively, these motifs greatly impact the outcome of AID expression by regulating its activity, substrate specificity, and subcellular trafficking.

The catalytic domain and the secondary catalytic residues catalyze the deamination reaction and stabilize the dC in the active site, respectively (refer to section 1.5.2.2 for further details) (Barreto & Magor, 2011; Conticello, 2008; Harris et al., 2002; King et al., 2015). The residues forming the substrate-binding groove(s) interact with the adjacent nucleotides; therefore, they establish the substrate specificity and hotspot motif (refer to section 1.5.1.3 for further details) (Qiao et al., 2017). The NLS, NES cytoplasmic retention residues regulate AID activity through modulating its nucleocytoplasmic shuttling (Brar et al., 2004; Hu et al., 2013; Ito et al., 2004; McBride et al., 2004; Patenaude et al., 2009). Specifically, the residues 19-RWAK-22, N51, and N53 generate a conformational classical NLS, while R8, K16, R19, and R171, revealed by mutational screening, are the most likely residues impacting human AID (Hs-AID) entry to nucleus and nucleoli (Hu et al., 2013; Patenaude et al., 2009). The last 16 residues at C-terminus of Hs-AID are essential for nuclear export and cytoplasmic retention (Brar et al., 2004; Hu et al., 2013; Ito et al., 2004; McBride et al., 2004; Patenaude et al., 2009). Similar to many other proteins and enzymes, phosphorylation is a post-translational mechanism that alters AID activity. Protein kinase A (PKA) phosphorylates AID at serine-38 (S38), mediating its interaction with the endonuclease APE1, which is necessary for CSR (Vuong et al., 2013). Phosphorylation of S3 (by PKC) inhibits SHM and CSR, while phosphorylated threonine-140 (T140) promotes

their occurrence (Chandra et al., 2015; Vaidyanathan et al., 2014). Considering the importance of these motifs in AID activity and regulation, it is reasonable to assume that these motifs/residues would be subject to high conservation throughout AID evolution.

1.5.1.2 The primary and secondary catalytic residues of AID

Like other zinc-dependent deaminases, four evolutionarily conserved amino acid residues within the H(A/V)$Ex_{(24-36)}$PCxxC motif form the catalytic core of AID where the two cysteines, the histidine (in Hs-AID: C87, C90, and H56), and a water molecule coordinate the Zn^{2+} ion while the catalytic glutamate (in Hs-AID: E58) donates a proton (Figure 1-3) (Barreto & Magor, 2011; Conticello, 2008; Harris et al., 2002; Holden et al., 2008; King et al., 2015; Qiao et al., 2017; Silvas & Schiffer, 2019). The deamination reaction occurs when the Zn^{2+}-activated water molecule (in the form of Zn-hydroxide) performs a direct nucleophilic attack at the amine group (*i.e.*, -NH_2 on the C4) of the dC pyrimidine ring (Conticello, 2008; Holden et al., 2008; Silvas & Schiffer, 2019). The interaction between the glutamate side chain and the N3 of the pyrimidine ring (*i.e.*, protonation of N3 nitrogen by the carboxyl group of catalytic glutamate [OE1]) facilitates this attack (Qiao et al., 2017; Silvas & Schiffer, 2019). This glutamate also transfers the proton from the Zn-hydroxide group to the leaving NH_3 molecule (ammonia) through its side chain (OE2). The result is the replacement of the amine group with oxygen (creating a carbonyl group [C=O]) on the C4 of the pyrimidine ring, which converts dC to dU. Protonation of the catalytic glutamate carboxyl group by a new water molecule that coordinates the Zn^{2+} ion to regenerate the Zn-hydroxide group resets the catalytic site (Silvas & Schiffer, 2019).

The proper positioning of dC inside the active site is necessary for efficient deamination activity. Prior to solving the crystal structure of AID, the computational modeling and DNA:protein docking revealed a network of amino acid residues that either contact and/or stabilize the dC in catalytic pocket (King et al., 2015). This network of amino acids was named secondary catalytic residues and consist of G23, R24, R25, E26, T27, L29, N51, K52, N53, G54, C55, V57, T82, W84, S85, P86, D89, Y114, F115, C116, and E122 in Hs-AID (King et al., 2015).[a] These residues form the "walls" and "floors" of the catalytic pocket and interact with substrate dC in several predicted protein conformations through hydrogen bonding, electrostatic interactions, and aromatic base stacking (King et al., 2015). Remarkably, the importance of direct interactions between some of the secondary catalytic residues and substrate DNA was validated when the crystal structure of AID was published. Among these are R24, R25, T27, N51, K52, W84, S85, Y114, and F115 (Qiao et al., 2017). The nature of these interactions is outlined in the following section. Additionally, Y114 and F115 may play a significant role in shaping the catalytic pocket and defining the substrate specificity of AID (Gajula et al., 2014). Nonetheless, the primary and secondary catalytic residues are both vital in effective enzymatic activity of AID.

[a] The substrate dC interacted with T27, N51, and W84 in 63%–75% of models. In 25%–50% of models, R25, V57, S85, P86, and Y114 also participated in these interactions. However, the interaction with R24, E26, K52, F115, G23, L29, N53, G54, C55, T82, D89, C116, and E122 were only noticed in 6%–18% of the models.

1.5.1.3 DNA and RNA binding groove(s) of AID

Multiple ssDNA and ssRNA binding grooves have been identified on the surface of AID (King & Larijani, 2020). Prior to the "near native X-ray crystal structure" of AID, DNA binding groove 1 and 2 were identified through "computational-evolutionary-biochemical" approach (King et al., 2015). The DNA binding groove 1 is mainly formed by α1-β1, β2-α2, and β4-α4 loops (ℓ2, ℓ4, and ℓ8, respectively). This DNA binding groove is positively charged and starts at the junction of ℓ2 and ℓ4, passes over the catalytic pocket, travels along ℓ2, and ends at the junction of ℓ2 and ℓ8. The ℓ2 interacts with the +1 position, while the ℓ8 interacts with the bases at the -1 and -2 positions (with respect to the dC) and defines the substrate specificity in the AID/APOBEC family (Gajula et al., 2014; Iyer et al., 2011; Kohli et al., 2009). Notably, through DNA:protein docking, the presence of the DNA binding groove 2 has been predicted to start at the junction of ℓ2 and ℓ4, to pass over the catalytic pocket, but to continue along the valley between the α2 and α3 (King et al., 2015).

The recent crystal structure of AID revealed the presence of a bifurcated substrate-binding surface that consists of the substrate channel and the assistance patch (Qiao et al., 2017). The substrate channel is identical to the previously identified DNA binding groove 1, while the assistant patch is a separate collection of positively-charged amino acids of α6 (Qiao et al., 2017). Mutating the assistant patch affected AID activity only on structured ssDNA and G4-containing substrates (Qiao et al., 2017).

Two putative RNA binding domains have also been identified on the surface of AID by mutagenesis and biochemical-computational approaches (King & Larijani, 2020). One RNA binding groove is predicted to be formed by amino acids 130 to 138 (in mouse AID) based on the homology with the G4 RNA binding domain of the RNA helicase associated with AU-rich element (RHAU) protein (Creacy et al., 2008; Vaughn et al., 2005; Zheng et al., 2015).[a] Interestingly, a single mutation in this region (*i.e.,* G133V) was found in HIGM patients manifesting lack of CSR (Mahdaviani et al., 2012). A second RNA binding groove was also predicted which overlaps with the first RNA binding groove but also includes amino acid residues from α7 (Abdouni et al., 2018). The second RNA binding groove was suggested to be involved in AID activity in the context of DNA/RNA hybrids (Abdouni et al., 2018). Although AID has no catalytic activity on pure RNA, its RNA binding grooves are thought to facilitate attraction of AID to R-loops or DNA/RNA hybrid structures which are abundant at the *Ig* loci during SHM and CSR (King & Larijani, 2020). Interestingly, the ability of AID to target dC in the context of diverse structures (*i.e.,* ssDNA bubbles, DNA/RNA hybrids, stem loops, and G4 structures) was attributed to the combinatorial usage of its multiple substrate binding motifs (King & Larijani, 2020). Previous studies have shown that the abovementioned structures are abundant at the AID-targeted *IgV* and *IgS* regions (Chaudhuri & Alt, 2004; Chaudhuri et al., 2003; Roy et al., 2008; Yu et al., 2003). Therefore, the presence of multiple substrate binding motifs was

[a] RHAU is also known as DHX36 or G4R1.

suggested as an evolutionary feature of AID structure to regulate its activity at various loci (King & Larijani, 2020).

1.5.2 Biochemical and enzymatic properties of AID

Previous studies estimated that the AID catalytic turnover rate was 1 to 4 minutes per reaction, defining AID as a lethargic enzyme compared with most other enzymes (King et al., 2015). The strong affinity of AID for ssDNA contributes to the long half-life of the AID-ssDNA complex (approximately eight minutes), resulting in the slow catalytic rate and high enzymatic processivity of AID (Choudhary et al., 2018; Larijani et al., 2007; Mak et al., 2013). Single-molecule resolution experiments revealed a random bidirectional short slides/hops movement where 80% of AID molecules remained bound to ssDNA for 25 s to 10 min (with an average time of 270 ± 30 s) (Senavirathne et al., 2015). AID is capable of forming a multimer complex, a characteristic that is critical for CSR but not SHM, potentially due to promoting a clustered mutation pattern. Interestingly, substrates forming G4 structures that resemble the structure of the mammalian *Ig* S regions proved to be the preferred substrate and induced the cooperative oligomerization of AID (Choudhary et al., 2018; Qiao et al., 2017).

A catalytic pocket occlusion was suggested as an internally built-in mechanism to regulate AID/APOBEC activity (King & Larijani, 2017). Using a combined computational-evolutionary-biochemical approach, this novel regulator of AID/APOBEC activity was described, where the catalytic pocket could transition between a closed (*i.e.*, catalytically inactive) and an open (*i.e.*, catalytically active) state due to the flexibility of the component loops (King & Larijani, 2017; King et al., 2015). This catalytic pocket

duality was termed “Schrodinger’s CATalytic pocket”. Based on this regulatory mechanism, the majority of Hs-AID conformations (~ 75 %) at any given time contain catalytic pockets that are closed and inaccessible for accommodating a dC. Furthermore, the majority of ssDNA:AID docking events resulted in non-productive binding modes (*i.e.*, the conformations where the substrate does not pass over the catalytic pocket) due to the highly positively charged surface of AID (King & Larijani, 2017; King et al., 2015).[a] Therefore, the frequent catalytic closure and sporadic ssDNA binding are significant bottlenecks for AID activity, such that < 1 % of all ssDNA:AID binding events translate into a cytidine deamination event (King & Larijani, 2017; King et al., 2015).

Nevertheless, the enzymatic robustness of AID catalytic activity is an important determinant of SHM and CSR efficiency (Larijani & Martin, 2012; Wang et al., 2009). Studying AID from different species demonstrated that its biochemical characteristics, such as catalytic rate and optimal temperature, vary significantly amongst different species (Dancyger et al., 2012; King et al., 2015). In general, the mutator activity of mammalian and avian AID is higher at 37 ºC, while the amphibian and bony fish AIDs are more active at lower temperatures. More importantly, various AIDs at their optimal temperature exhibit significantly different catalytic rates (Barreto et al., 2005; Conticello et al., 2005; Dancyger et al., 2012; Ichikawa et al., 2006; Wakae et al., 2006). For example, at their optimal temperature, zebrafish (*Danio rerio*) AID (Dr-AID) is catalytically more robust than Hs-AID, which is more active than *I. punctatus* AID (Ip-AID) (Abdouni et al., 2013; Dancyger

[a] Hs-AID has the surface charge of +10.25 at pH 7, which is the highest positive surface charge amongst AID/APOBECs members.

et al., 2012; King et al., 2015). Since AID is the key enzyme initiating the secondary antibody diversification process, the biochemical properties of AID may greatly delineate the outcome of the humoral immune response.

Intriguingly, previous studies showed that AID might also deaminate 5-methyl dC (5m-C) although less efficiently than dC. Based on this observation, it was suggested that AID might play a role in epigenetics and genetic reprogramming. For example, AID-mediated deamination of 5-mC has been reported in induced pluripotent stem (iPS) cells, primordial germ cells, B cells, cancerous cell lines, and bovine and zebrafish embryo (Ao et al., 2016; Bhutani et al., 2013; Dominguez et al., 2015; R. Kumar et al., 2013; Moon et al., 2016; Munoz et al., 2013; Popp et al., 2010; Rai et al., 2008). However, this hypothesis has been challenged by a growing body of evidence (Habib et al., 2014; Hogenbirk et al., 2013; Kunimoto et al., 2017; Ramiro & Barreto, 2016; Shimamoto et al., 2014; Shimoda et al., 2014). Besides these *in vivo* evidence, some *in vitro* studies have claimed that Hs-AID efficiently deaminates 5-mC, while others showed that it was inefficient (Abdouni et al., 2013; Larijani, Frieder, Sonbuchner, et al., 2005; Morgan et al., 2004; Nabel et al., 2012; Wijesinghe & Bhagwat, 2012). For instance, it was previously shown that Dr-AID, Hs-AID, and Ip-AID have different deamination efficiency ratio of dC/5m-C substrates. While Dr-AID is the most robust enzyme on 5m-C (2/1), Hs- and Ip-AIDs were not efficient in deaminating 5m-C (Abdouni et al., 2013). In general, all AIDs studied from various species thus far showed less activity on 5m-C compared with dC. This observation led to the suggestion that methylation protects dC from AID targeting, this protection, however, is more restricted in humans compared with zebrafish due to the enzymatic

differences between their AIDs (Abdouni et al., 2013; Larijani, Frieder, Sonbuchner, et al., 2005).

1.5.3 Co-evolution of AID substrate specificity with *Ig* genes

Sequencing analyses of *IgV* genes and biochemical studies of AID from different species have defined the WRC motif as its favored target motif (Dancyger et al., 2012; Gajula et al., 2014; Hackney et al., 2009; Larijani, Frieder, Basit, et al., 2005; Larijani & Martin, 2007; Malecek et al., 2005; Marianes & Zimmerman, 2011; Yang et al., 2006).[a] Specifically, *in vivo* analyses revealed that AGCT is the AID preferred motif in *IgV* genes, and WRCH/DGYW motifs are mildly enriched in mammalian *IgV* regions (Hackney et al., 2009). Since the frequency of SHM is correlated with the recurrence of the AID hotspot in the *IgV* regions, a co-evolution between AID substrate specificity and the *Ig* gene sequences were proposed (Choudhary et al., 2018). This co-evolution has been observed in mammals, birds, amphibians, as well as bony and cartilaginous fish (Conticello et al., 2005; Detanico et al., 2016; Golub & Charlemagne, 1998; Jolly et al., 1996; Oreste & Coscia, 2002; Wagner et al., 1995; Wei et al., 2015). The analysis of the human *IGHV3-23*01* region revealed an accumulation of overlapping AID (especially AGCT) and Polη (WA) hotspots in the CDR1 and 2 compared to the framework (FRs) regions, suggestive of a co-evolution between *IgV* sequence and the SHM machinery (Wei et al., 2015).[b] Using deep sequencing, it has been shown that the replacement of the hotspots with neutral or

[a] However, more distant homologs such as cartilaginous fish and lamprey AID exhibit divergent patterns of sequence specificity, often favoring non-WRC motifs (Quinlan *et al.*, 2017).
[b] CDR3 was excluded from the analyses.

coldspots reduced mutation frequency in the CDR1 and 2 as well as the entire *IgV* region (Wei et al., 2015). Additionally, when dividing serine codons into AGY (WRC) and TCN (non-WRC), a clear preference for AGY over TCN was observed in *IgV* CDRs *vs.* FRs (Detanico et al., 2016; Golub & Charlemagne, 1998; Jolly et al., 1996; Wagner et al., 1995). Moreover, analyses of the antibody-antigen (Ab-Ag) crystal structure of human and mouse revealed that somatic mutations in AGY codons in CDRs are responsible for generating 4 out of 7 of the most abundant residues involved in Ab-Ag interactions (Detanico et al., 2016). Additionally, the WGCW motifs, which contain AID hotspots on both strands, have been suggested to attract AID to the *IgV* regions (Hwang et al., 2017; Ohm-Laursen & Barington, 2007; Wei et al., 2015; Yeap et al., 2015). The analyses of WGCW distribution revealed these overlapping motifs as a key evolutionary feature of IgV_H genes in human (Tang et al., 2020). This apparent co-evolution of AID substrate specificity and the sequence of *Ig* genes may play a significant role in targeting AID activity towards *Ig* genes.

1.5.4 AID transcript and expression pattern

Thus far, there have been reports of alternative AID transcripts and isoforms in several, but not all, studied vertebrates. In bony fish, no AID alternative splice variant has been reported in *I. punctatus* nor *D. reiro* (Saunders & Magor, 2004; Zhao et al., 2005). In amphibians, cloning of the *Pleurodeles waltl* (Iberian ribbed newt) AID cDNA revealed the presence of three potential poly-A sites and two isoforms, one of which is missing the first exon (Bascove & Frippiat, 2010). Two different AID transcripts were found in *X. laevis* spleen (2 and 1.3 kb) (Marr et al., 2007). Only one AID transcript was reported in dogs (*Canis lupus familiaris*) and cows (*Bos taurus*) while two AID transcripts were

identified in a murine B cell line (CH12F3-2), both containing full-size AID ORF but utilizing different poly-A sites (Muramatsu et al., 1999; Ohmori et al., 2004; Verma et al., 2010). In humans, five different splice variants of AID have been detected: Full-length AID (AID-FL), exclusion of the beginning of exon 4 (AID-ΔE4a), exclusion of exon 4 (AID- ΔE4), exon 3 and 4 exclusion (AID- ΔE3E4), and inclusion of intron 3 containing a stop codon (AID-ivs3) (Albesiano et al., 2003; Babbage et al., 2004; Greeve et al., 2003; McCarthy et al., 2003; Noguchi et al., 2001; Oppezzo et al., 2003; Wu et al., 2008). Noteworthy, individual human B cells only express one of the AID splice variants (Wu et al., 2008). Since Hs-AID splice variants have different functional properties in carrying SHM or CSR, it was suggested that differential splicing of AID in normal and malignant B cells might play a crucial role in antibody maturation regulation and tumor suppression (Wu et al., 2008).

AID expression can be induced during B cell activation, either through interaction of peptide-MHC II complex and CD40 on B cells with TCR and CD40L on T_H cells (*i.e.*, TD B cell activation), or through dual engagement of B cell receptor and TLRs on B cells with antigens such as LPS (*i.e.*, TI B cell activation) (DeFranco, 2016; Hou et al., 2011; Kasturi et al., 2011; Pone et al., 2012; Stavnezer & Schrader, 2014). Importantly, the effect of TI activation of B cells in AID expression and CSR induction is comparable with that induced by the TD pathway. AID induction through the TI pathway peaked between 24 to 48 hours (100-fold increase) in stimulated murine B cells (Pone et al., 2012). Therefore, although both pathways lead to comparable AID expression, the latter pathway takes place early in immune response when T_H cell assistance is not yet available (Pone et al., 2012).

Consistent with its role in secondary antibody diversification, all studies conducted on vertebrates have identified lymph node and spleen (where TD B cell activation occur) as the main AID expressing tissues (Bascove & Frippiat, 2010; Marr et al., 2007; Muramatsu et al., 1999; Muto et al., 2000; Ohmori et al., 2004; Saunders & Magor, 2004; Verma et al., 2010). AID is mainly expressed in activated GC B cells (Muramatsu et al., 1999). Canonical GCs in the lymph node of mammals and spleen of birds are the main sites of TD B cell activation. Although reptiles and amphibians lack the conventional GC, TD activation of B cells occurs in their spleen (Boehm et al., 2012). A previous study on *I. punctatus* identified melano-macrophage clusters (MMCs) as the main site of AID-expressing B cells in early gnathostome vertebrates (Saunders et al., 2010). In most fish species, these clusters exist in the spleen and posterior kidney, and they contain large macrophage aggregates and pigment-containing cells (Agius & Roberts, 2003). MMCs have been suggested as the antigen-trapping sites where the antigens may persist for a long-term similar to the birds' and mammalian germinal centers (Ellis, 1980; Lamers, 1986). Therefore, these clusters have been suggested as the primitive analogues of the germinal centers in fish (Agius & Roberts, 2003). However, lower and variable levels of AID expression have also been reported in thymus, pancreas, kidney, liver, and lungs of mammals (Muto et al., 2000; Ohmori et al., 2004; Verma et al., 2010). Likewise, low levels of AID expression have been observed in the brain, intestine, kidney, liver, and lungs of amphibians, and in the intestine, fin, posterior, and anterior kidney of fish (Bascove & Frippiat, 2010; Marr et al., 2007; Saunders & Magor, 2004).

A controversial role for AID in epigenetics reprogramming has been suggested through the deamination 5-mC leading to the CpG motif demethylation (Bhutani et al., 2013; Dominguez et al., 2015; Moon et al., 2016; Popp et al., 2010; Rai et al., 2008). Thus far, Dr-AID is the only AID homolog that efficiently deaminates 5m-C (Abdouni et al., 2013; Larijani, Frieder, Sonbuchner, et al., 2005; Nabel et al., 2012; Wijesinghe & Bhagwat, 2012). Interestingly, AID expression was reported during most embryonic stages in zebrafish, where AID knockdown by morpholinos caused loss of neurons (Rai et al., 2008). However, these findings were reported to be unreproducible in a later publication (Shimoda et al., 2014). AID expression was also observed in the early stages of embryogenesis in Iberian ribbed newt (*Pleurodeles waltl*) and early larval stages in African clawed frog (Bascove & Frippiat, 2010; Marr et al., 2007).

1.5.5 AID regulation and targeting

Despite the central role of AID in humoral immune responses, its off-target activity would be costly for the cells (Choudhary et al., 2018; Lindley et al., 2016; Silvas & Schiffer, 2019). Therefore, AID expression and activity is highly regulated and mostly directed towards *Ig* genes. Many mechanisms have been identified that regulate AID expression and activity. *Aicda* expression and AID shuttling to the nucleus are mainly restricted to activated B cells in the DZ of GC (de Yebenes & Ramiro, 2006; Mai et al., 2010; Owen, 2019). In these cells, *aicda* expression is regulated through *cis*- and *trans*-

acting factors, such as Stat6, Smad3 and 4, Pax5, E2A, BATF, NF-κB, HoxC4,[a] Myb, and E2F transcription factors,[b] and the stability of its mRNA is governed through micro-RNAs, such as miR-155, miR-181b, and miR-93 (Zan & Casali, 2013).[c] In the cytoplasm, AID protein exists in a high molecular mass complex with other proteins, such as Hsp90 and translation elongation factor 1α (eEF1α) to prevent its degradation and regulate its entry into the nucleus (Häsler et al., 2012). Moreover, it was demonstrated that phosphorylation of AID^{S38} by PKA permits its association with replication protein A (RPA), which enhances AID activity (Basu et al., 2005; Basu et al., 2008; Chaudhuri & Alt, 2004; Methot & Di Noia, 2017).[d] Also, the inefficiency of AID to deaminate dC even at preferred hotspot motifs (~ 3%) contributes to protecting genomic DNA from excessive AID-mediated mutations. This phenomenon is mostly due to its lethargic catalytic rate and extremely high substrate binding affinity (Chi et al., 2020; King & Larijani, 2017; Larijani & Martin, 2012; Mak et al., 2013). These various levels of regulation are crucial to prevent off-target activity of AID.

Beside regulation of AID expression and activity, numerous studies examined the molecular basis of AID targeting towards *Ig* genes. Many factors have been proposed to

---

[a] Interestingly, it was shown that estrogen enhances antibody and autoantibody responses by increasing AID expression through inducing the expression of HoxC4, a critical *aicda* gene activator. This phenomenon was suggested to contribute to more robust antibody responses in females (Mai *et al.*, 2010).

[b] Among these transcription factors, Myb and E2F inhibit *aicda* expression, while others induce its expression.

[c] It was suggested that these micro-RNAs protect resting B cells and non-B cells against AID-mediated mutations by reducing AID protein level. Accordingly, Burkitt's lymphoma patients are deficient in miR-155 and show high levels of somatic mutations and chromosomal translocations.

[d] Interestingly, given the importance of serine 38 phosphorylation in CSR, the lack of this serine residues in bony fish AIDs, and absence of CSR in bony fish, it was suggested that serine 38 and its phosphorylation are evolutionary adaptations to emergence of CSR in higher vertebrates (Basu *et al.*, 2008).

define the selectivity of AID targeting towards *Ig* genes, such as the target sequence, transcription, and protein co-factors (Choudhary et al., 2018). Studies have suggested that *Ig* gene primary sequence may direct AID activity towards CDRs and S regions (Choudhary et al., 2018; Conticello et al., 2005; Detanico et al., 2016; Golub & Charlemagne, 1998; Hackney et al., 2009; Jolly et al., 1996; Oreste & Coscia, 2002; Wagner et al., 1995; Wei et al., 2015). As described in section 1.5.3, a co-evolution of AID substrate specificity with *IgV* primary sequence has been previously proposed (Choudhary et al., 2018). S regions are also moderately enriched with WRC motifs, AID's hotspots. However, replacement of *IgV* and S regions with heterogenous sequences would not diminish SHM and CSR (de Yebenes & Ramiro, 2006). Also, not all hotspots are targeted equally which means that some other local sequences or higher-order structures are also involved. Recently, the plasticity in AID's substrate choice, due to containing multiple substrate binding motifs on its surface, has also been proposed as a regulatory mechanism of its activity at various loci (refer to section 1.5.1.3) (King & Larijani, 2020). Therefore, it seems that WRC enrichment and higher abundance of structured substrates (*e.g.,* ssDNA bubbles, R-loops, DNA/RNA hybrids, and G4) at the AID-targeted *IgV* and *IgS* regions may contribute to AID targeting towards *Ig* genes (Chaudhuri & Alt, 2004; Chaudhuri et al., 2003; Roy et al., 2008; Yu et al., 2003).

Previous studies have shown that active transcription of *Ig* genes is required for both SHM and CSR (Betz et al., 1994; de Yebenes & Ramiro, 2006; Fukita et al., 1998; Goyenechea et al., 1997; Mandler et al., 1993; Peters & Storb, 1996; Pinaud et al., 2001; Rothenfluh et al., 1993; Storb et al., 1998; Xu et al., 1993; Zhang et al., 1993). Unique

transcription dynamic features, such as strong enhancers and bi-directional transcription at *Ig* loci (Meng et al., 2014; Qian et al., 2014) have emerged as features that explain AID's genome targeting patterns and preference for targeting *Ig* loci. Facilitation of AID targeting through transcription may happen through de-chromatinization of DNA (Kodgire et al., 2012; Kodgire et al., 2013; Shen et al., 2009) and generation of ssDNA in the context of AID-preferred structured substrates (*e.g.,* ssDNA bubbles, R-loops, DNA/RNA hybrids, and G4) (Branton et al., 2020; Fugmann & Schatz, 2003; Kim & Jinks-Robertson, 2012; Yu et al., 2003; Yu et al., 2005).

Many studies also considered that in addition to ssDNA generation, another way transcription might facilitate AID targeting is through association of AID with the RNAP complex and/or transcription machinery-associated protein co-factors. Thus far, many protein cofactors have been proposed to recruit AID to *Ig* genes. Example of these proposed co-factors are: RNAPII (Nambu et al., 2003), the ssDNA binding protein Replication protein A (RPA) (Chaudhuri et al., 2004), the transcription elongation factor Spt5 (Pavri et al., 2010), RNAPII associated factor I (PAF1) (Willmann et al., 2012), spliceosome-associated factor CTNNBL1 (Conticello et al., 2008), RNA binding heterogeneous nuclear ribonucleoproteins (hnRNP) (Hu et al., 2015), splicing regulator polypyrimidine tract binding protein 2 (PTBP2) (Nowak et al., 2011), splicing factor SRSF1-3, (Kumar Singh et al., 2019), and the chromatin-associated SUV4-20H2 (Rodríguez-Cortez et al., 2017). Though in different instances some of these co-factors may be involved in guiding AID to a specific target, none could fully explain preferential targeting of AID to the *Ig* loci while at the same time accounting for its genome-wide targeting and lack of specificity.

Additionally, the distribution of the proposed co-factors at the *Ig* genes, the small size of AID (only 198 amino acids in human), and its highly positively charged surface were used to dispute the role of co-factors in targeting AID (King & Larijani, 2017). Recently, the earlier observation that AID can indeed act efficiently on supercoiled dsDNA in the absence of transcription was confirmed using an unbiased PCR-based assay. Furthermore, it was shown that AID can also act on relaxed linear dsDNA in the absence of transcription, and that even the most optimal transcription conditions only modestly enhances AID activity on supercoiled dsDNA (Branton et al., 2020). Based on these findings, it was suggested that the association between transcription and AID targeting may indeed be due to transcription being a corollary of de-chromatinized naked loci rendered accessible for AID to target breathing ssDNA regions naturally found in supercoiled dsDNA, as well as transcription being a direct generator of ssDNA (Branton et al., 2020).

Nevertheless, despite the tight regulation of AID expression and activity, AID may off-target oncogenes resulting in somatic mutations, chromosomal translocation, and subsequent cell transformation and tumor development (Choudhary et al., 2018; Lindley et al., 2016; Silvas & Schiffer, 2019).[a] Indeed, a source of genome instability and mutations in B cells is the mis-targeted activity of AID (Choudhary et al., 2018). For instance, AID expression and activity have been suggested as a main contributing factor in *IgH-cMyc* translocations manifested in the patients with Burkitt's lymphomas (Takizawa et al., 2008). AID-mediated mutations are also identified in serous ovarian adenocarcinoma and chronic

[a] AID off-targets other genes such as *cd95*, *cd79a*, *cd79b*, *pim1*, c-*myc*, *rhoh*, and *pax5* genes.

lymphocytic leukemia (CLL) (Burns et al., 2017; Lindley et al., 2016). In patients with chronic myeloid leukemia (CML), AID-mediated hypermutation of tumor repressor and DNA repair genes have been associated with progression into fatal B lymphoid blast crisis and Imatinib-resistance phenotype (Klemm et al., 2009). In diffuse large B cell lymphomas (DLBCL), somatic hypermutation (SHM) off-targeting has been reported in proto-oncogenes (Seifert et al., 2019). There has also been evidence of AID-mediated carcinogenesis in GC B cells as the result of Epstein-Barr virus (EBV)-induced AID expression (Mohri et al., 2017). Interestingly, under strong inflammatory stimuli, the premature expression of *AID* during B-cell development creates an opportunity for cooperation between RAG and AID to drive the clonal evolution of childhood B cell acute lymphoblastic leukemia (B-ALL) (Swaminathan et al., 2015). It was proposed that aberrant AID-mediated mutations in CpG islands would create T:G mismatches which would attract RAG complex activity, causing genome instabilities. AID- and APOBEC3-mediated mutations have been observed in many types of cancers, such as breast, ovarian, and lung cancers, as the driving mutation and potentially cancer-progression associated signatures (Leonard et al., 2013; Lindley et al., 2016; Ruder et al., 2019; Sasaki et al., 2014; Zou et al., 2017). Taken together, AID which is used by the adaptive immune system towards antigen receptor diversification, also mediates considerable collateral mutation and damage to the host cell's genome, and is therefore aptly considered to be a double-edged sword.

1.6 Evolution of the AID/APOBEC family

AID belongs to the vertebrate-specific polynucleotide cytidine deaminase family of the *apolipoprotein B mRNA editing enzyme catalytic polypeptide* (APOBEC) (Methot

& Di Noia, 2017). Controversial to this, BLAST search results revealed the presence of the AID/APOBEC-like deaminases in *Wolbachia* endosymbiont (parasitic bacteria), nematodes, and distantly related algal lineages (Iyer et al., 2011). Nevertheless, the AID/APOBEC family contains 11 members in humans: AID, APOBEC1, APOBEC2, the APOBEC3 sub-branch (A-H, excluding E), and APOBEC4. AID and APOBEC3s act on DNA and are involved in antibody maturation and viral protection, respectively. APOBEC1 participates in lipid transport by editing the apolipoprotein B mRNA, while the roles of APOBEC2 and 4 are still unknown (Conticello, 2008; Silvas & Schiffer, 2019).[a]

In a comprehensive phylogenetic analysis, a bacterial toxin deaminase, capable of binding metal ions and a nucleotide or a related molecule, was suggested as the ancestor of all deaminases from which two deaminase divisions of the C-terminal hairpin and the Helix-4 were derived. The β sheet four (β4) and β5 are anti-parallel in the C-terminal hairpin division while the presence of the intervening α-helix four (α4) causes β4 and β5 to be parallel in the Helix-4 division, including all tRNA deaminases (TADs), adenosine deaminases acting on RNA (ADARs), and the AID/APOBEC family (Iyer et al., 2011).[b]

It is suggested that, at the beginning of the vertebrate radiation, the AID/APOBECs family has evolved from the tRNA adenosine deaminases containing the consensus motif (C/H)xEx$_n$PCxxC (x is any given amino acid) as their catalytic domain (Conticello, 2008; Torres et al., 2014). The shift in substrate specificity from adenine to cytidine during the

[a] The editing of apolipoprotein B mRNA by APOBEC1 results in a stop codon, producing a truncated apoB protein that is essential for lipid transport from the intestine to other organs.
[b] All the proteins in the Helix-4 division share a HxE motif in their α2.

divergence of the AID/APOBEC family from Tad2/TadA deaminases has been attributed to the expansion of the α4-β4 loop (*i.e.*, ℓ8) and a conserved tyrosine in this loop. The larger ℓ8 decreases the size of the substrate-binding pocket, and the conserved tyrosine could participate in base-stacking interactions (Iyer et al., 2011). Moreover, the $HxEx_nPCxxC$ motif is the conserved catalytic domain shared by the AID/APOBEC family in which the glutamate (E) acts as a proton donor and the histidine (H) with two cysteines (C) coordinate a Zn^{2+} ion with the help of a water molecule (Qiao et al., 2017; Silvas & Schiffer, 2019).

The evolution of the AID/APOBEC family within the vertebrate class starts with the divergence of AID-like and the APOBEC4-like clades where the fourth Zn^{2+}-coordinating agent is a water molecule or a cysteine residue (located between β2 and α2), respectively (Iyer et al., 2011; Qiao et al., 2017). In jawless vertebrate, the AID-like branch then gave rise to PmCDA1 and PmCDA2. In jawed vertebrates, this branch has further diverged into AID and APOBEC2 (at the base of jawed vertebrates), APOBEC3 (in tetrapod) and APOBEC1 (in mammals) (Iyer et al., 2011). Interestingly, the involvement of PmCDA1 in diversifying the lamprey's immune receptors and the continuing of a similar role for AID in the jawed vertebrates indicates that the acquisition of this role by the AID-like branch had already occurred before the further divergence of this branch within vertebrates (Emma M. Quinlan, 2017; Iyer et al., 2011).

1.7 Evolution of antibody maturation within the vertebrate class

Functional and genomic analysis of antibody repertoires in various vertebrates revealed the emergence of the antibody affinity maturation process as early as cartilaginous fish (Betz et al., 1993; Bromage et al., 2006; Cain et al., 2002; Diaz et al., 1999; Dooley &

Flajnik, 2005; Dooley et al., 2006; Hsu, 2016; Jenne et al., 2003; Kaattari et al., 2002; Lee et al., 2002; Malecek et al., 2005; Marianes & Zimmerman, 2011; Mehr et al., 2004; Wilson et al., 1992; Yang et al., 2006). Specifically, AID-mediated mutations were identified in the CDRs of *Ig* genes in studied poikilotherms. In one study on *Xenopus* (frog), a five to 10-fold increase in antibody affinity was observed four weeks after immunization with 2,4-dinitrophenyl-keyhole limpet hemocyanin (DNP-KLH). DNP-KLH is a highly immunogenic TD antigen that can be used to study the T cell-dependent immune response in animals (Kojima et al., 2013). In the same study, point mutations were detected in the V_H1 region with 4 to 7-fold lower frequency than that reported for mice (Wilson et al., 1992). In *Oncorhynchus mykiss* (rainbow trout), a 2 to 3-fold increase in antibody affinity by week 14 after immunization with TD antigen (FITC-KLH) was reported (Cain et al., 2002). In a more detailed study in the same species, Kaattari *et al.* discovered the emergence of higher affinity antibodies later in the immune response, which suggests the presence of antibody affinity maturation process (Kaattari et al., 2002). G to A and C to T mutations in RGYW motifs were observed in *I. punctatus* CDR regions, however analyzing synonymous *vs.* nonsynonymous mutations showed no evidence of antigen-driven B cell selection (Yang et al., 2006). In 2011, the contribution of AID-mediated mutations in antibody diversification of *D. rerio* was confirmed by mutational analysis of the *IgL* cDNA library from a healthy individual. In this study, WRCH/DGYW motifs were described as the primary target of mutations in CDR regions (Marianes & Zimmerman, 2011). High frequency of somatic mutations has also been reported in nurse shark (*Ginglymostoma cirratum*) *Ig* genes. These somatic mutations could increase antibody affinity up to 10-fold

(Dooley et al., 2006). Therefore, while the extent of AM varies among studied vertebrates, the occurrence of SHM in their *Ig* genes and its contribution to AM seems to be somewhat conserved.

1.8 The genetically altered immune system of Gadiformes lineage

Ray-finned fishes (class Actinopterygii), with 33792 validated extant species, is the largest group of vertebrates and they inhabit every marine and freshwater habitat. Within the Actinopterygii class, the vast majority of species belong to the teleost lineage (Ron Fricke; Sallan, 2014; Solbakken et al., 2017). The recent genomic sequence of non-model fish species revealed a remarkable heterogenicity in the teleost's innate and adaptive immune systems, particularly within Gadiformes order. These variabilities include gene losses and/or expansions of *tlrs*, *mhc I* and *II*, *cd4*, invariant chain (also known as *cd74*), and Myxovirus resistance (*Mx*) genes (Malmstrom et al., 2016; Solbakken, Rise, et al., 2016; Solbakken, Torresen, et al., 2016; Solbakken et al., 2017). Although the functional consequences of these gene losses and expansions are still unclear, alternative immune strategies might have successfully replaced the classical immune system in Gadiformes species.

Past environmental changes are powerful evolutionary factors diversifying the vertebrates' immune system (Solbakken et al., 2017). In a series of publications, Solbakken and colleagues showed that the immune gene losses and expansions in teleost lineage overlap with major paleoclimatic and geological events. They associated the loss of *Mx* gene in the Gadiformes and *Stylephorus chordates* ancestor, and the loss of *mhc II* gene in the common ancestor of Gadiformes with the first (~120 million years ago [Ma]) and the

second global anoxia events (~95 Ma), respectively (Solbakken, Rise, et al., 2016). They also showed that the *tlr* expansions within teleost correlate with latitudinal distribution and the maximum depth, while *tlr* losses in the order of Gadiformes reflects the global ocean anoxia and the geography of the Atlantic Ocean in the past (Solbakken et al., 2017). It was suggested that the adaptability of the teleosts' immune system in response to the major changes in their habitat played a crucial role in their successful radiation and speciation (Malmstrom et al., 2016; Solbakken et al., 2017).

Intriguingly, the functional analyses of the Atlantic cod (*Gadus morhua*; a member of Gadiformes group) humoral responses showed high levels of low-affinity IgM and lack of robust antigen-specific antibody response upon immunization (Arnesen et al., 2002; Magnadottir et al., 1999; Magnadottir et al., 2001; Solem & Stenvik, 2006). Yet other studies claimed that Atlantic cod antibody response to *Aeromonas salmonicida* was comparable to that in salmon, and Atlantic cod produced specific antibody responses against *Francisella* and different *Vibrio anguillarum* serotypes (Lund et al., 2008; Lund et al., 2006; Schroder et al., 2009). However, antibodies measured in these studies were mainly LPS-specific with some of the serum pools reacting towards O-polysaccharide, which indicates the B cell activation through the TI pathway. Despite the loss of central genes required for TD B cell activation, evidence of TI B cell activation has been reported in this species (Malmstrom et al., 2013; Solbakken, Jentoft, Reitan, Mikkelsen, Gregers, et al., 2019; Solbakken, Jentoft, Reitan, Mikkelsen, Jakobsen, et al., 2019). Generally, the TD activated B cells almost exclusively secrete the highly pathogen-specific antibodies. The lack of TD pathway is consistent with the drastic re-modeling of immune genes in this

species. Taken together with the genetic re-modeling of the immune system, it appears that the Atlantic cod immune system has a unique gene structure and tactics which require a more detailed investigation.

1.9 Research hypothesis and objectives

AID is the functional initiator and master switch without which antibody affinity maturation is genetically not possible. The collective functional and genetic evidence is highly suggestive that the Atlantic cod humoral immune response is less robust than other studied bony fish. This phenomenon is most likely due to little or lack of antibody affinity maturation in this species. The sequencing of the Atlantic cod genome revealed the presence of a putative AID gene. However, it is not clear whether this gene expresses a functional enzyme during an immune response. Also, there has been a drastic remodeling of the immune system in the cod-like lineage of Gadiformes. Given the unique antibody responses of the Atlantic cod and the lack of robust AM, we asked whether its AID enzyme, the master switch for initiation of the molecular events of AM, may also be involved. Therefore, this thesis aims to shed light on the evolutionary plasticity of AID within the Gadiformes group, with an emphasis on Atlantic cod. By combining state-of-the-art *in vivo*, *in vitro*, and *in silico* analyses, we attempted to comprehensively examine the genetics, expression, and function of AID in Atlantic cod and address the evolutionary trajectory of this enzyme within the Gadiformes group. Therefore, this thesis has three specific objectives that are addressed in the three following chapters.

In chapter 2, we examined the AID gene structure, synteny, expression, and immune responsiveness. Specifically, we compared the genetic structure of AID and its

gene synteny with other studied vertebrates. We then characterized Atlantic cod AID mRNA and its expression pattern in a panel of different tissues. We also examined its expression upon immune stimulation in adult Atlantic cod individuals and during Atlantic cod embryogenesis. We concluded that Atlantic cod AID showed a conserved gene structure and transcript expression as compared with the previously studied species such as channel catfish and zebrafish.

In chapter 3, we examined the enzymatic properties of Atlantic cod AID. Since the Atlantic cod AID expression profile was similar to that of other studied species, we sought to assess its catalytic activity in comparison with AID from other species. For the first time, we reported that this enzyme has evolved to become nearly inactive in Atlantic cod, mirroring its lack of affinity matured antibodies. Correspondingly, we observed a significantly lower level of AID target sequences in the Atlantic cod *Ig* loci compared to other vertebrates. This phenomenon indirectly confirms the functional impairment of Atlantic cod AID during evolution. We also used computational modeling and DNA:protein docking to pinpoint the underlying molecular reason(s) for the lethargic activity of Atlantic cod AID.

In chapter 4, we investigated the plasticity of AID function among Gadiformes species by measuring the catalytic activity of 36 species within and outside of the Gadiformes lineage. We then predicted the ancestral sequence of AIDs within the Gadiformes family using Ancestral Sequence Reconstruction (ASR)- a powerful bioinformatics method. By comparing the ancestral AIDs, we showed that the catalytic activity of AID was drastically reduced in the ancestor of the Gadidae while its sister group

had retained a functional AID. In this light, our findings suggest that the Gadidae ancestor may represent an instance in the evolution of immunity wherein AID has become nearly inactive to reflect lesser reliance on high-affinity antibody responses.

Chapter 2:

Characterization of *aicda* gene structure, synteny, and expression in Atlantic cod (*Gadus morhua*)

2.1 Abstract

Activation-induced cytidine deaminase (AID; encoded by *aicda* gene) converts deoxycytidine (dC) into deoxyuracil (dU) at immunoglobulin (*Ig*) loci, initiating antibody affinity maturation. It was previously assumed that antibody affinity maturation existed in all jawed vertebrates. However, it was recently showed that the Atlantic cod was an exception since it lacks affinity-matured antibodies. Since AID is the key enzyme in generating a high affinity antigen-specific antibody response, we sought to examine the genetics and expression of *aicda* in Atlantic cod. Our data showed that Atlantic cod *aicda* locus conserved its synteny with other teleost species. In Atlantic cod immune-related tissues, we identified two *aicda* transcripts, one of which is missing the first exon. This truncated isoform, if translated, lacks the first 21 amino acids suggesting it is inactive as a cytidine deaminase. Comparison of the Atlantic cod AID amino acid sequence with that of other studied vertebrate species uncovered the presence of all AID's hallmark functional motifs. However, we noticed a potentially important difference in one of the predicted secondary catalytic residues in Atlantic cod AID's catalytic motif. Based on the structure-function knowledge of AID's catalytic pocket, this difference would likely affect Atlantic cod AID's activity as a cytidine deaminase. We found that a highly evolutionary conserved amino acid residues of E122 in human AID (Hs-AID) is a histidine in Atlantic cod (H136). The important role of secondary catalytic residues in stabilizing dC in the catalytic pocket of AID, the conservation of this amino acid in all AIDs studied thus far, and the previously shown functional impairment of Hs-AID^{E122A} mutant, are highly suggestive that the enzymatic activity of AID might have been compromised during the evolution of Atlantic

cod species. These findings are consistent with the lack of affinity-matured antibodies in Atlantic cod.

2.2 Introduction

Functional analyses of immune responses have indicated the presence of antibody immune response and antibody affinity maturation prior to the divergence of cartilaginous and bony fish (Abos et al., 2018; Bromage et al., 2006; Cain et al., 2002; Covello et al., 2013; Davidson et al., 1997; Dooley & Flajnik, 2005; Dooley et al., 2006; Hsu, 2016; Jenne et al., 2003; Kaattari et al., 2002; Malecek et al., 2005; Marianes & Zimmerman, 2011; Mehr et al., 2004; Wiens et al., 2003; Wilson et al., 1992; Yang et al., 2006; Zwollo et al., 2017). Specifically, a high frequency of somatic mutations resulting in antibody affinity maturation has been detected in IgM and the immunoglobulin new antigen receptor (IgNAR) of the immunized nurse shark (*Ginglymostoma cirratum*), improving affinity up to 10-fold (Dooley & Flajnik, 2005; Dooley et al., 2006). In rainbow trout (*Oncorhynchus mykiss*), the emergence of higher affinity antibodies (2- to 3-fold increase in affinity) by week 14 after immunization with T cell-dependent antigen (*i.e.*, FITC-KLH; fluorescein isothiocyanate [FITC] conjugated to keyhole-limpet hemocyanin [KLH]) has been reported (Cain et al., 2002; Kaattari et al., 2002). In immunized Atlantic salmon (*Salmo salar*), it was observed that the antibody affinity increased less than 10-fold (Solem & Stenvik, 2006). In African clawed frog (*Xenopus laevis*), a 5- to 10-fold increase in antibody affinity was detected four weeks after immunization with DNP-KLH (2,4-Dinitrophenyl [DNP] hapten conjugated to KLH protein through lysine) (Wilson et al., 1992). These reports support the idea that antibody affinity maturation is an ancient process dating back to the ancestor of jawed vertebrates.

Antibody affinity maturation is initiated when the enzyme AID introduces somatic hypermutation (SHM) in immunoglobulin (*Ig*) genes (Bransteitter et al., 2003; Kolar et al., 2007; Larijani, Frieder, Basit, et al., 2005; Meffre et al., 2001; Muramatsu et al., 1999; Muto et al., 2000). AID is mainly expressed in activated B lymphocytes where it converts deoxycytidine (dC) to deoxyuridine (dU) in *Ig* variable (V) genes, preferentially in the context of WRC (W=A/T; R=A/G) motifs (Bransteitter et al., 2003; Emma M. Quinlan, 2017; Larijani, Frieder, Basit, et al., 2005; Meffre et al., 2001). Studies have shown that in the mammalian model, AID-mediated SHM can enhance the affinity of antibodies for the cognate antigen as high as 1000-fold (Magor, 2015). Moreover, AID deficiency in mice and humans results in hyper IgM immunodeficiency characterized by a lack of affinity matured antibodies (Minegishi et al., 2000; Revy et al., 2000). AID-mediated SHM has also been reported in *IgV* genes of immunized frog, channel catfish (*Ictalurus punctatus*), zebrafish (*Danio rerio*), and the nurse shark (Dooley et al., 2006; Hsu, 2016; Marianes & Zimmerman, 2011; Mehr et al., 2004; Wilson et al., 1992; Yang et al., 2006). Point mutations were detected in the V_H1 region of *Ig* genes in the frog (*Xenopus*) with 4- to 7-fold lower frequency than that reported for mice (Wilson et al., 1992). In channel catfish, G-to-A and C-to-T mutations were observed in RGYW motifs of complementarity-determining regions (CDRs) (Yang et al., 2006). Mutational analyses of the *IgL* cDNA library from a healthy individual zebrafish confirmed the contribution of AID-mediated mutations in antibody diversification of this species. In this study, WRCH/DGYW motifs were the primary target of mutations in CDRs (Marianes & Zimmerman, 2011). Taken together, while the extent of increase in antibody affinity during immune response varies

among studied vertebrates, the occurrence of AID-mediated SHM in *IgV* regions appears to be a universal phenomenon that has been found in all vertebrate species in which it has been sought.

The Atlantic cod (*Gadus morhua*) antibody response has been shown to be different than that of other bony fish. Numerous studies have reported only low-affinity antibodies and a lack of a robust antigen-specific antibody response upon immunization, concluding that Atlantic cod has a weak humoral immune response (Arnesen et al., 2002; Corripio-Miyar et al., 2007; Magnadottir et al., 1999; Magnadottir et al., 2001; Solem & Stenvik, 2006). Intriguingly, previous studies have shown that the only antigen-specific antibody response detected in Atlantic cod, if any, is T cell-independent and mainly against LPS (Ellingsen et al., 2011; Espelid et al., 1991; Lund et al., 2008; Nymo et al., 2016). LPS induces a broad and evolutionary conserved B cell response that does not depend on the intricate processes of T-cell/B-cell interactions, specific antibody production, and antibody affinity maturation (AM) (Futoma-Kołoch, 2016; Uchiyama, 1982). Specifically, the anti-LPS humoral response was detected during Atlantic cod infection with *Brucella pinnipedialis, Francisella noatunensis,* and *Vibrio salmonicida* (Ellingsen et al., 2011; Lund et al., 2006; Nymo et al., 2016). In line with these functional observations, the Atlantic cod's genome is unique in that it lacks several essential genes required for T-cell/B-cell interactions that initiate the antibody affinity maturation program in B cells. Notably absent from the Atlantic cod genome are major histocompatibility complex (*mhc*) class II, cluster of differentiation 4 (*cd4*; pseudogene), and invariant chain (*Ii*) genes. In contrast, its *mhc I* and some Toll-like receptor (*tlr*) loci are significantly expanded relative

to other vertebrates (Malmstrom et al., 2016; Parham, 2015, 2016; Solbakken, Rise, et al., 2016; Solbakken, Torresen, et al., 2016; Star et al., 2011; Torresen et al., 2017).

Taken together, the collective functional and genetic evidence is highly suggestive that the Atlantic cod humoral immune response lacks the process of AM, making it potentially less specific and robust than that of other studied bony fish. Since AID is a key initiator of antibody affinity maturation, we sought to examine its gene synteny and expression in Atlantic cod. Here, we report that, like other studied vertebrates, Atlantic cod *aicda* locus synteny has been conserved during Teleostei evolution. Moreover, our gene expression analyses show that Atlantic cod *aicda* is expressed in immune-related tissues, and its splenic expression is upregulated in response to immune stimulations. We also find two *aicda* transcript isoforms in Atlantic cod, one of which lacks the first exon resulting in predicted truncation of the first 21 amino acids and possibly loss of function, if translated. However, the translation of the full-length Atlantic cod AID transcripts divulged a drastic change in a conserved amino acid (Gm-AID^{H136}) that may compromise its enzymatic activity compared to other studied AIDs.

2.3 Methods

2.3.1 Synteny analysis of *aicda*

The aicda gene synteny was assessed both manually and using the synteny database. The 1-Mb regions containing *aicda* locus in Atlantic cod, three-spined stickleback (*Gasterosteus aculeatus*), Japanese pufferfish (*Takifugu rubripes*), zebrafish (*Danio rerio*), spotted gar (*Lepisosteus oculatus*), coelacanth (*Latimeria chalumnae*), green anole (*Anolis carolinensis*), chicken (*Gallus gallus*), mouse (*Mus musculus*), and human (*Homo spiens*) were derived using the assemblies from the Ensembl database (https://uswest.ensembl.org/index.html). In the case of the tropical clawed frog (*X. tropicalis*), the genomic region was retrieved from Xenbase database (http://www.xenbase.org/entry/). The annotated genes within this 1-Mb region were then manually inspected to obtain Figure 2-2. Additionally, using the synteny database (http://syntenydb.uoregon.edu/synteny_db/) (Catchen et al., 2009). The chromosomal location of zebrafish AID (*Dr-aicda*) was compared to that of Japanese pufferfish, three-spined stickleback, spotted gar, mouse, and human. Also, *Hs-aicda* synteny was compared to that of the mouse, spotted gar, and the tropical clawed frog.

2.3.2 Animals

All animal maintenance and sampling conducted for this study was approved by the Memorial University of Newfoundland's Institutional Animal Care Committee following the Canadian Council for Animal Care guidelines. Ten different families of passive integrated transponder-tagged Atlantic cod (juvenile life stage; ~ 60 g; ~ 30 fish per family) from the Atlantic cod Genomics and Broodstock Development Project (CGP)

year-class 3 (YC3) were transported to the Ocean Sciences Center of Memorial University of Newfoundland. Fish were obtained from the Huntsman Marine Science Center in St. Andrew's, New Brunswick in October of 2008 and kept in a 3000-L flow-through seawater tank at 10 °C and > 90 % oxygen saturation. During one month of acclimation, fish were fed to apparent satiation.

2.3.2.1 Immune stimulated spleen tissues

Samples used to investigate the *Gm-aicda* transcript response to the immune stimulation were collected for a previously published study (Hori et al., 2012; Hori et al., 2013). After one month of acclimation, the fish were divided between eight 500-L tanks at 10 °C and > 90 % oxygen saturation. Approximately equal numbers of fish from each family was transferred into each tank (~ 36 fish per tank). After two weeks of acclimation in the 500-L tanks, fish were intraperitoneally injected with polyinosinic-polycytidylic acid (pIC; a synthetic dsRNA viral mimic) or formalin-killed typical *A. salmonicida* (ASAL) in sterile phosphate-buffered saline (PBS). The control group was injected with PBS alone. Fish were sacrificed 6 or 24 hours post-injection (HPI) by submersion in an anesthetic bath containing tricaine methanesulfonate (MS-222, 400 mg L^{-1}, Syndel Laboratories, Canada). Spleen samples were collected in certified RNase-free 1.5 ml tubes, flash-frozen in liquid nitrogen, and stored at -80 °C.

2.3.2.2 Sampling for tissue panel experiment

The Atlantic cod used for tissue expression and developmental experiments were kept in a 21 m^3 flow-through tanks in the Dr. Joe Brown Aquatic Research Building (JBARB) of the Ocean Sciences Center (OSC, Memorial University of Newfoundland).

The tank provided the conditions of 5.2 to 6.4 °C, > 95 % oxygen saturation, and an ambient photoperiod. The fish (2.29 ± 0.42 kg [mean ± SE]) were fed a commercial diet (Skretting, BC, Canada; crude protein 50 %, crude fat 18 %, and crude fiber 1.5 %) three times per week at 1 % body weight per day.

To investigate the *Gm-aicda* tissue expression pattern, its mRNA expression was studied in 19 tissues extracted from four healthy-appearing individual adults (2 males and 2 females). The fish were not fed for 24 h before euthanizing with MS-222 (as described above). Dissection tools and surfaces were cleaned with RNase Away solution (Sigma). From each fish, 19 tissues were collected: blood, brain, eye, fin, gill, gonad, hindgut, midgut, heart, head kidney, posterior kidney, liver, dorsal muscle, ventral muscle, pyloric caecum, dorsal skin, ventral skin, spleen, and stomach. Samples were immediately flash-frozen in liquid nitrogen and kept at -80 °C until RNA extraction.

2.3.2.3 Sampling for developmental experiments

The Broodstock fish used in these experiments were kept in the same conditions as the tissue panel experiment except their diet. These fish were fed mackerel, herring, and squid diet supplemented with vitamins twice per week before and during the spawning season. To assess *Gm-aicda* transcript expression during embryogenesis and early larval development, a mixture of fertilized eggs and cleavage-stage embryos (1.4 L, 2-cell to 64-cell embryos) were automatically collected after communal spawning. The collected floating fertilized eggs (0-days post-fertilization [DPF], *i.e.*, day 0) were distributed into three 50-L conical incubator tanks (350 ml of eggs per tank). The tanks were kept at 5.5 to 6.1 °C, with a 25 L h^{-1} flow rate, gentle aeration, and under an ambient photoperiod (Rise

et al., 2012). Using 500 μm Nitex, a mixture of ~ 180 eggs/embryos (~ 0.5 ml of embryos or ~ 0.4 ml of larvae) were collected daily from each tank until the yolk-sac absorption stage (*i.e.*, day 20; before active feeding). Samples were immediately flash-frozen using liquid nitrogen and kept at -80 °C for RNA extraction. The developmental stage of embryos was also examined every day (Hall et al., 2004). The blastula/gastrula stages were observed from day 1 to 6. The segmentation period started on day 7, and the golden eye stage was noticed on day 12. On day 15, hatching began and completed for all embryos on day 18 (Eslamloo et al., 2019).

2.3.3 Macrophage isolation and immune stimulation

The immune stimulated Atlantic cod macrophage samples were used as the negative control for *aicda* expression experiments (Eslamloo et al., 2018; Eslamloo et al., 2016). The macrophage-like cells were isolated from the head kidneys of 5 healthy-appearing individual fish kept in the same condition as the ones used for the tissue panel experiments (Eslamloo et al., 2016). Throughout the experiment, the Leibovitz L-15 medium (Gibco, Carlsbad, CA) supplemented with 2 mM L-glutamine, 4.2 mM $NaHCO_3$, 25 mM HEPES, 1.8 mM glucose, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin (Gibco) and 1% fetal bovine serum (FBS, Gibco) was used (L-15^{+}). The blood was removed from the caudal vein of each fish after euthanizing with MS-222. The hematopoietic kidney (*i.e.*, head kidney) was then dissected out. The cell suspension in L-15^{+} culture medium was made by mincing the samples through a 100-μm nylon cell strainer (Fisherbrand™, Thermo Fisher Scientific, Waltham, MA, USA). The macrophage-enriched interface was collected after a centrifugation step on a discontinuous 25/51 % Percoll gradient (GE

Healthcare, Uppsala, Sweden) at 300 × g for 40 min at 4 °C. The isolated cells were then washed twice in L-15^{+} and centrifuged at 300 × g for 15 min at 4 °C. Following this step, cells were suspended in the L-15^{+} medium containing 1 % fetal bovine serum (FBS; Gibco) and without heparin. Viability of > 96 % was recorded for the isolated cells using a hemocytometer and a trypan blue (Sigma-Aldrich) exclusion test. These cells were then cultured in 6-well plates (Corning, Corning, NY) in L-15^{+} medium at the initial density of 3×10^{7} cells (in 2 ml of L-15^{+}) per well. After overnight incubation at 10 °C, the wells were washed 3 times with L-15^{+} to remove the non-adherent cells. 24 hours after harvesting, the cells were exposed to 50 µg ml^{-1} pIC (the stock solution was made in PBS [pH 7.2] at 10 mg ml^{-1} concentration) for 24 hours. At 24-hours post-stimulation (24 HPS), the media was removed, and 800 µl of TRIzol (Invitrogen, Burlington, ON) was added into each well to lyse the cells. The TRIzol-lysed cell suspensions were kept at -80 °C until RNA extraction.

2.3.4 Total RNA extraction and purification

The total RNA was extracted from flash-frozen samples (~ 100 mg of tissue samples) using TRIzol reagent following the manufacturer's protocol. Briefly, one ml of TRIzol was added to ~ 100 mg of tissue. To homogenize firm tissues (*i.e.*, eye, gill, heart, stomach, pyloric caecum, midgut, hindgut, dorsal skin, ventral skin, dorsal muscle, ventral muscle, and fin) ceramic mortars and pestles, baked at 220 °C for seven hours, were used, while disruption of other samples was accomplished using RNase-free disposable pellet pestles (Fisherbrand). Following sample disruption, the QIAshredder spin columns (QIAGEN, Mississauga, ON) were used to homogenize the sample according to the

manufacturer's protocol. For each sample, chloroform (0.2 ml) was then added to the collected supernatant, mixed, and incubated at room temperature for two to three minutes. After centrifuging the sample at 4 °C (15 min at 12000 × g), the aqueous phase was transferred into a new tube. Isopropanol (0.5 ml) was then mixed with the aqueous phase. After 10 min of incubation at room temperature, the sample was centrifuged for 10 min at 12000 × g and 4 °C. The RNA pellet was then washed using 75 % ethanol (1 ml). After centrifugation for 5 min at 7500 × g at 4 °C and removal of the supernatant, the RNA pellet was air-dried, then re-suspended in 100 µl of RNase/DNase free water (Gibco). Liver samples were re-purified through standard phenol-chloroform extraction and ethanol precipitation. To remove any genomic DNA contamination, 30 µg of each extracted RNA sample was treated with DNase-I (6.8 Kunitz U, RNase-free DNase Set, Qiagen, Valencia, CA) following the manufacturer's protocol. The RNA was purified from salts, proteins, and nucleotides using the RNeasy MinElute clean-up kit (Qiagen) according to the kit instructions. The quality and quantity of the purified RNA were measured using NanoDrop spectrophotometry (ND-1000), and the RNA integrity was assessed by 1 % agarose gel electrophoresis. RNA samples with A260/230 > 2, A260/280 > 1.8, and tight 18S and 28S rRNA bands were used for further analyses.

2.3.5 cDNA synthesis

cDNA synthesis was performed on 1 µg or 5 µg of clean total RNA using either SuperScript III Reverse Transcriptase (SuperScript III-RT, Invitrogen) or M-MLV Reverse Transcriptase (M-MLV RT, Invitrogen) as recommended by the manufacturer's manuals. Specifically, 1 or 5 µg of total clean RNA was reverse transcribed at 50 °C for 1 h using

SuperScript III RT (200 U) in a 20-µl reaction containing 250 ng random hexamer primers (Invitrogen), 1 µl of dNTPs (10 mM each), 1 × first stand buffer, 40 U of RNaseOUT, and 5 mM DTT. The same conditions were used for M-MLV RT (200 U) except that reactions were incubated at 37 °C for 50 min in the presence of 10 mM DTT. The cDNA was diluted 10 × using RNase/DNase free water.

2.3.6 Characterization of *Gm-aicda* transcript(s)

Based on AID expression pattern studied thus far, GCs are the main site of AID expressing B cells in mammals and birds (Bascove & Frippiat, 2010; Marr et al., 2007; Muramatsu et al., 1999; Muto et al., 2000; Ohmori et al., 2004; Saunders & Magor, 2004; Verma et al., 2010). Previous studies have reported the melano-macrophage clusters in the spleen of fish as the alternative to the canonical germinal centers in mammals and birds. We, therefore, used the pIC-stimulated splenic total RNA to characterize the possible AID transcript(s) in Atlantic cod (Agius & Roberts, 2003; Boehm et al., 2012; Saunders et al., 2010).

2.3.6.1 Preliminary validation of *Gm-aicda* transcript expression

To confirm the expression of *Gm-aicda* transcript(s), gene-specific primers (Table 2-1) were designed based on the predicted AID ORF sequence in the Atlantic cod genome project using Primer3web v4.0.0 (http://primer3.ut.ee/). SuperScript III-RT was used to synthesis first-strand cDNA from 1 µg of total RNA, as described in section 2.3.5. In a 25-µl PCR reaction, 1 µl of undiluted cDNA (equivalent to ~ 100 ng of initial total RNA) was amplified using 0.625 U of TopTaq DNA polymerase (QIAGEN), 0.2 µM of each primer, 0.2 mM of each dNTP, 1 × TopTaq PCR buffer, 1 × CoralLoad, and 1 × Q-solution. No-

template and no-RT reactions were included as well. Touchdown PCR cycling conditions were an initial denaturation step for 3 min at 94 °C followed by 35 cycles of [30 s at 94 °C; 30 s at 65 °C → 54.5 °C, decreasing 0.3 °C per cycle; and 1 min at 72 °C] and 10 min at 72 °C. After examining the PCR products on a 1.5 % agarose gel, the PCR band was gel extracted using the MinElute gel extraction kit (QIAGEN) following the manufacturer's instructions. The gel-extracted PCR products were then TA-cloned into the pCR 2.1-TOPO TA vector (TOPO TA Cloning Kit, Invitrogen, USA) as per the kit's recommended protocol. Briefly, in a 6-µl reaction, 3 µl of the extracted PCR band was mixed with 1 µl of the vector and 1 µl of the salt solution. Reactions were incubated at room temperature (22 to 23 °C) for 30 min. Performing chemical transformation protocol, 2 µl of the TOPO cloning reaction was transformed into One Shot TOP10 competent cells (chemically competent *E. coli*, Invitrogen) following the kit's instructions. After overnight culture of transformed bacteria at 37 °C, 6 white colonies were picked and cultured in 5 ml of Luria-Bertani (LB) broth medium containing 50 µg ml^{-1} ampicillin (for ~ 16 h at 37 °C and 225 rpm). The cultured colonies were then purified using the QIAprep spin miniprep kit (QIAGEN) as per the manufacture's protocol. The purified TA-cloned plasmid preparations were Sanger sequenced (Macrogen, South Korea).

2.3.6.2 Identification of the full-length *Gm-aicda* mRNA(s)

To obtain full-length mRNA, rapid amplification of cDNA ends (RACE) PCR was performed. Sequencing results from the previous step were used to design gene-specific RACE-PCR primers (Table 2-1) using Primer3web v4.0.0 (http://primer3.wi.mit.edu). Splenic RNA extracted from pIC stimulated fish (24 HPI) was used, and RACE-PCR was

carried out using the SMARTer RACE cDNA amplification kit (Clontech, Takara Bio Company, USA). To obtain 3′/5′-RACE-Ready cDNA, 1 µg of cleaned RNA was reverse transcribed. The produced cDNA was then diluted 3 × in Tricine-EDTA buffer. For 3′-RACE and 5′-RACE PCR, 2.5 µl of diluted 3′ or 5′-RACE-Ready cDNA (equivalent to ~ 75 ng of initial RNA) was amplified in a 50-µl reaction containing 1 × Advantage 2 polymerase mix (Clontech), 1 × Advantage 2 PCR buffer, 0.2 mM of each dNTPs, 0.2 µM of gene-specific primers, and 0.2 µM of the Universal Primer A mix. A touch-down PCR program of 1 min at 95 °C; 5 cycles of (94 °C for 30 s, 72 °C for 3 min); 5 cycles of (94 °C for 30 s, 70 °C for 30 s, 72 °C for 3 min); 25 cycles of (94 °C for 30 s, 68 °C for 30 s, 72 °C for 3 min); and a final extension cycle of 72 °C for 10 min was conducted. These primary PCR products were then gel extracted using the MinElute gel extraction kit. For nested 3′-RACE or 5′-RACE, 5 µl of 50 × diluted primary PCR product (~ 400 pg μl^{-1}) were re-amplified using the same conditions, except the Nested Universal Primer A mix, and nested gene-specific primers were used. The nested PCR program consists of 1 min at 95 °C, 25 cycles of [30 sec at 94 °C; 30 sec at 68 °C; 3 min at 72 °C], and 10 min at 72 °C. PCR bands were then gel extracted and sequenced as described above.

Sequencing data were assembled and analyzed using Lasergene 7 MegAlign software (DNASTAR, Inc., USA). The ATGpr website (https://atgpr.dbcls.jp/cgi-bin/atgpr.cgi) was used to identify the initiation codon, coding sequence (CDS), and the stop codon. The CDS with the highest reliability score was reported.

To confirm the presence of two *Gm-aicda* transcripts, nested RT-PCR was performed on splenic RNA extracted from 11 pIC-stimulated fish using the manually

designed isoform-specific primers (ISPs, Table 2-1). Using the SuperScript III-RT kit, 1 µg of clean total RNA was reverse transcribed as per section 2.3.5. In a 25-µl reaction, the primary PCR was performed using 2.5 µl of 10 × diluted cDNA of pIC stimulated spleen samples (equivalent to 25 ng initial RNA), ISPs (0.2 µM), and TopTaq DNA polymerase (0.625 U per reaction) following the manufacturer's recommended protocol. In the second round of PCR, 2.5 µl of the first-round PCR reaction was further amplified in the same reaction condition as the first PCR except the nested ISPs were used. For the full-length *Gm-aicda* (*Gm-aicda*) isoform, both first and nested PCR reactions were incubated at 94 °C for 3 min, followed by 10 cycles of [94 °C for 30 sec; 55 °C → 50 °C for 30 sec, decreasing 0.5 °C per cycle; 72 °C for 90 sec] and 25 cycles of [94 °C for 30 sec; 50 °C for 30 sec; 72 °C for 90 sec] and 72 °C for 10 min. For truncated *Gm-aicda* (*T- Gm-aicda*), 53 °C was used as the initial annealing temperature. PCR products were gel extracted, TA-cloned, and 10 colonies for each spleen sample and isoform were sequenced as detailed in the previous paragraphs.

2.3.7 Delineation of *Gm-aicda* transcripts expression in adult tissues, embryonic, and early larval life stages

The transcript expression of the elongation factor 1-α (*ef1-α*) was studied alongside *Gm-aicda* isoforms as a normalizer gene (Inkpen et al., 2015). In these experiments, we also used splenic cDNA of immune challenged individual Atlantic cod (24 HPI) as a positive control for *aicda* transcript expression. Due to almost exclusive expression of *aicda* in activated B cells, RNA obtained from immune stimulated Atlantic cod macrophages (24 HPS, pIC) was used as a negative control (Eslamloo et al., 2016).

To investigate the *Gm-aicda* tissue expression pattern, 19 tissues from 4 healthy adult Atlantic cod (two males: 758 and 1260 gr; two females: 1520 and 890 gr) were extracted as described in 2.3.2.2 section. To assess *Gm-aicda* transcripts expression during embryogenesis and early larval development, a mixture of fertilized eggs and cleavage-stage embryos were collected after communal spawning and distributed into three incubators. A mixture of ~ 180 eggs/embryos was collected daily from each tank until embryos reach the yolk-sac absorption stage (section 2.3.2.3).

In both experiments, total RNA was extracted and cleaned as per section 2.3.4, and the cDNA was synthesized using 5 µg of clean total RNA and M-MLV kit (refer to section 2.3.5). Using TopTaq DNA polymerase kit, 2 µl of 10 × diluted cDNA was amplified in a 25-µl reaction containing TopTaq DNA polymerase (0.625 U), 1 × TopTaq PCR buffer, 1 × CoralLoad, and 1 × Q-solution, 0.2 mM of each dNTP, and 0.2 µM of *Gm-aicda* ISPs or *ef1-α* primers (Table 2-1). PCR cycling conditions were an initial denaturation step for 5 min at 94 °C followed by 35 cycles of [30 sec at 94 °C; 30 sec at 54 °C; and 30 sec at 72 °C] and 5 min at 72 °C. Amplicons were then visualized on 2.5 % agarose gel.

2.3.8 Immune responsiveness of *Gm-aicda* transcript levels

To measure the changes in *aicda* transcription in response to immune stimulation, reverse transcription – fluorescence-based quantitative real-time PCR (RT-qPCR) was performed. The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines was followed to conduct, analyze, and report the RT-qPCR results. In these series of experiments, splenic clean RNA extracted from pIC, ASAL, and PBS treated fish were used (6 HPI and 24 HPI; 10 fish per treatment, section

3.2.1) (Hori et al., 2012; Hori et al., 2013). The flash-frozen tissues were stored in -80 °C for 3 years. For these experiments, the total RNA was isolated, DNase treated, and cleaned up from each frozen sample as detailed in section 2.3.4. M-MLV RT was used to synthesize cDNA from 5 µg of the clean total RNA (section 2.3.5). cDNA was stored at -20 °C and only thawed twice.

Prior to the qPCR assays, primer quality control was conducted using the splenic cDNA pool of pIC and ASAL stimulated samples. A 5-point and 3 × dilution standard curve of cDNA (starting from 10 ng of input RNA) was used to test the quality and efficiency of primer pairs (Table 2-1). Three fish per treatment and time point were used to select normalizers with stable expression. Two different sets of ISPs and four sets of normalizer primers were tested. The same ISPs, as described above, along with gene-specific primers for 60S acidic ribosomal protein P1 (*rplp1*) (Eslamloo et al., 2016) and ATP synthase H^+ transporting, mitochondrial Fo complex, subunit F2 (*atps*) (Hori et al., 2012), were qualified for qPCR analysis (Table 2-1). Two microliters of 10 × diluted cDNA (10 ng input RNA) were amplified in a 13-µl reaction containing 6.5 µl of Power SYBR Green master mix (Applied Biosystems), and 0.52 µl of each primer (1.25 µM). Q-PCR was carried out on a ViiA7 System (Applied Biosystems, Burlington, Ontario). Cycling conditions were one cycle of [2 min at 50 °C; 10 min at 95 °C], 40 cycles of [15 sec at 95 °C; 30 sec at 55 °C; 1 min at 60 °C]. The dissociation curves were created to confirm the homogeneity of the PCR products. The qPCR assays were performed in 384-well plates, and consistency of the assays between plates was checked using linker samples (C_T values

were < 1 cycle between plates). All the samples, linkers, and no-template controls were carried out in triplicate.

To analyze the q-PCR results, ViiA 7 Software v1.2 was used. The expression of *Gm-aicda* isoforms (C_T values) was normalized to the expression level of *rplp1* and *atps*, with the incorporation of amplification efficiency of primer pair. Then, the relative quantity (RQ) of each transcript was calculated using a calibrator sample. For each transcript, the lowest expression sample was considered as the calibrator (RQ set as 1). Statistical analysis was conducted using IBM SPSS Statistics 20 software. The expression of *Gm-aicda* isoforms at each immune stimulated condition was compared to that of PBS injected control using a nonparametric T-test for independent samples.

Table 2-1: The sequence of primers used in this chapter

Gene		Direction	Primer sequence (5' to 3')	Amplification efficiency (%)	R^2	Amplicon size (bp)	Application
Activation induced cytidine deaminase (aicda); *Gm-aicda*	Set 1	Forward	TAGTAAGCTAGACAGTGTGCTCTTGG	NA	NA	608	Detecting *Gm-aicda* ORF
		Reverse	CATCTCTTAAATCTTCTGTTTCACATGG				
	Set 2	Forward	CTCTGCTTCGTAGTAAAGAGAAGGC	NA	NA	473	
		Reverse	AGTTTTCTTGACAGACGCACATAATTGG				
Gm-aicda	First PCR	Forward	GACTTCGGACACCTACGCAATCGCACTGGC	NA	NA	NA	3' RACE-PCR
		Reverse	CCTCAGGTCCCTCAAGCCCTCTACATGCGG				5' RACE-PCR
	Nested PCR	Forward	CGCAATCGCACTGGCTGCCACGCAGAGCTG	NA	NA	NA	3' RACE-PCR
		Reverse	GCCCTCTACATGCGGACTGCCCTCCAGGTC				5' RACE-PCR
Gm-aicda	First PCR	Forward	GACTTTCAAAATGATTAGTAAGCTAGACAG	NA	NA	780[i]	Confirming *Gm-aicda* isoforms
T-Gm-aicda		Forward	GAATGGTTGATGATTACAGACCC			822	
Gm-aicda -3'UTR-r1		Reverse	TTGGACTACATAGGCGGTTTCAC				
Gm-aicda	Nested PCR	Forward	TAAGCTAGACAGTGTGCTCTTGG	NA	NA	747[ii]	
T-Gm-aicda		Forward	GATTACAGACCCTTACCGCAG			799	
Gm-aicda-3'UTR-r1		Reverse	GGTTTCACAAAGTTCTACAGTTTGC				
Eukaryotic translation elongation factor 1 alpha (*ef1-α*)[iii]		Forward	CCCTCCAGGACGTCTACAAG	NA	NA	150	Tissue and developmental panel (normalizer)
		Reverse	GAGACTCGTGGTGCATCTCA				

Gene	Direction	Primer sequence (5' to 3')	Amplification efficiency (%)	R^2	Amplicon size (bp)	Application
Gm-aicda	Forward	AGTAAGCTAGACAGTGTGCTC	101.57	0.989	125	Tissue and developmental panel; qPCR
	Reverse	CAGGTCCAAGCCTTCTCTT				
T-Gm-aicda	Forward	TTCTCTCCTATGTCTCAGTGTGC	100.47	0.989	133	
	Reverse	GGAATCAGGTCCAAGCCTTC				
60S acidic ribosomal protein P1 (*rplp1*)[iii]	Forward	TCTGAAGCTAAGGCCCTCAA	104.8	0.998	141	qPCR (normalizers)
	Reverse	ATCGTCGTGGAGGATCAGAG				
ATP synthase H+ transporting, mitochondrial Fo complex, subunit F2 (*atps*)[iv]	Forward	ACATGGATAAATGGCTTTTTGC	99.43	0.994	155	
	Reverse	TTGAAGAAGTAGTGTGGCTGGA				

[i]: If used with *Gm-aicda*-3'UTR-r1
[ii]: If used with *Gm-aicda*-3'UTR-r2
[iii]: The primer sequences for these genes were previously published in Inkpen *et al.*, (2015)
[iv]: The primer sequences for these genes were previously published in Hori *et al.,* (2012)

2.3.9 Protein Structure prediction

Five APOBEC structures and a partial near-native AID structure were chosen as templates for homology modeling (Table 2-2) (Bohn et al., 2013; Byeon et al., 2013; Hayashi, 2009; Holden et al., 2008; Kitamura et al., 2012; Qiao et al., 2017). The template AID/APOBEC structures were obtained from the protein databank (http://www.rcsb.org) and visualized using PyMOL v1.7.6 (http://www.pymol.org/). The computational homology modeling of each AID homologs was done using the default parameters of I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy et al., 2010; Yang et al., 2015; Zhang, 2008). Ramachandran plots were created using Rampage and used to evaluate the quality of the proteins on an individual residue basis based on their stereochemical angles (Lovell et al., 2003). The catalytic pocket was defined by the indented space containing the Zn-coordinating and catalytic residues (Hs-AID: H56, E58, C87, and C90; Dr-AID: H60, E62, C99, and C102; Ip-AID: H59, E61, C98, and C101; Gm-AID: H60, E62, C100, and C103). The catalytically accessible models were defined by the accessibility of catalytic glutamate to the surface of the protein. The pKa values were calculated using PDB2PQR (http://apbs-rest-test.westus2.cloudapp.azure.com/pdb2pqr or http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0/) (Dolinsky et al., 2004; Olsson et al., 2011).

Table 2-2: APOBEC and AID structures used as templates for homology modeling

Species	AID/APOBEC	Method	PDB ID
Mouse	APOBEC2	NMR	2RPZ
Human	APOBEC3A	NMR	2M65
Human	APOBEC3C	X-ray	3VOW
Human	APOBEC3F-CTD	X-ray	4IOU
Human	APOBEC3G-CTD	X-ray	3E1U
Human	AID	X-ray	5W1C, 5W0R, 5W0U, and 5W0Z

2.4 Results

2.4.1 Genomic features of Atlantic cod *aicda* locus

Annotation of the Atlantic cod genome project revealed a putative *aicda* gene with a 5-exon genomic structure (Star et al., 2011; Torresen et al., 2017). Figure 2-1 and Table 2-3 illustrate the Atlantic cod *aicda* locus structure in comparison with other species. Previous studies, as well as our analysis of available sequencing data on NCBI and Ensembl genome browser 89, revealed that this genomic structure is conserved in all studied species except African clawed frog and tropical clawed frog in which exon 2 and 3 are fused (Bascove & Frippiat, 2010). Based on the *aicda* genomic structure in Atlantic cod, the predicted five exons are 20, 166, 283, 116, and 54 bp in length and make up a 642-nucleotide coding sequence (CDS) encoding a 213-aa protein. The size of the introns is reported as 412, 206, 2080, and 146 bp; however, the third intron is not fully sequenced, and our attempts to sequence this intron were unsuccessful as well.

To assess the conservation of the *Gm-aicda* chromosomal location in comparison with other vertebrates, we performed gene synteny analysis. We observed that *aicda* has a similar synteny within Teleostei and Mammalia (Figure 2-2 and Figure 2-3). Table 2-4 illustrates the regions which were used to generate these analyses.

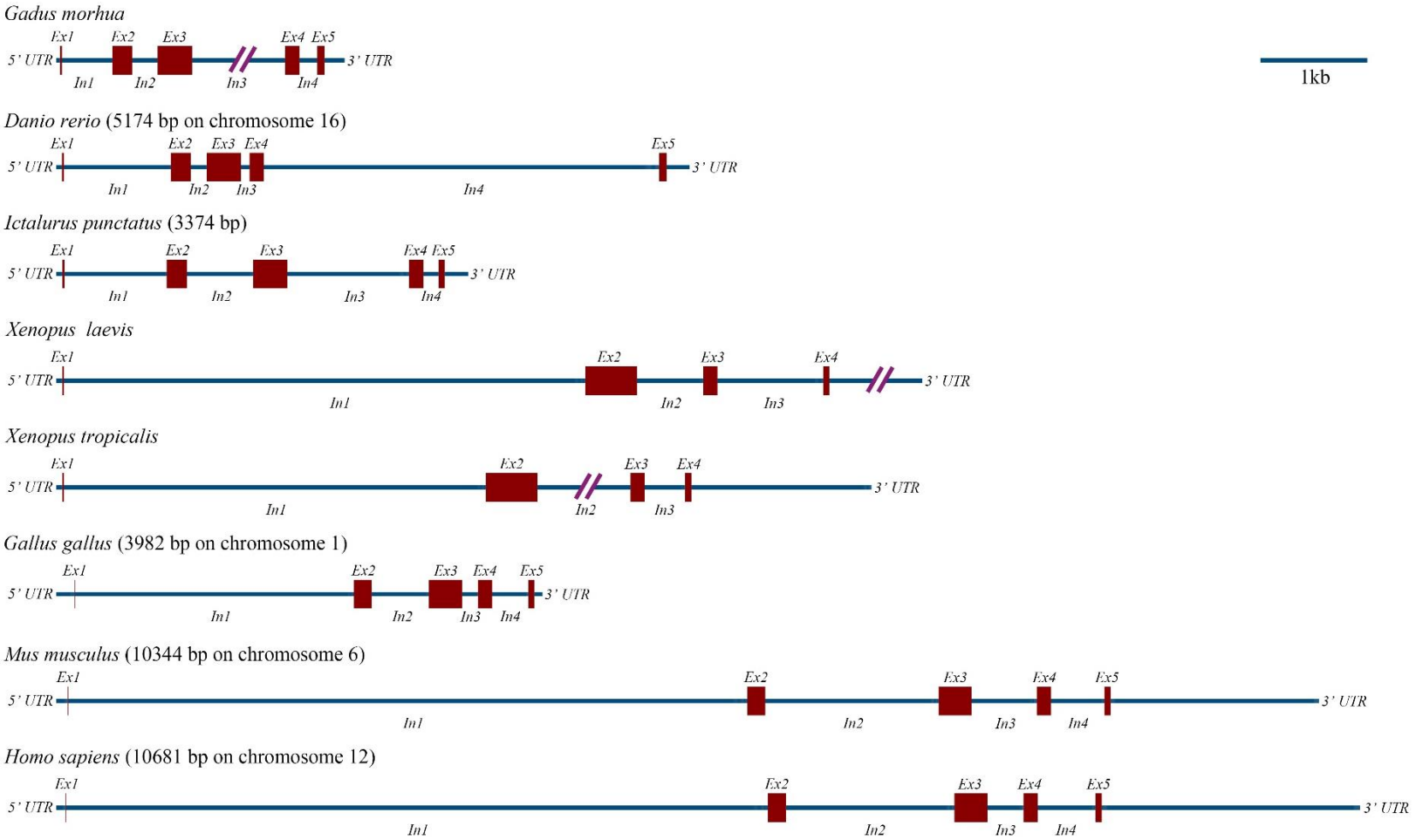


Figure 2-1: Comparison of the aicda genomic structure amongst vertebrates. Proportional schematic of the aicda locus exon-intron structure. Exons and introns are shown as red boxes and blue lines, respectively. Discontinued lines represent introns or untranslated regions (UTRs) that are not fully sequenced.

Table 2-3: Comparison of aicda locus amongst different species

	Exons (bp)					Introns (bp)				UTRs (bp)	
	1	2	3	4	5[a]	1	2	3	4	5'	3'[b]
Dr-aicda	20	166	280	116	51	870	129	73	3235	44	190
Gm-aicda	20	166	283	116	57	412	206	-	146	27	162
T-Gm-aicda	NA	123	283	116	57	NA	206	-	146	151	162
Ip-aicda	17	166	280	116	51	839	545	998	122	51	189
Xl-aicda[c]	14	422	116	54[a]	NA	4269	545	866	NA	-	-
Xt-aicda[c]	14	422	116	54[a]	NA	3452	-	332	NA	-	1475
Gg-aicda	8	148	271	116	54	2282	464	128	292	155	64[d]
Mm-aicda	8	148	271	116	54	5561	1422	529	436	93	1706
Hs-aicda	8	148	271	116	54	5747	1379	292	469	79	2118

[a]: including stop codon
[b]: excluding poly-A tail
[c]: exon 2 and 3 are fused together in these species
[d]: no poly-A tail is reported in mRNA sequence
-: no sequencing data available

Table 2-4: Genomic regions used in synteny analysis

Species	Database	Gene ID	Location	Region shown (bp)
Gadus morhua	Ensemble	ENSGMOG00000004114.1	GeneScaffold_1960: 226,520-229,999-forward strand gadMor1:HE567552.1	1-728260
Gasterosteus aculeatus	Ensemble	ENSGACG00000010521.1	groupXX: 12,050,972-12,052,426-forward strand	11551699-12551699
Takifugu rubripes	Ensemble	ENSTRUG00000007079.2	Primary_assembly 7: 13,080,121-13,081,769-forward strand FUGU5:HE602541.1	12580945-13580945
Danio rerio	Ensemble	ENSDARG00000015734.9	Chromosome 16: 12,660,477-12,665,652-forward strand GRCz11:CM002900.2	12163064-13163064
Lepisosteus oculatus	Ensemble	ENSLOCG00000008158.1	Chromosome LG26: 13,213,594-13,215,049-reverse strand LepOcu1:CM001429.1	12714321-13714321
Latimeria chalumnae	Ensemble	ENSLACG00000009320.1	Scaffold JH127875.1: 411,538-412,095-reverse strand	1-655812
Xenopus tropicalis	Xenbase	XM_002941202.4	Chr07:662023-669842-forward strand	165932-1165932
Anolis carolinensis	Ensemble	ENSACAG00000017441.2	Chromosome 2: 81,518,110-81,535,131-forward strand AnoCar2.0:CM000938.1	81026620-82026620
Gallus gallus	Ensemble	ENSGALG00000014280.6	Chromosome 1: 75,632,084-75,637,754-reverse strand GRCg6a:CM000093.5	75134919-76134919
Mus musculus	Ensemble	ENSMUSG00000040627.14	Chromosome 6: 122,553,801-122,564,180-forward strand GRCm38:CM000999.2	122043990-123043990
Homo sapiens	Ensemble	ENSG00000111732.11	Chromosome 12: 8,602,170-8,612,867-reverse strand GRCh38:CM000674.2	8107518-9107518

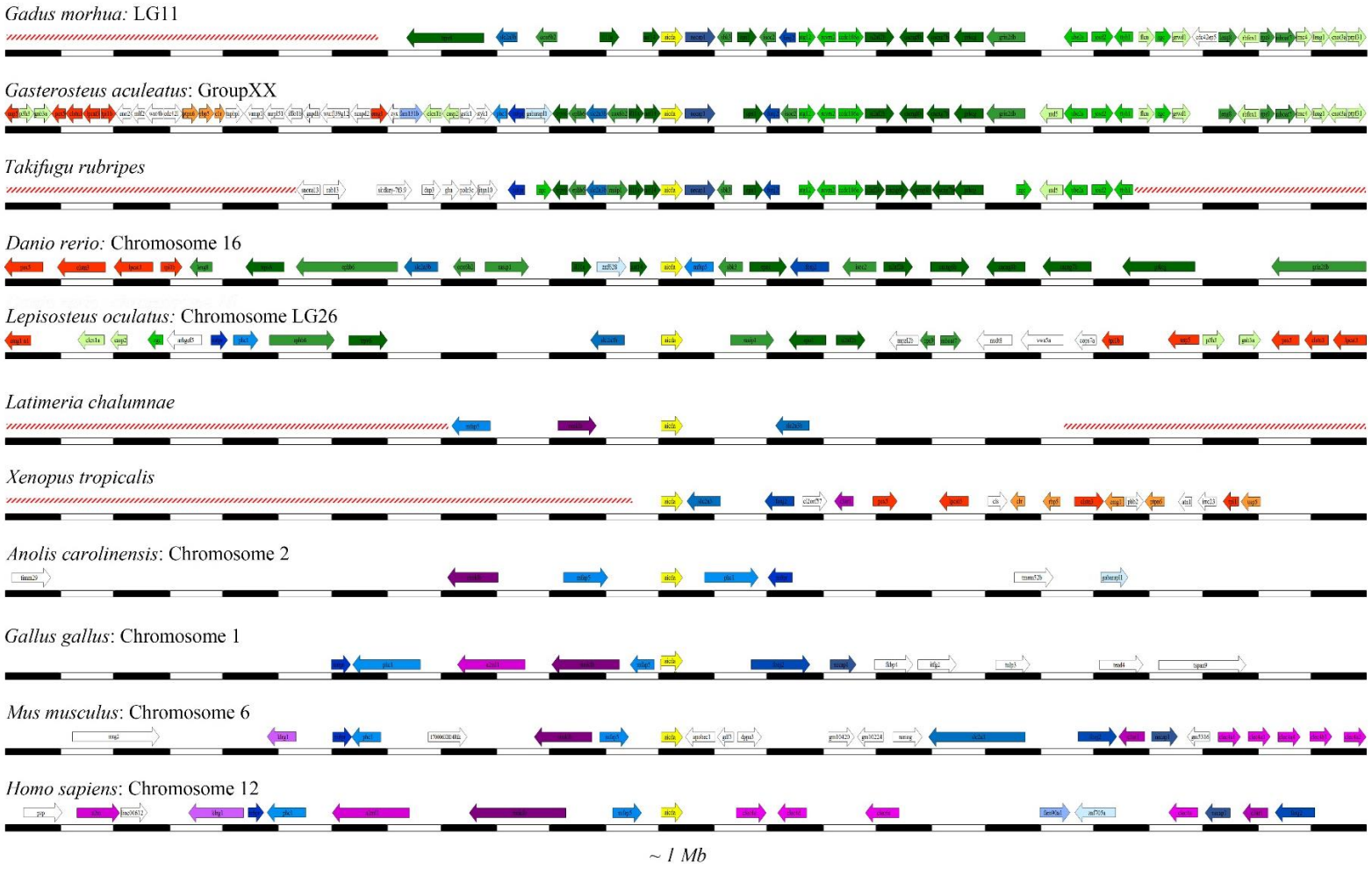


Figure 2-2: Comparison of the aicda synteny amongst vertebrates. Approximately 1 Mb region surrounding the aicda locus (colored in yellow) was retrieved from Ensembl genome browser 89. Red diagonal striped lines represent regions of genomic DNA with no sequencing data available. Genes conserved in all vertebrates, or only in tetrapods, or in bony fish are colored blue, violet, or green, respectively. Genes colored different shades of orange represent those found in selected bony fish and amphibian species.

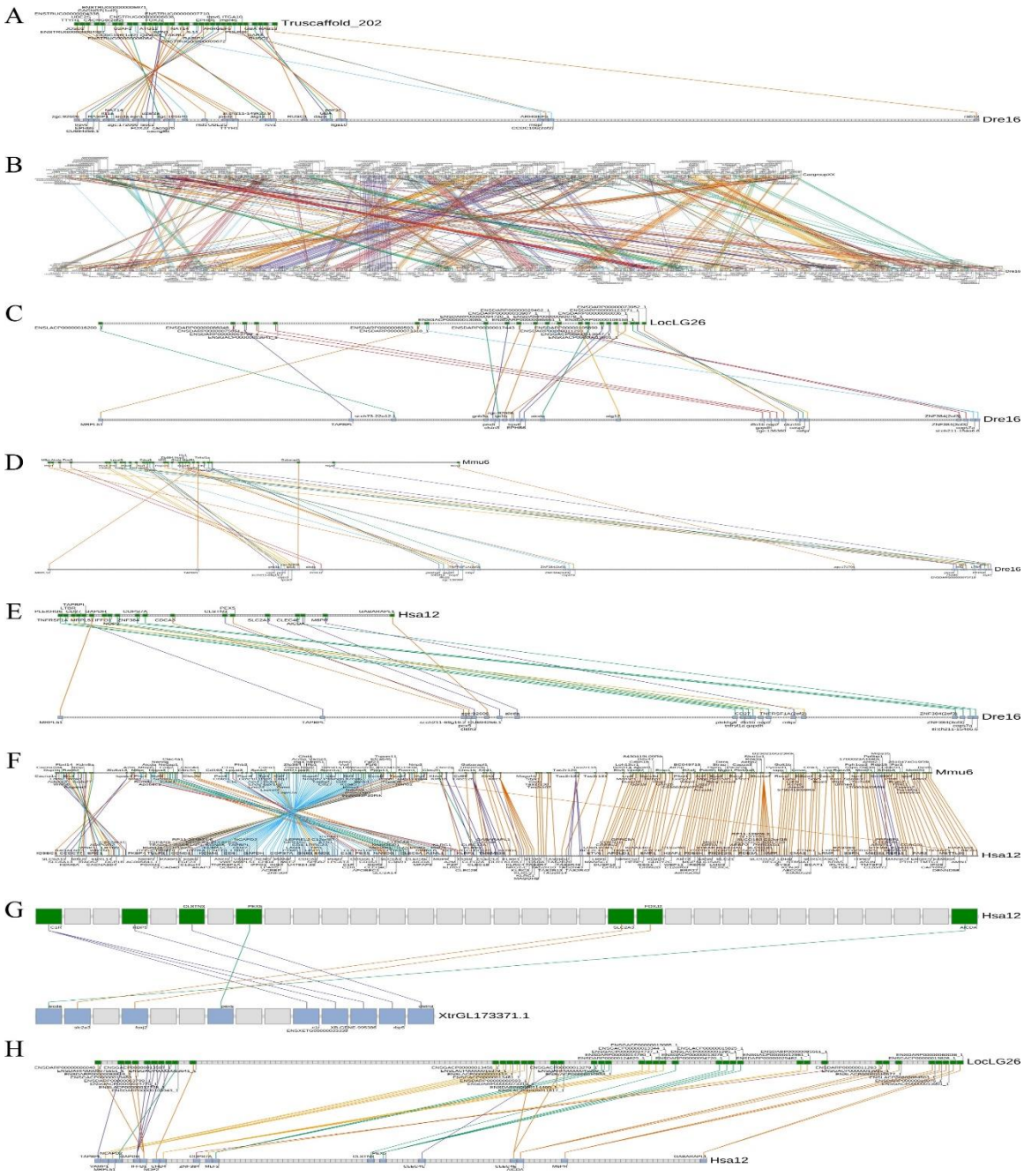


Figure 2-3: Aicda gene synteny. Aicda synteny analysis was performed using a synteny database based on Ensembl version 70 dataset. Dr-aicda chromosomal location was compared to that of the Japanese pufferfish (panel A), the three-spined stickleback (panel B), the spotted gar (panel C), mouse (panel D), and human (panel E). Also, Hs-AID synteny was compared to that of the mouse (panel F) and spotted gar (panel G), and the tropical clawed frog aicda synteny was compared with the human (panel H). Results showed a conserved micro-synteny across the vertebrate class.

2.4.2 *Aicda* transcript(s) expressed in adult Atlantic cod immune tissues

To confirm the expression of *aicda* gene in Atlantic cod, two sets of gene-specific primers (GSP) were designed based on the predicted *aicda* gene in the Atlantic cod genome project. Using these GSPs, RT-PCR was performed to detect the putative Atlantic cod *aicda* transcript(s) in splenic RNA samples extracted from pIC immune stimulated individuals (Figure 2-4 A). Sequencing confirmed a 473-nt fragment spanning position 97-570 of the predicted Atlantic cod *aicda* gene.

To obtain the full-length mRNA, rapid amplification of cDNA ends (RACE) nested PCR was conducted using primers designed based on the aforementioned transcript sequence (Figure 2-4 B). Assembly of RACE-PCR sequencing revealed two distinct *aicda* transcripts. One transcript of 830-bp contains all five predicted exons and encodes for a full-length 642-bp ORF. The other transcript is 892 bp long and lacks the first exon encoding for a truncated 579-bp ORF (Figure 2-4 C).

The full-length and truncated versions, hereafter respectively referred to as *Gm-aicda* (encodes Gm-AID) and *T-Gm-aicda* (potentially encodes T-Gm-AID), share the same 162-bp untranslated region at their 3′ end (*i.e.*, 3′-UTR) in which the polyadenylation signal (AAUAAA) is observed 13 bp upstream of the poly-A tail. However, the two transcripts differ in their 5′-UTR where a 27-bp and a 151-bp precede the ATG start codon in the *Gm-aicda* and *T-Gm-aicda* transcripts, respectively (Figure 2-5 and Table 2-5). Comparison of the *Gm-aicda* genomic region and the identified transcripts showed different transcription start site utilization among the two transcripts resulting in the absence of the first exon in the *T-Gm-aicda* isoform. Moreover, assessment of the exon-

intron boundaries revealed conserved sequences on these junctions in Atlantic cod compared with other vertebrate species (Figure 2-6). To further confirm the expression of both *Gm-aicda* transcripts, isoform-specific primers (ISP) were designed. PCR amplification and sequencing confirmed that both transcripts were indeed present in splenic cDNA of 11 Atlantic cod individuals (Figure 2-7).

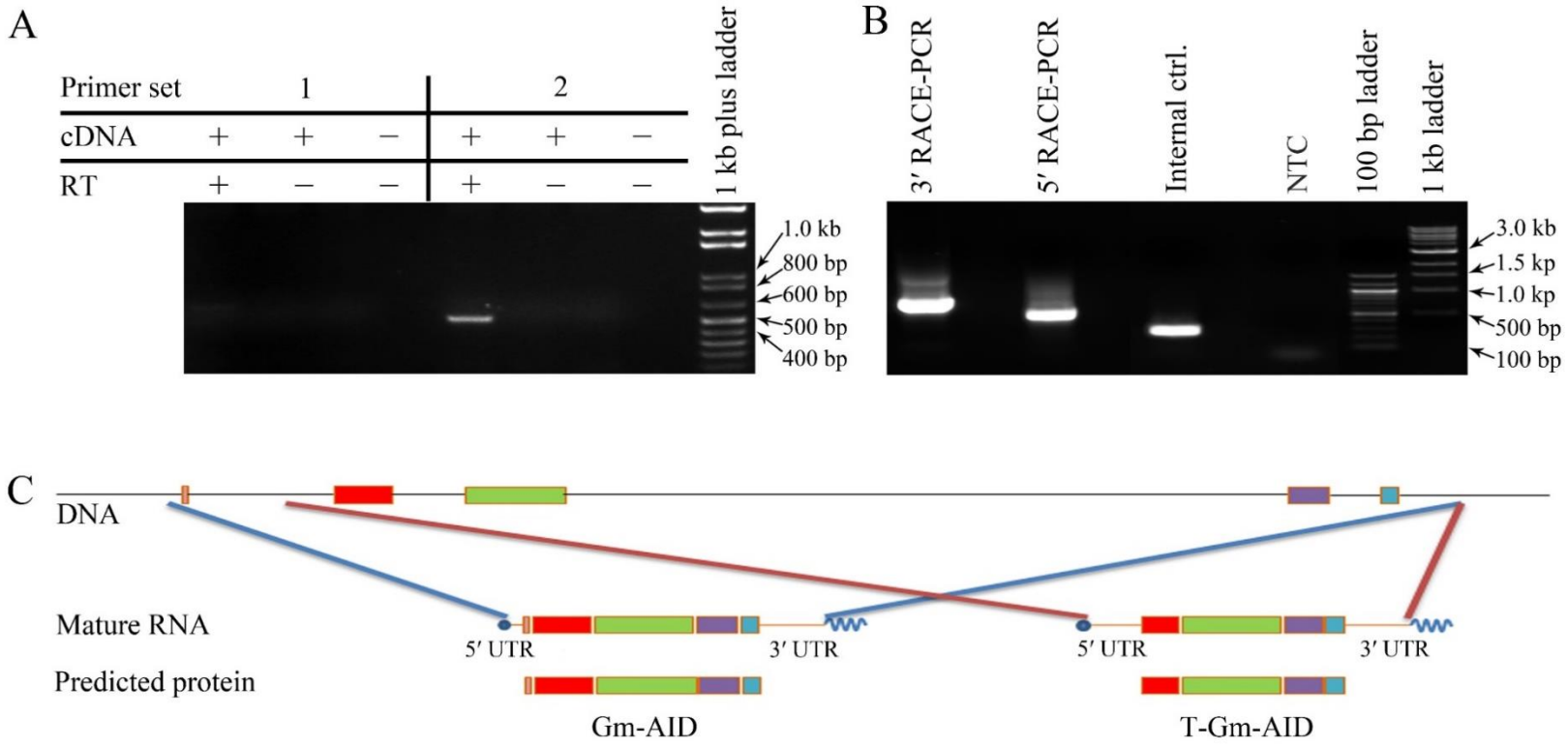
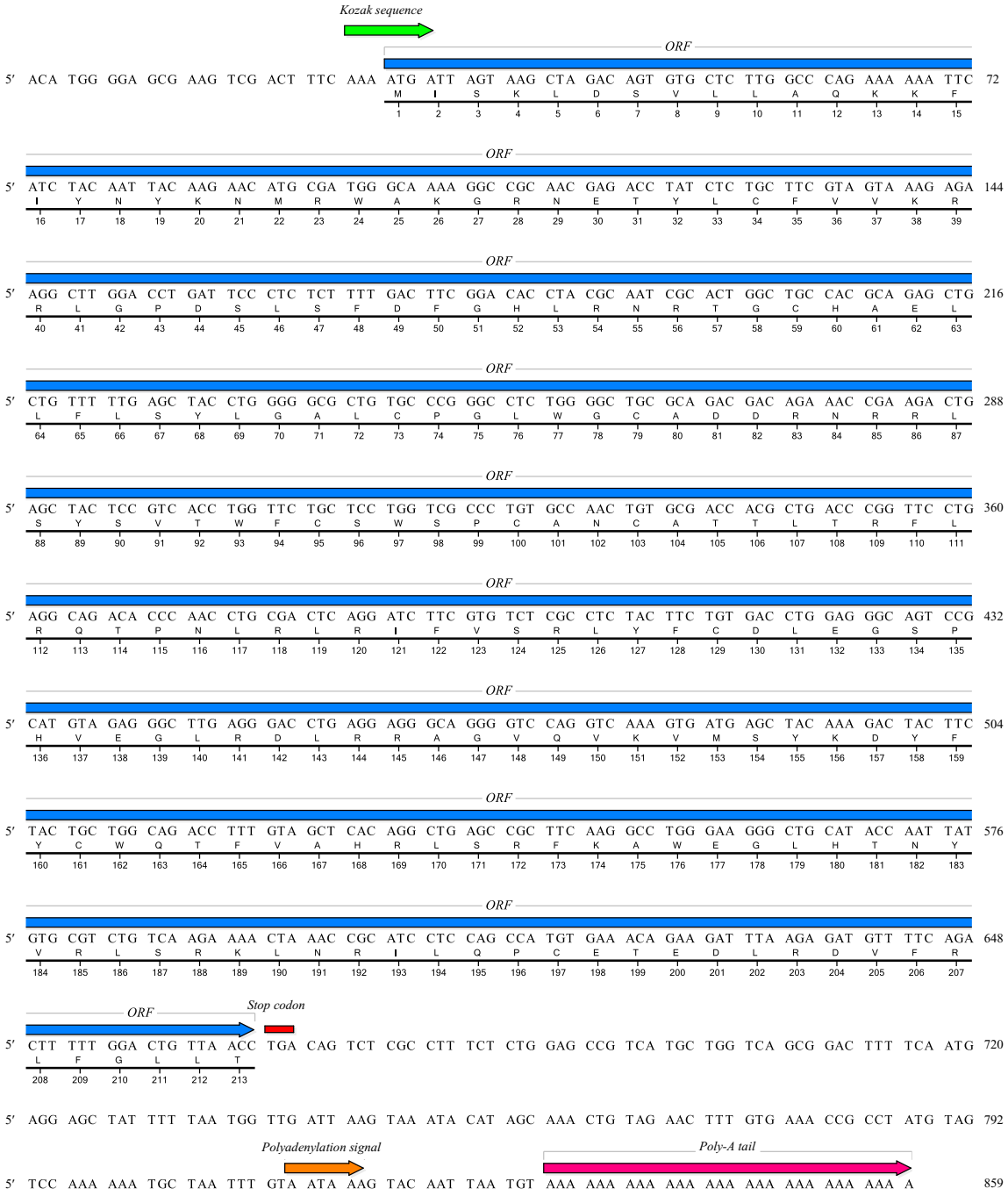


Figure 2-4: Identification and characterization of Atlantic cod aicda transcript(s). A) Amplification of partial aicda CDS using splenic total RNA and two sets of primers. The + and – refer to the presence and absence of each component in the PCR reaction, respectively. B) Amplification of full-length aicda mRNA(s) through RACE-PCR from splenic total RNA. C) Schematic representation of Atlantic cod aicda transcripts identified through RACE-PCR. Exons and introns are shown as boxes and lines, respectively.

A


Kozak sequence
ORF
Stop codon
Polyadenylation signal
Poly-A tail
5′ ACA TGG GGA GCG AAG TCG ACT TTC AAA ATG ATT AGT AAG CTA GAC AGT GTG CTC TTG GCC CAG AAA AAA TTC 72
M I S K L D S V L L A Q K K F
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
5′ ATC TAC AAT TAC AAG AAC ATG CGA TGG GCA AAA GGC CGC AAC GAG ACC TAT CTC TGC TTC GTA GTA AAG AGA 144
I Y N Y K N M R W A K G R N E T Y L C F V V K R
16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
5′ AGG CTT GGA CCT GAT TCC CTC TCT TTT GAC TTC GGA CAC CTA CGC AAT CGC ACT GGC TGC CAC GCA GAG CTG 216
R L G P D S L S F D F G H L R N R T G C H A E L
40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63
5′ CTG TTT TTG AGC TAC CTG GGG GCG CTG TGC CCG GGC CTC TGG GGC TGC GCA GAC GAC AGA AAC CGA AGA CTG 288
L F L S Y L G A L C P G L W G C A D D R N R R L
64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87
5′ AGC TAC TCC GTC ACC TGG TTC TGC TCC TGG TCG CCC TGT GCC AAC TGT GCG ACC ACG CTG ACC CGG TTC CTG 360
S Y S V T W F C S W S P C A N C A T T L T R F L
88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111
5′ AGG CAG ACA CCC AAC CTG CGA CTC AGG ATC TTC GTG TCT CGC CTC TAC TTC TGT GAC CTG GAG GGC AGT CCG 432
R Q T P N L R L R I F V S R L Y F C D L E G S P
112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135
5′ CAT GTA GAG GGC TTG AGG GAC CTG AGG AGG GCA GGG GTC CAG GTC AAA GTG ATG AGC TAC AAA GAC TAC TTC 504
H V E G L R D L R R A G V Q V K V M S Y K D Y F
136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159
5′ TAC TGC TGG CAG ACC TTT GTA GCT CAC AGG CTG AGC CGC TTC AAG GCC TGG GAA GGG CTG CAT ACC AAT TAT 576
Y C W Q T F V A H R L S R F K A W E G L H T N Y
160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183
5′ GTG CGT CTG TCA AGA AAA CTA AAC CGC ATC CTC CAG CCA TGT GAA ACA GAA GAT TTA AGA GAT GTT TTC AGA 648
V R L S R K L N R I L Q P C E T E D L R D V F R
184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207
5′ CTT TTT GGA CTG TTA ACC TGA CAG TCT CGC CTT TCT CTG GAG CCG TCA TGC TGG TCA GCG GAC TTT TCA ATG 720
L F G L L T
208 209 210 211 212 213
5′ AGG AGC TAT TTT TAA TGG TTG ATT AAG TAA ATA CAT AGC AAA CTG TAG AAC TTT GTG AAA CCG CCT ATG TAG 792
5′ TCC AAA AAA TGC TAA TTT GTA ATA AAG TAC AAT TAA TGT AAA AAA AAA AAA AAA AAA AAA AAA AAA A 859

B

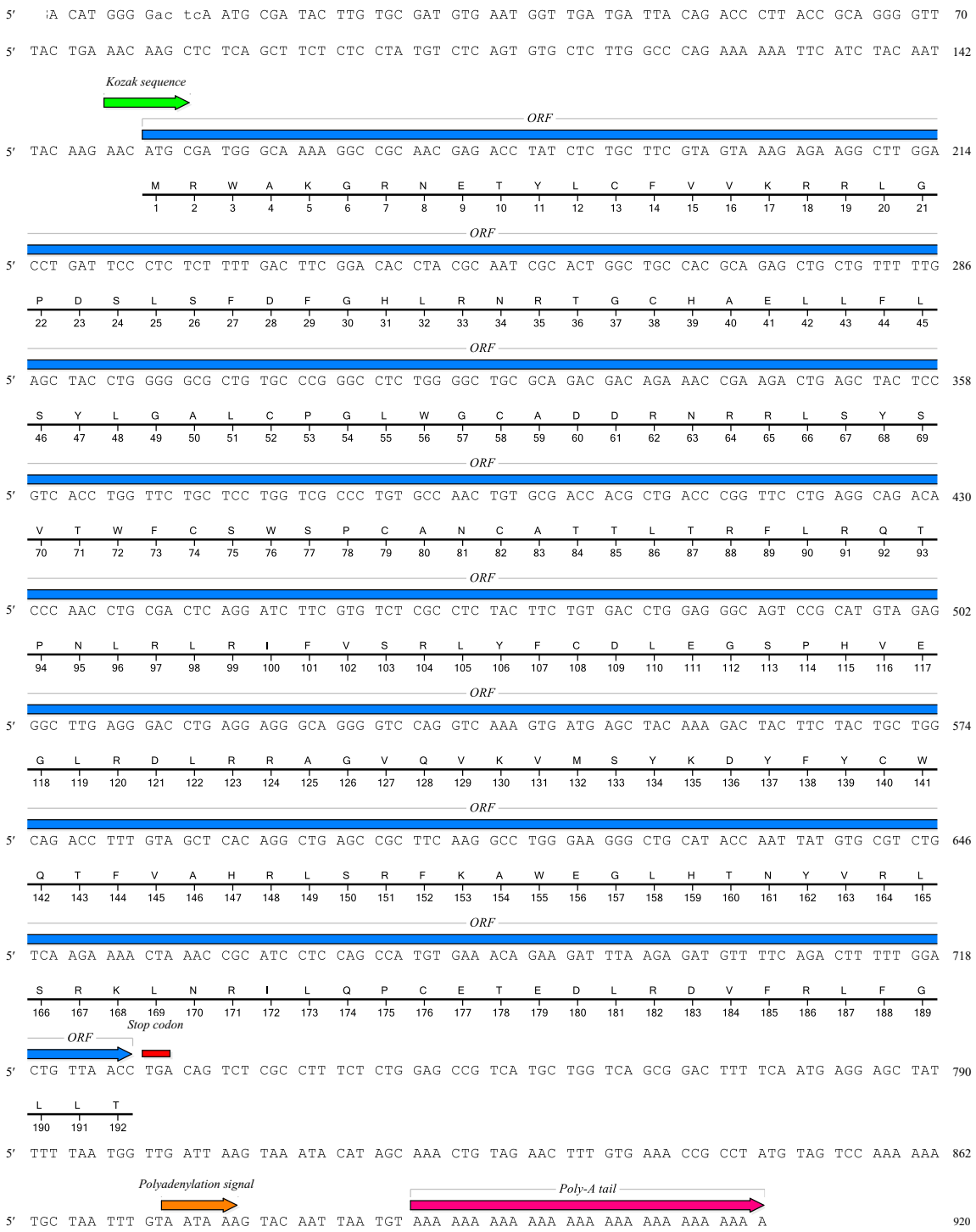


Figure 2-5: Sequence of the identified Atlantic cod aicda mRNA transcripts. Analyses of the sequencing data revealed two mRNA transcripts encoding a full-length aicda (A) and a truncated isoform (B).

Table 2-5: Characteristics of identified aicda transcripts predicted by ATGpr website

Isoform	# of ATG from 5′ end	Reliability score	Identity to Kozak rule A/GXXATGG	Start (bp)	Finish (bp)	ORF length (aa)	Stop codon found?	Protein sequence
Gm-aicda	2	0.35	AXXATGa	28	666	213	Yes	MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHAELLFLSYLGALCPGLWGCADDRNRRLSYSVTWFCSWSPCANCATTLTRFLRQTPNLRLRIFVSRLYFCDLEGSPHVEGLRDLRRAGVQVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNYVRLSRKLNRILQPCETEDLRDVFRLFGLLT
T-Gm-aicda	7	0.23	AXXATGc	152	727	192	Yes	MRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHAELLFLSYLGALCPGLWGCADDRNRRLSYSVTWFCSWSPCANCATTLTRFLRQTPNLRLRIFVSRLYFCDLEGSPHVEGLRDLRRAGVQVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNYVRLSRKLNRILQPCETEDLRDVFRLFGLLT

Abbreviations: *Gm-aicda*: Atlantic cod *aicda*; *T-Gm-aicda*: Atlantic cod truncated *aicda* isoform.

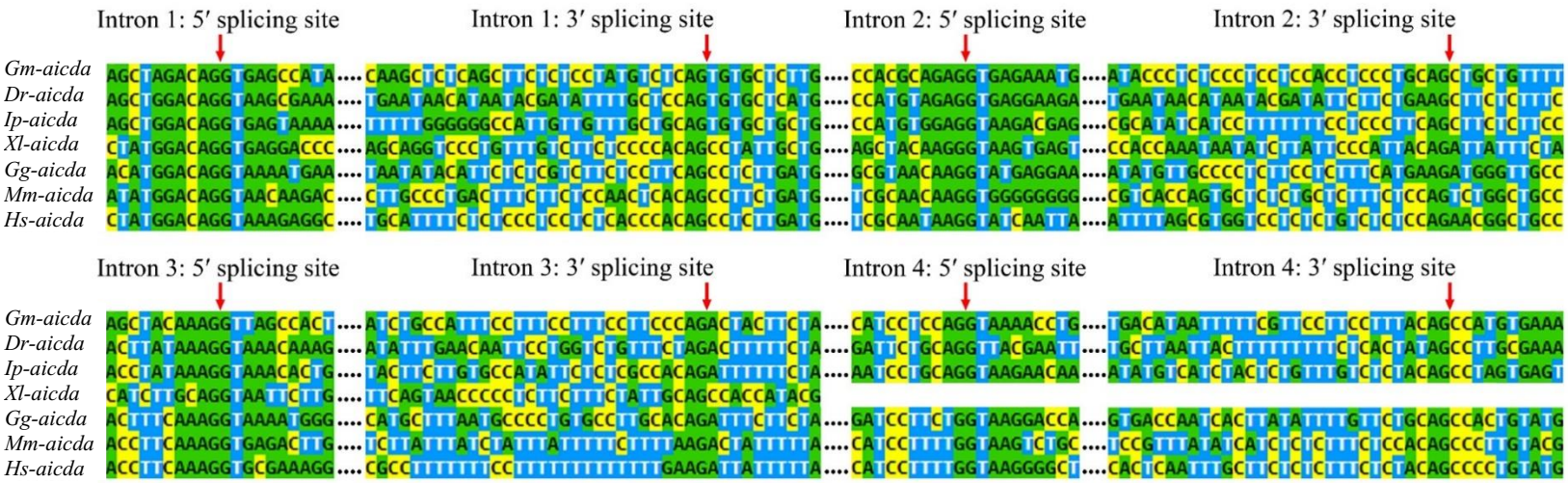


Figure 2-6: Alignment of splicing sites of aicda transcripts in different species. The red arrows show the exon-intron boundaries.

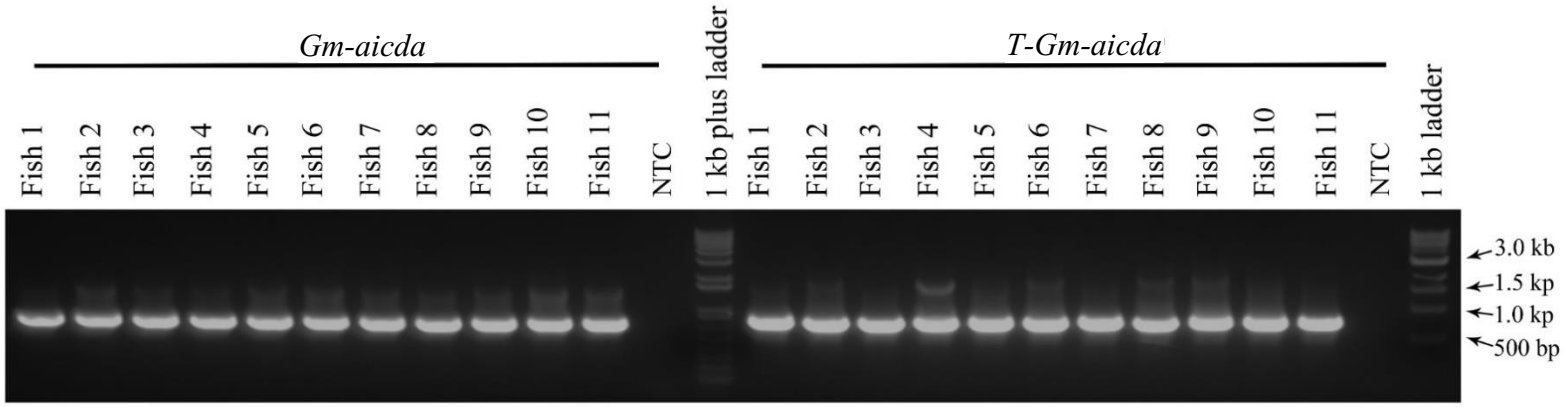


Figure 2-7: Confirmation of the presence of both aicda transcripts in several Atlantic cod individuals through RT-PCR. ISP were used to amplify both transcripts in immune-stimulated splenic cDNA samples.

2.4.3 The Atlantic cod *aicda* expression profile in adult tissues, embryonic, and early larval life stages

The Atlantic cod *aicda* expression pattern of both isoforms was investigated in 19 different tissues using ISPs. The transcript expression of *ef1-α* was also assessed as a normalizer gene. Our RT-PCR analyses revealed that both *Gm-aicda* transcripts were expressed in immune-related tissues and no expression was detected in pIC-stimulated Atlantic cod macrophages (Figure 2-8 A). *Gm-aicda* showed a moderate level of transcript expression compared to *ef1-α* in the spleen, head kidney, and gill. *Gm-aicda* was also expressed in the posterior kidney at moderate to a low level, and in blood and heart, at low levels. The *T-Gm-aicda* transcript was expressed only at low levels in some immune-related tissues, notably spleen. Moreover, *Gm-aicda* transcripts were not expressed in any mucosa-associated lymphoid tissues (MALT) except for the gill-associated lymphoid tissue (GIALT) (Salinas, 2015). Interestingly, *T-Gm-aicda* but not *Gm-aicda* transcript was also detected in male but not female reproductive tissues.

In one other bony fish, the zebrafish, an epigenetic-regulatory role for AID has been suggested during embryogenesis (Abdouni et al., 2013; Rai et al., 2008). To assess the potential role of AID during Atlantic cod embryogenesis, the expression of both *aicda* transcripts was studied in fertilized eggs and early larval stages. *Gm-aicda* isoforms were amplified in total RNA samples extracted from 0-DPF until the yolk-sac absorption stage using RT-PCR and ISPs. The results showed no detectable expression of either *aicda* isoforms in Atlantic cod embryos (Figure 2-8 B).

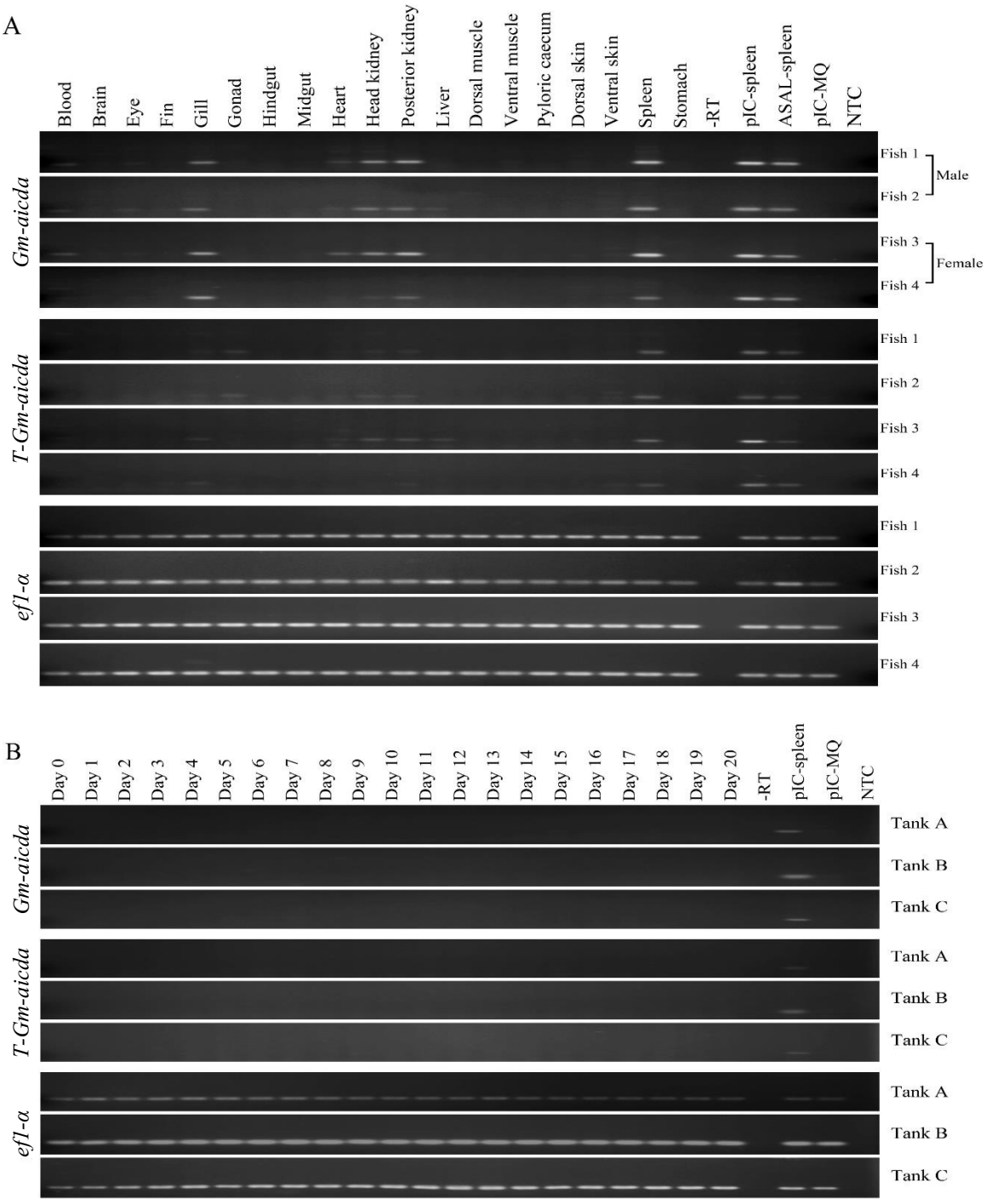


Figure 2-8: Atlantic cod aicda expression profile in adult tissues and embryonic stages. A) Expression of Atlantic cod aicda transcripts was analysed in 19 different tissues extracted from two male and two female Atlantic cod individuals through RT-PCR. Transcript expression of Gm-aicda (top panel), and T-Gm-aicda (middle panel) were compared to ef1-α (bottom panel). B) Gm-aicda transcripts expression during Atlantic cod embryogenesis. No aicda transcript expression was detected.

2.4.4 Atlantic cod *aicda* expression in response to immune stimulation

We then assessed the splenic expression of the *Gm-aicda* transcript in response to immune stimulation by viral (pIC) and bacterial (ASAL) antigens. *Aicda* expression can be induced during B cell activation, either through the interaction of peptide-MHC II complex and CD40 on B cells with T cell receptor and CD40L on $CD4^+$ T helper (T_H) cells or through the dual engagement of B cell receptor and TLRs on B cells with antigens such as LPS (DeFranco, 2016; Hou et al., 2011; Kasturi et al., 2011; Pone et al., 2012; Stavnezer & Schrader, 2014). Although both pathways lead to *aicda* expression, the latter pathway takes place early in immune response when T_H cell assistance is not yet available (Pone et al., 2012). Since the loss of *cd4* and *mhc II* in the Atlantic cod genome are highly suggestive of impaired canonical T_H cell function, we sought to investigate *Gm-aicda* expression in early immune response (Solbakken, Jentoft, Reitan, Mikkelsen, Gregers, et al., 2019; Star et al., 2011; Torresen et al., 2017). In response to pIC and ASAL at 6 HPI, we observed approximately 3- and 2-fold higher expression of *Gm-aicda* transcript, respectively. However, this difference in expression was not detected at 24 HPI (Figure 2-9). In contrast, splenic expression of *T-Gm-aicda* did not significantly change in response to immune stimulation (Figure 2-9). These results indicate that splenic expression of *Gm-aicda* is immune-inducible.

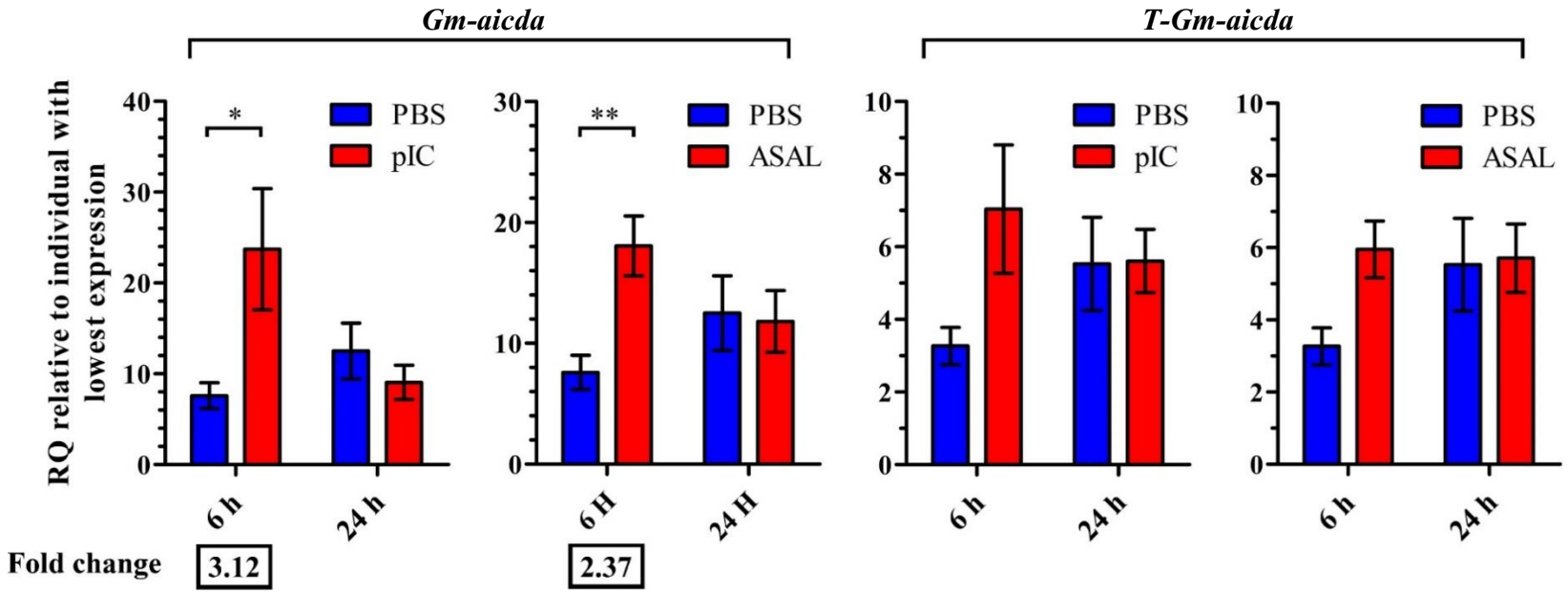


*Figure 2-9: Analysis of Atlantic cod aicda transcripts upon immune stimulation. Gm-aicda transcript expression was normalized to rplp1 and atps expression, and the sample with the lowest normalized expression was used as calibrator. Data are represented as mean ±SEM (n=10). Asterisks represent a significant difference between an immune-challenged group and the corresponding PBS-injected control group. The expression fold-change values are shown below the figures. Gm-aicda and T-Gm-aicda expression were studied at 6 and 24 HPI with pIC or ASAL. Significantly higher expression was only observed at 6 HPI for Gm-aicda transcript (n=10; *: $p < 0.05$; **: $p < 0.01$). Abbreviations: Gm-aicda: full-length Atlantic cod aicda transcript; T-Gm-aicda: truncated aicda transcript identified in Atlantic cod.*

2.4.5 Predicted structural features of Atlantic cod AID protein

Translation of identified *Gm-aicda* CDSs revealed that *Gm-aicda* encodes for a full-length AID protein homologous to AID of other bony fish, whilst T-Gm-AID is missing the N-terminal 21 amino acids (Figure 2-10 A). As expected, Gm-AID exhibited the highest identity and similarity with other bony fish AIDs (Table 2-6). Akin to the other bony fish, Gm-AID contains the bony fish-specific loop five inserts (bony fish insert), as well as an N-terminal extension (Figure 2-10 A and B) (King et al., 2015; Zhao et al., 2005). Unlike other bony fish AIDs, Gm-AID possess extra leucine (L) and threonine (T) amino acids at the C-terminus end making Gm-AID (213 aa) the longest AID identified thus far.

Amino acid alignment of AID homologs revealed that Gm-AID contains all of AID's hallmark functional motifs, including the Zn-coordinating and catalytic residues, secondary catalytic residues, nuclear localization signal, nuclear export signal, and phosphorylation sites (Figure 2-10 A) (Barreto & Magor, 2011; Brar et al., 2004; Chandra et al., 2015; Hu et al., 2013; Ito et al., 2004; King et al., 2015; McBride et al., 2004; Patenaude et al., 2009). Within the AID/APOBEC family, the core catalytic motif is comprised of H[A/V]E-X[24-36]-PCXXC motif in which the histidine (H) and the two cysteines (C) coordinate the catalytic Zn^{2+} and the glutamate (E) acts as proton donor in the deamination reaction (Conticello, 2008). We have previously presented a functional and native structure for Hs-AID using a combined computational-biochemical method, which has been confirmed by later-published X-ray crystal structures of Hs-AID (King & Larijani, 2017; King et al., 2015; Qiao et al., 2017). Using the same methodology, we

generated a predicted structure of Gm-AID and carried out comparisons to Hs-AID, Dr-AID, and Ip-AID. We found that the overall structural architecture of Gm-AID was similar to that of other homologs (Figure 2-10 B). Also, Gm-AID was predicted to form a viable catalytic pocket with equivalent catalytic pocket residues (H60, E62, C100, and C103, equivalent to H56, C87, E58, and C90 in Hs-AID; Figure 2-10 C) (Barreto & Magor, 2011; Brar et al., 2004; Chandra et al., 2015; Hu et al., 2013; Ito et al., 2004; King et al., 2015; McBride et al., 2004; Patenaude et al., 2009).

We previously demonstrated that Hs-AID's catalytic pocket accessibility is determined by 21 secondary catalytic residues that are located on flexible loops which form the walls and floor of the catalytic pocket (King et al., 2015). These amino acids are G23, R24, R25, E26, T27, L29, N51, K52, N53, G54, C55, V57, T82, W84, S85, P86, D89, Y114, F115, C116, and E122 in Hs-AID. In addition to the four core catalytic residues which carry out the enzymatic reaction of deamination, these secondary catalytic residues function in a supporting role to stabilize the target dC in catalytic pocket (King et al., 2015). Although most secondary catalytic residues are highly conserved amongst studied species, we noted that Hs-AID^{E122} is different in Gm-AID (*i.e.*, H136). Moreover, our computational modeling divulged a potential local conformational change around the catalytic pocket of Gm-AID where we noticed that Y127 could potentially protrude into the catalytic pocket of Atlantic cod AID and block the dC entrance by closing the catalytic pocket (Figure 2-11). If confirmed, this conformational change could hamper catalytic activity of Gm-AID. However, ssDNA:AID docking simulation and characterizing the

biochemical properties of purified wild type and mutant Gm-AID^{H136E} is required to confirm this hypothesis.

Our predicted models revealed that the protrusion of Y127 into Gm-AID catalytic pocket is most likely due to preferred T-shape interaction between the side chain of H136 and Y127. This tyrosine is located on AID/APOBECs' substrate specificity loop (*i.e.*, ℓ8) and is fully conserved amongst AID homologs as well as AID/APOBEC family members (Figure 2-10 A) (Abdouni et al., 2013; Iyer et al., 2011). Indeed, several previous studies have emphasized the importance of this tyrosine residue (Abdouni et al., 2013; Iyer et al., 2011; Wijesinghe & Bhagwat, 2012). Interestingly, the substrate specificity transition from adenine to cytidine during the emergence of the AID/APOBECs from adenosine deaminases has been attributed to the expansion of ℓ8 and the base-stacking ability of the abovementioned conserved tyrosine (Iyer et al., 2011). Remarkably, in APOBEC3A, the greater distance of this tyrosine' side chain (*i.e.,* Hs-A3A^{Y130}) from the catalytic pocket compared with that of Hs-AID^{Y114}, was postulated as the basis of Hs-A3A ability to efficiently deaminate 5m-C (Wijesinghe & Bhagwat, 2012). Additionally, when modeled based on NMR structure of APOBECs, this tyrosine was noted to rotate ~ 180 ° and shifted away from the catalytic pocket in Hs-AID and Dr-AID (Abdouni et al., 2013). In the "away" conformation, the steric hindrance between the side chain of this tyrosine and 5m-C would be eliminated. Since, compared to Hs-AID, ℓ8 in Dr-AID has an extra negatively charge amino acid (E130) and, therefore, it is more flexible, it was suggested that the "away" conformation of this tyrosine in Dr-AID may occupy a lower energy state

compared to that of Hs-AID^{Y114}, explaining the higher efficiency of Dr-AID in deaminating 5m-C compared with that of Hs-AID (Abdouni et al., 2013).

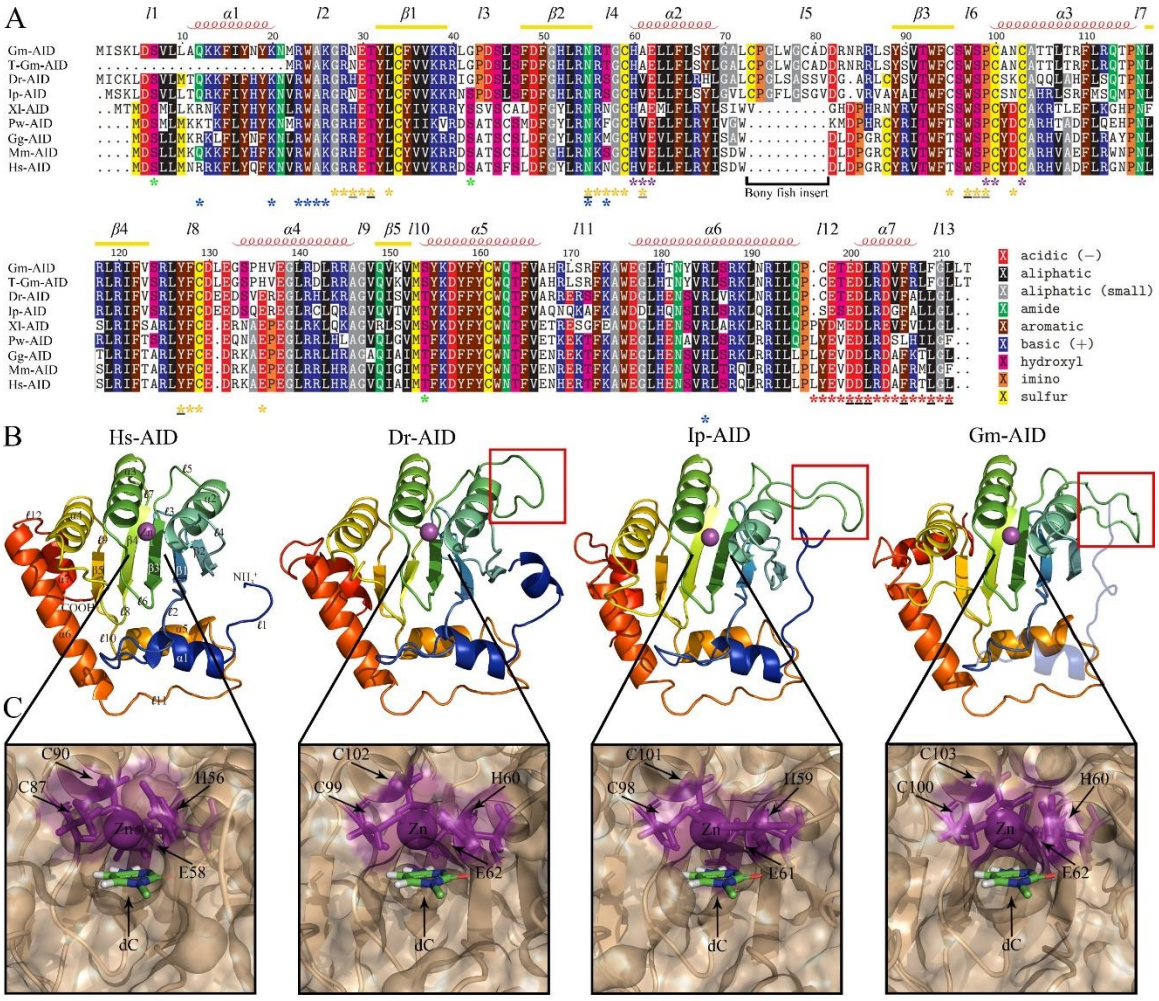


Figure 2-10: General structural features of Atlantic cod AID. A) Sequence alignment of Gm-AID and T-Gm-AID with representative AID homologs from different classes of vertebrates. The approximate secondary structure of α-helical, β-strand, and loop regions are shown. Residues which comprise well-established AID functional domains are labelled with asterisks: main Zn2+-coordinating and catalytic residues (purple), secondary catalytic residues (yellow), nuclear localization signal (blue), nuclear export signal (red), and phosphorylation sites (green) are labeled with asterisks. Residues are colored according to chemical properties of the side chain. B) Representative ribbon model of predicated Gm-AID structure with that of solved Hs-AID structure and predicted structures of two other bony fish (Dr-AID and Ip-AID). In each model, blue to red color change indicates N to C terminus progression and the catalytic pocket zinc is shown in purple. Loops, β-strands, and α-helices are labeled in the Hs-AID model. The bony fish insert is shown in a red box in predicted models of bony fish AIDs. The first 21 amino acid-long motif missing from T-Gm-AID is transparently shown (last right panel). Comparison of predicted structure of Gm-AID with that of other AID homologs revealed no major differences in overall structural architecture. C) AID core catalytic motif. Comparison of this motif amongst different AID homologs revealed that Gm-AID forms a classical and potentially viable AID catalytic pocket. Abbreviations: Gm-AID: Atlantic cod AID; T-Gm-AID: truncated isoform of Atlantic cod AID; Dr-AID: zebrafish AID; Ip-AID: channel catfish AID; Xl-AID: South African clawed toad AID; Pw-AID: the Iberian ribbed newt AID; Gg-AID: chicken AID; Mm-AID: mouse AID; Hs-AID: human AID.

Table 2-6: Comparison of AID amino acid identity and similarity amongst different species

Identity

	Gm-AID	Dr-AID	Ip-AID	Xl-AID	Pw-AID	Gg-AID	Mm-AID	Hs-AID
Gm-AID		77%	73%	62%	60%	60%	61%	60%
Dr-AID	83%		78%	63%	63%	62%	67%	63%
Ip-AID	82%	88%		61%	61%	59%	60%	60%
Xl-AID	72%	73%	74%		71%	67%	69%	68%
Pw-AID	69%	74%	75%	86%		77%	72%	77%
Gg-AID	68%	73%	71%	84%	88%		88%	90%
Mm-AID	70%	75%	73%	86%	91%	94%		92%
Hs-AID	68%	74%	73%	85%	87%	94%	95%	

Similarity

Abbreviations: Gm-AID: Atlantic cod AID; Dr-AID: zebrafish AID; Ip-AID: channel catfish AID; Xl-AID: South African clawed toad; Pw-AID: the Iberian ribbed newt AID; Gg-AID: chicken AID; Mm-AID: mouse AID; Hs-AID: human AID.

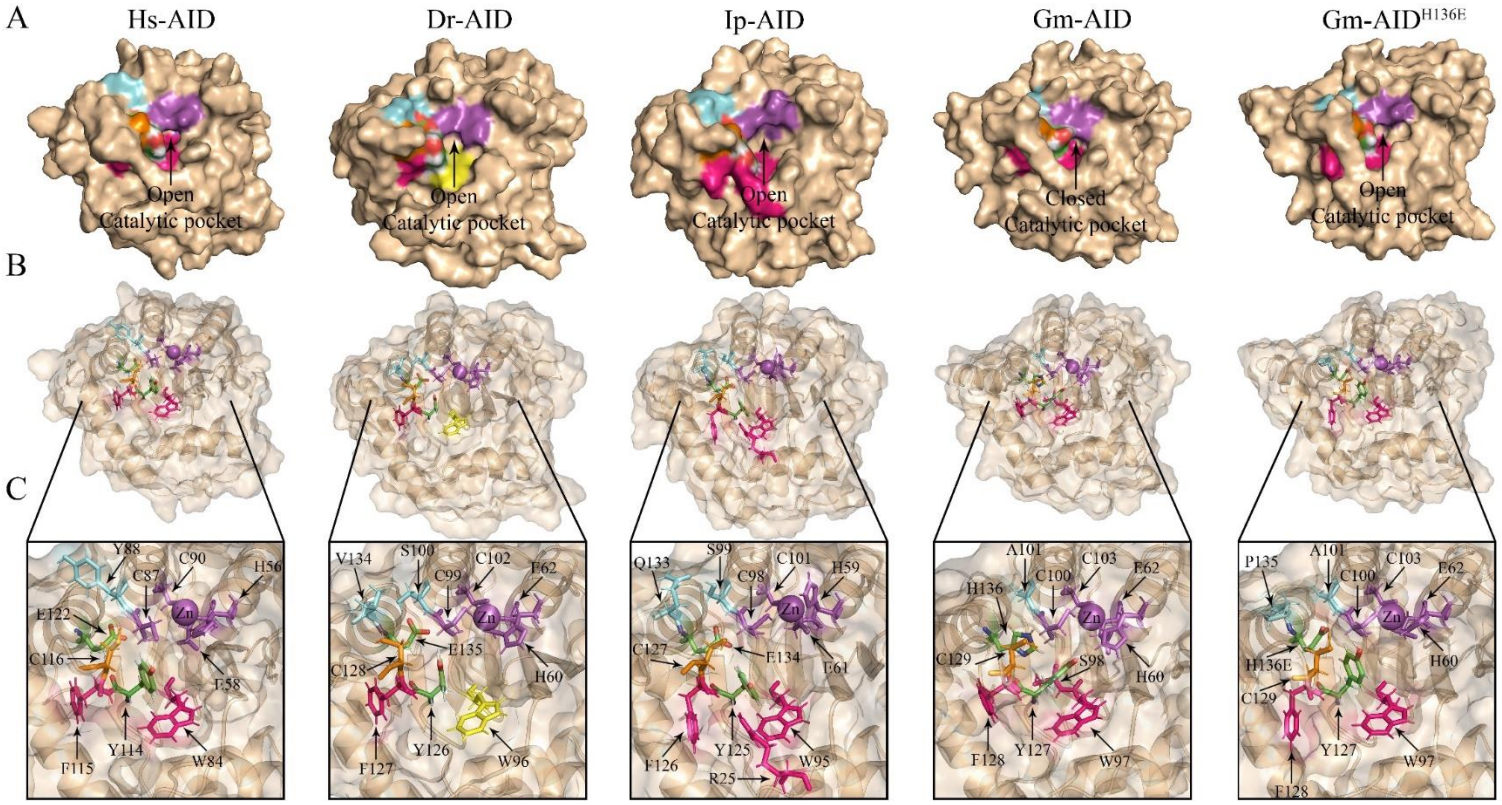


Figure 2-11: Potential conformational changes induced by H136 in Atlantic cod AID compared to the corresponding glutamic acid (E) in other AID homologs. A) Representative surface model of predicated AID structures based on solved Hs-AID structure, showing the closed catalytic pocket of Gm-AID due to protrusion of Y127 into the pocket. B) zoomed out and C) zoomed in detailed conformational changes induced by Gm-AID^{H136} vs. its corresponding amino acid in other AID homologs (Hs-AID^{E122}, Dr-AID^{E135}, Ip-AID^{E134}, and Gm-AID^{H136E}). In each model, the catalytic pocket zinc and the core catalytic motif residues are shown in purple. In all models, the amino acids in 4 A° radius of Gm-AID^{H136}, Gm-AID^{Y127}, and both residues or the corresponding amino acids in other AID homologs are colored in cyan, pink, and orange. In all models, the Gm-AID^{H136} and Gm-AID^{Y127} or their corresponding amino acids in other AID homologs are within 4 A° distance of each other. In all AID models, the Gm-AID^{C100} or its corresponding amino acid in other AID homologs is within 4 A° distance from Gm-AID^{H136}. In all AID models except Ip-AID, the Gm-AID^{C100} or its corresponding amino acid in other AID homologs is within 4 A° distance from Gm-AID^{Y127}. Only in Gm-AID, H60 is within 4 A° distance of Gm-AID^{Y127}. In Dr-AID, W96 is within 4.5 A° distance of Dr-AID^{Y126}. H136E mutation in Gm-AID can reverse the Y127 protrusion into the catalytic pocket.

2.5 Discussion:

Antibody affinity maturation has been observed across vertebrates (Bromage et al., 2006; Cain et al., 2002; Dooley & Flajnik, 2005; Dooley et al., 2006; Hsu, 2016; Jenne et al., 2003; Kaattari et al., 2002; Malecek et al., 2005; Marianes & Zimmerman, 2011; Mehr et al., 2004; Wilson et al., 1992; Yang et al., 2006). Among studied bony fish, the Atlantic cod has emerged as a unique species that lacks antigen-specific antibody affinity maturation (Arnesen et al., 2002; Lund et al., 2008; Lund et al., 2006; Magnadottir et al., 2001; Schroder et al., 2009; Solem & Stenvik, 2006). Since AID is the initiator of secondary antibody diversification, we explored Atlantic cod *aicda* tissue expression pattern, expression during embryogenesis, and transcript expression in early immune response (Sernandez et al., 2008; Wang et al., 2009). Here, we show that the chromosomal location of Atlantic cod *aicda* locus has a conserved synteny compared to other Teleostei *aicda*. We also report that Atlantic cod expresses two distinct *aicda* transcripts one of which is missing the first exon. Although both transcripts are predominantly expressed in immune-related tissues with no detectable expression during embryogenesis, only expression of the full-size transcript increases in the context of immune stimulation. Our computational protein modeling also reveals that the full-length Atlantic cod AID protein contains all the conserved structural properties of other studied AID homologs. However, we noticed a drastic change in one of the secondary catalytic residues in Atlantic cod (*i.e.,* Gm-AID^{H136} equivalent to Hs-AID^{E122}) which, if confirmed, could impair the enzymatic activity of Gm-AID.

Our synteny analyses revealed a conserved synteny for *aicda* locus among Teleostei species suggesting the possibility of similar expression and gene regulation compared to other teleost species. Lack of complete genomic sequence of earlier Sarcopterygii species (*i.e.*, coelacanth [*Latimeria chalumnae*]) along with the previously reported potential loss of *aicda* gene from lungfish (*Protopterus dolloi* and *P. annectens*; Sarcopterygii: Dipnoi) impeded a definitive conclusion whether *aicda* synteny was conserved in the entire Sarcopterygii class (Tacchi et al., 2015). Nevertheless, these results suggest that during the teleost-specific whole-genome duplication (TS-WGD) event, a different copy of the *aicda* has been retained in teleost species compared to the tetrapod group (Glasauer & Neuhauss, 2014).

Previous studies have reported the presence of different *aicda* isoforms in several vertebrate species but not in the two bony fish (channel catfish and zebrafish) whose *aicda* transcripts have been well-studied (Saunders & Magor, 2004; Zhao et al., 2005). In Iberian ribbed newt (*Pleurodeles waltl*), three potential poly-A sites and two *aicda* isoforms, one of which, like the *T-Gm-aicda*, is missing the first exon, have been described (Bascove & Frippiat, 2010). In African clawed frog, two different *aicda* transcripts of 2 and 1.3 kb-length were found (Marr et al., 2007). In dogs (*Canis lupus familiaris*) and cows (*Bos taurus*) only one *aicda* transcript was reported while in mice two *aicda* transcripts containing complete *aicda* CDS but utilizing different poly-A sites were identified (Muramatsu et al., 1999; Ohmori et al., 2004; Verma et al., 2010). In human, five different splice variants of *aicda* have been detected where individual human B cells only express one of the *aicda* splice variants. These splice variants are: Full-length AID (AID-FL),

exclusion of the beginning of exon 4 (AID-ΔE4a), exclusion of exon 4 (AID-ΔE4), exon 3 and 4 exclusion (AID-ΔE3E4), and inclusion of intron 3 containing a stop codon (AID-ivs3) (Albesiano et al., 2003; McCarthy et al., 2003; Wu et al., 2008). In this chapter, we found two *aicda* isoforms one of which is missing the first exon. Unlike *Hs-aicda*, *Gm-aicda* isoforms are the result of different transcription start site usage rather than alternative splicing, suggesting involvement of different transcription factors.

Previous studies conducted on vertebrates have identified lymph node and spleen as the main *aicda*-expressing tissues (Bascove & Frippiat, 2010; Marr et al., 2007; Muramatsu et al., 1999; Muto et al., 2000; Ohmori et al., 2004; Saunders & Magor, 2004; Verma et al., 2010). Lower and variable levels of *aicda* expression have also been reported in thymus, pancreas, kidney, liver, and lung of mammals (Muto et al., 2000; Ohmori et al., 2004; Verma et al., 2010). Likewise, low levels of *aicda* expression have been observed in the brain, intestine, kidney, liver, and lung of amphibians, and the intestine, fin, posterior and anterior kidney of fish (Bascove & Frippiat, 2010; Marr et al., 2007; Saunders & Magor, 2004). In this chapter, we identified the immune-related tissues (*i.e.*, spleen, head kidney, and gill) as the main site of *Gm-aicda* expression. This is consistent with the previous study where the melano-macrophage clusters have been identified as the main site of *aicda*-expressing B cells in early gnathostome vertebrates (Saunders et al., 2010). In most fish species, these clusters mainly exist in the spleen and kidney and to lesser extent in the liver and intestine (Agius & Roberts, 2003; Arciuli et al., 2017; Diaz-Satizabal & Magor, 2015). Compared to *Gm-aicda* transcript, *T-Gm-aicda* transcript was only expressed at low levels and mainly in the spleen that suggests that *Gm-aicda* is the main

aicda transcript in Atlantic cod. Therefore, we concluded that *aicda* transcripts were mostly but not exclusively expressed in immune-related tissues in adult fish.

Interestingly, we detected low levels of *T-Gm-aicda* isoform transcript in male gonad. *Gm-aicda* is located on Linkage Group (LG) 11 which has recently been proposed to contain the majority of the Atlantic cod sex-locus (Star et al., 2016). One possible explanation is that *aicda* expression in male gonad tissue could partly be due to its proximity to the sex-locus. In this scenario, its expression as *T-Gm-aicda* isoform, which most likely lacks catalytic activity, might be due to lack of proper transcription factor(s) and it could be a safeguard to protect the genome in male gonad tissue. Also, since *aicda* transcript and protein expression, and *apobec4* transcript expression have been detected in human spermatocytes and testis, respectively, *T-Gm-aicda* transcript expression in male gonad might be a remnant of an ancient unknown role of AID (Marino et al., 2016; Rogozin et al., 2005; Schreck et al., 2006).

Besides the established role of AID in antibody affinity maturation, a controversial role for AID was suggested in embryonic development in zebrafish (Rai et al., 2008; Shimoda et al., 2014). *Aicda* expression was also observed in the early stages of embryogenesis in Iberian ribbed newt and early larval stages in African clawed frog (Bascove & Frippiat, 2010; Marr et al., 2007). In this study, we observed no *aicda* expression during Atlantic cod embryogenesis. Therefore, we concluded that *aicda* transcripts were unlikely to play a role during embryogenesis.

AID expression can be induced during both T cell-dependent and T cell-independent B cell activation (TD and TI pathways, respectively). During TD B cell

activation, peptide-MHC II complex and CD40 on B cells interact with TCR and CD40L on T helper cells (*i.e.*, T_H cell), while the dual engagement of B cell receptor and TLRs on B cells with antigens such as LPS activate B cell without T_H cell assistance (DeFranco, 2016; Hou et al., 2011; Kasturi et al., 2011; Pone et al., 2012; Stavnezer & Schrader, 2014). The lack of key genes involved in T-cell/B-cell interactions from Atlantic cod genome, may have compromised the *aicda* expression through TD pathway. Therefore, we investigated *Gm-aicda* expression during the early immune response (*i.e.*, TI pathway). It should be noted that a previous study has shown *aicda* expression in murine B cells through both pathways where expression through the TI pathway peaked at 24 to 48 hours post immune challenge (100-fold increase) (Pone et al., 2012). We observed a moderate increase (2- to 3-fold) in *Gm-aicda* transcript expression only in early response to pIC and ASAL stimulation (6 HPI). This observed higher expression of *Gm-aicda* upon immune stimulation could be due to upregulation of its expression and/or increased number of activated B cells that express *aicda*. To distinguish between these two scenarios, further studies are required. Importantly, the expression of *T-Gm-aicda* was not affected during this time frame which indicates the lack of immune-related role for the truncated isoform. Nevertheless, the conserved *Gm-aicda* gene synteny compared to other teleosts and the observed increase in *Gm-aicda* transcript expression upon immune stimulation could indicate that the regulation and transcript expression of *aicda* might be evolutionary conserved in Atlantic cod.

T-Gm-AID, if translated, lacks the first 21 amino acid residues compared to Gm-AID. These residues are involved in stabilization of the core of the enzyme, stabilization

of the surface DNA binding residues, and contains potential DNA binding residues (King & Larijani, 2017; King et al., 2015). Moreover, truncation of the first 10 or 20 amino acids from Hs-AID impaired its nuclear import by reducing its affinity for importin-α3 (Hu et al., 2013; Patenaude et al., 2009). Due to the importance of the AID N-terminal amino acids in AID activity, we predicted that T-Gm-AID, if translated, was inactive and would not localize into the nucleus.

Although we confirmed the presence of all AID's well-known functional motifs in Gm-AID, we detected a drastic change from glutamic acid (E) to histidine (H) in one of the secondary catalytic residues. These residues function in a supporting role to stabilize the target dC in catalytic pocket (King et al., 2015). Moreover, Hs-AID^{E122} resides in the substrate-specificity loop (ℓ8), a motif that has previously been shown to be critical regulator of cytidine deamination activity and DNA targeting specificity across all AID/APOBEC family members (Gajula et al., 2014; Iyer et al., 2011; Kohli et al., 2009). In Hs-AID, E122 stabilizes other secondary catalytic residues (residues from the N-termini of ℓ8 and conserved residues from ℓ6) and plays a role in stabilizing dC and neighboring ssDNA in AID:DNA complexes. Accordingly, perturbation of this residue in Hs-AID resulted in a drastic reduction in activity, consistent with its role in regulation of catalytic pocket dynamics (Gajula et al., 2014).

Our computational modeling revealed that due to the proximity of Hs-AID^{E122} to Hs-AID^{Y114}, its replacement with histidine in Gm-AID (*i.e.,* H136) could cause a protrusion of Y127 into the catalytic pocket, thereby blocking the catalytic pocket and potentially producing a catalytically inaccessible conformation (*i.e.*, closed conformation).

Similar conformations of the Gm-AID^{Y127} equivalent residue in Hs-AID (*i.e.*, Y114) were shown to restrict the catalytic pocket (King et al., 2015). Additionally, the recent AID crystal structure has shown that Y114 assisted in holding cytidine in place by interacting with dC O5′ and the interactions between Y114 and F115 contributed to shape the catalytic pocket and defining the substrate specificity of AID (Gajula et al., 2014).

Due to the dissimilar chemical properties of histidine and glutamic acid, the replacement of Hs-AID^{E122} with Gm-AID^{H136} could lead to significant conformational changes. E is a negatively charged amino acid while H is mostly neutral at the physiological pH. Previous studies have shown that the π-π interactions between H and the aromatic amino acids contributes significantly in protein stability and reduces the protein solubility (Hou et al., 2018). The optimum imidazole-benzene interactions are in a T-shaped conformation (Kumar et al., 2018; Schaeffer, 2008; Trachsel et al., 2015). Histidine is capable of forming both N–H … π and C–H … π interactions with aromatic amino acids where the N–H … π is more stable than C–H … π interactions (-14.0 kcal mol^{-1}*vs.* -11.5 kcal mol^{-1}) presumably due to the increased polarity of the N–H bond (Kumar et al., 2018; Trachsel et al., 2015). However, data-mining studies uncovered a 4:1 ratio of C–H … π to N–H … π interactions in the T-shaped interactions (Trachsel et al., 2015). On the other hand, the positively charged edge of an aromatic ring can interact with an anion to form an anion-π interaction. These edgewise interactions can produce stabilizing interactions with -2 to -7.3 kcal mol^{-1} contributing to the overall structural stability of the proteins (Chakravarty et al., 2018; Newberry & Raines, 2019; Philip et al., 2011). Many of these

anion-π interactions were also involved in the nearby π-π interactions, creating anion-π-π triads (Philip et al., 2011).

Hs-AID^{E122}, Dr-AID^{E135}, and Ip-AID^{E134} could potentially participate in an anion-π interaction with Hs-AID^{Y114}, Dr-AID^{Y126}, and Ip-AID^{Y125}, respectively (Figure 2-11). In these interactions, the preferred orientation is when the carboxyl group of E is in a close-to-parallel conformation with respect to the interacting aromatic plane (Lucas et al., 2016). Therefore, tyrosine is positioned out of the respective catalytic pocket. However, in Gm-AID, Propka analyses showed the pKa of 5.63 for H136, leaving it mostly neutral at pH 7 (Dolinsky et al., 2004; Olsson et al., 2011). In this case, H136 side chain (imidazole ring) can potentially participate in π-π interactions with Y127 in a T-shaped conformation, causing the protrusion of Y127 into the catalytic pocket. However, more studies are required to assess the potential impact of this drastic change (*i.e.*, Hs-AID^{E122} to Gm-AID^{H136}) in the enzymatic activity of Atlantic cod AID.

In summary, here for the first time, we showed that although Atlantic cod has lost MHC II pathway, it increases *aicda* expression in the context of innate immune system in response to immune stimulation. These results indicate that likely, *aicda* expression but maybe not its function, have been conserved during the evolution of Atlantic cod.

Chapter 3:

Impairment of the enzymatic function of activation induced cytidine deaminase (AID) in Atlantic cod (*Gadus morhua*)

3.1 Abstract

In vertebrates, the enzyme activation-induced cytidine deaminase (AID, encoded by the *aicda* gene) introduces somatic mutations at the immunoglobulin (*Ig*) loci to instigate the process of antibody affinity maturation, generating high affinity antibodies. Unlike other studied bony fish, the Atlantic cod (*Gadus morhua*) humoral response lacks affinity-matured antibodies. In the previous chapter, we showed that the Atlantic cod *aicda* gene locus is conserved compared to other studied teleost species and that it encodes two transcripts which are expressed in immune-related tissues. Here we sought to investigate the enzymatic properties of the Atlantic cod AID protein (Gm-AID) to shed light on the molecular basis responsible for the lack of antibody affinity maturation in this species. Our biochemical analyses of the purified Gm-AID proteins showed that the truncated isoform is inactive and the full-length Gm-AID, despite the ability to bind DNA like other AID homologs, is two to three orders of magnitude less catalytically active, exhibiting barely detectible enzymatic activity. Gm-AID also exhibits the coldest temperature adaptation of any purified vertebrate DNA/RNA-editing enzyme studied to date, with an optimal activity range of 4 to 8 °C. AID preferentially mutates WRC (W=A/T, R=A/G) motifs. Accordingly, the complementarity determining region (CDR) of *Ig* variable genes (*IgV*) of mammals and fish are enriched in WRC motifs, reflecting substrate:enzyme co-evolution. We found that the Atlantic cod *Ig* gene CDRs exhibit a reduced level of WRC enrichment, consistent with compromised Gm-AID functionality. Taken together, our findings suggest that the Atlantic cod may represent a unique instance in evolution of immunity wherein

AID has become nearly inactive to reflect lesser reliance on high affinity antibody responses.

3.2 Introduction

Activation induced cytidine deaminase (AID) is a member of the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (AID/APOBEC) family of proteins. AID mutates deoxycytidine (dC) to deoxyuridine (dU) on single-stranded DNA (ssDNA), preferentially in the context of WRC (W=A/T; R=A/G) motifs (Bransteitter et al., 2003; Dancyger et al., 2012; Emma M. Quinlan, 2017; Larijani, Frieder, Basit, et al., 2005; Larijani et al., 2007; Meffre et al., 2001). AID is mainly expressed in mature activated B lymphocytes where it introduces mutations in the antibody gene V and C regions, thereby mediating somatic hyper-mutation (SHM) and class switch recombination (CSR) of antibody genes, leading to secondary antibody diversification (Bransteitter et al., 2003; Bransteitter et al., 2006; Frieder et al., 2006; Kolar et al., 2007; Larijani, Frieder, Basit, et al., 2005; Meffre et al., 2001; Muramatsu et al., 2000; Muramatsu et al., 1999; Muto et al., 2000; Nagaoka et al., 2002). The absolute requirement of AID for secondary antibody diversification is apparent in the case of hyper IgM syndrome type II (HIGM II) patients manifesting lack of SHM and CSR caused by deficiency in AID gene (Minegishi et al., 2000; Revy et al., 2000).

AID is a small positively charged protein that binds its ssDNA substrate with ~nM-range binding affinity (Larijani et al., 2007). Previous studies using a computational-evolutionary-biochemical approach as well as the X-ray crystal structure of AID revealed the presence of three DNA-binding grooves on AID's surface (King & Larijani, 2020). Amongst these grooves, the DNA-binding groove 1 and the assistance patch create AID's bifurcated substrate-binding surface, explaining AID's preference for the G-quadruplex

(G4) substrate (Qiao et al., 2017). Since the residues forming the substrate-binding groove 1 directly interact with ssDNA, they establish AID's substrate specificity (Qiao et al., 2017). The presence of the DNA-binding groove 2 was predicted through DNA:protein docking simulations (King et al., 2015). Although the DNA-binding groove 1 seems to be the main substrate binding domain, the ssDNA bound into the DNA-binding groove 2 also passes over the catalytic pocket, potentially positioning the dC properly in the catalytic pocket (King & Larijani, 2020; King et al., 2015).

Within AID, the conserved catalytic domain of H(A/V)EX$_{(24-36)}$PCXXC and the secondary catalytic residues are responsible for catalyzing the deamination reaction and stabilizing the dC in the active site, respectively (Barreto & Magor, 2011; Conticello, 2008; Harris et al., 2002; King et al., 2015). Despite the conserved overall arrangement of catalytic residues, AID's catalytic rate varies significantly between different species, potentially due to subtle breathing dynamics of the catalytic pocket (Barreto et al., 2005; Conticello et al., 2005; Dancyger et al., 2012; Ichikawa et al., 2006; Wakae et al., 2006). Thus far, studies have shown that zebrafish (*Danio rerio*) AID (Dr-AID) is catalytically the most robust AID and Hs-AID is more active than channel catfish (*Ictalurus punctatus*) AID (Ip-AID) (Abdouni et al., 2013; Dancyger et al., 2012; King et al., 2015). Besides activity on dC, Dr-AID is also uniquely capable of efficiently deaminating 5-methyl dC (5-mC), potentially underling a unique role for Dr-AID in epigenetic remodeling through demethylation of CpG motifs during embryogenesis (Abdouni et al., 2013; Rai et al., 2008). While Dr-AID is the most robust enzyme on 5m-C (2/1), Hs-AID and Ip-AID are not efficient in deaminating 5m-C (Abdouni et al., 2013). Nevertheless, since all AIDs

studied thus far showed less activity on 5m-C, it has been suggested that methylation protects dC from AID targeting (Abdouni et al., 2013; Larijani, Frieder, Sonbuchner, et al., 2005).

In addition to catalytic rate and activity on 5-mC, AID homologs from different species show different optimal temperature and substrate specificity (Abdouni et al., 2013; Dancyger et al., 2012; Emma M. Quinlan, 2017; King et al., 2015; Larijani, Frieder, Sonbuchner, et al., 2005; Nabel et al., 2012). Mammalian and avian AIDs exhibit the highest deamination activity at higher temperatures (*i.e.*, around 37^{o}C), while AIDs from amphibians and bony fish are more active at lower temperatures like 18 ^{o}C (Barreto et al., 2005; Conticello et al., 2005; Dancyger et al., 2012; Emma M. Quinlan, 2017). Sequencing analyses of *IgV* genes and biochemical studies have defined the WRC motif as AID's favored target motif (Dancyger et al., 2012; Gajula et al., 2014; Hackney et al., 2009; Larijani, Frieder, Basit, et al., 2005; Larijani & Martin, 2007; Malecek et al., 2005; Marianes & Zimmerman, 2011; Yang et al., 2006). However, more distant homologs such as cartilaginous fish and lamprey AID exhibit divergent patterns of sequence specificity, sometimes favoring non-WRC motifs (Emma M. Quinlan, 2017). WRC enrichment in the complementary-determining regions (CDRs) of the *Ig* genes of mammals, birds, amphibians, bony fish, and cartilaginous fish has been observed (Conticello et al., 2005; Detanico et al., 2016; Golub & Charlemagne, 1998; Jolly et al., 1996; Oreste & Coscia, 2002; Wagner et al., 1995; Wei et al., 2015). This co-evolution of AID substrate specificity and the sequence of *Ig* genes may play a significant role in ensuring efficient AID activity at *Ig* genes.

Unlike other studied vertebrates, functional analyses of the Atlantic cod humoral immune responses revealed no evidence of antibody affinity maturation. Specifically, many studies have shown high levels of low affinity serum IgM in Atlantic cod, and a lack of robust antigen-specific antibody responses upon immunization (Arnesen et al., 2002; Magnadottir et al., 1999; Magnadottir et al., 2001; Solem & Stenvik, 2006). Moreover, full sequencing of the Atlantic cod genome revealed a unique gene structure of its immune system, namely loss of *mhc II*, *cd4*, invariant chain (*Ii*), *tlr1/2/5/21β*, and *Mx* genes and expansion of *mhc I* and *tlr7/8/9/22/25* (Malmstrom et al., 2016; Solbakken, Rise, et al., 2016; Solbakken, Torresen, et al., 2016; Star et al., 2011; Torresen et al., 2017). In line with the loss of central genes required for T cell-dependent B cell activation, thus far only T cell-independent B cell activation has been reported in this species (Malmstrom et al., 2013; Solbakken, Jentoft, Reitan, Mikkelsen, Gregers, et al., 2019).

In the previous chapter, we showed that the putative *aicda* gene locus (encodes AID protein) in Atlantic cod exhibits conserved synteny with other teleosts species. We also found that the Atlantic cod *aicda* transcript is expressed mostly in immune-related tissues in the form of two distinctive isoforms. The main mRNA transcript encodes for a full-length AID protein (*i.e.*, 213 amino acids; Gm-AID) while the second mRNA encodes a truncated isoform (*i.e.*, 192 amino acids; T-Gm-AID). We also found that the expression of full-length transcript is increased during immune stimulation with viral or bacterial mimics. However, the expression of the truncated transcript is unresponsive to immune stimulation. In this chapter, we sought to explore the functional enzymatic properties of Atlantic cod AID isoforms to pinpoint the molecular basis behind the lack of antibody

maturation in this species. Here, we report that, the T-Gm-AID, if translated, is an inactive cytidine deaminase. In contrast, Gm-AID is a *bona fide* cytidine deaminase with the coldest optimal temperature reported for any AIDs thus far. However, we found that the enzymatic activity of AID is drastically reduced in Atlantic cod and we did not observe WRC enrichment in Atlantic cod CDRs to levels found in other vertebrates.

3.3 Methods

3.3.1 AID expression and purification

Gm-AID was expressed along with human AID (Hs-AID), zebrafish AID (Dr-AID), channel catfish AID (Ip-AID) for biochemical analyses. Prokaryotic expression and purification of glutathione S-transferase (GST)-AID were performed as described in a well-established protocol (Abdouni et al., 2013; Dancyger et al., 2012; Emma M. Quinlan, 2017; Larijani et al., 2007). The GST-AID expression vector was constructed by inserting the coding sequence of each AID homolog into the pGEX-5x-3 vector (GE Healthcare, Waukesha, WI, USA) using the EcoRI enzyme restriction site located in the multiple cloning site downstream of the GST-encoding sequence. Site-directed mutagenesis and PCR-based manipulations were conducted to create single point mutants and T-Gm-AID, respectively. For each GST-AID construct, between two to six independent protein preparations were purified from *E. coli* Bl21(DE3) cells (Abdouni et al., 2013; Dancyger et al., 2012; Emma M. Quinlan, 2017; Larijani et al., 2007). A 500-ml culture of DE3 cells containing GST-AID expression vector was grown at 37 °C and 225 rpm in the presence of 100 µg/ml ampicillin. When the culture reached the log phase (an OD of 0.6), 1 mM of Isopropyl β-d-1-thiogalactopyranoside (IPTG) and 100 µg/ml ampicillin were added. Bacterial cultures were then incubated at 16 °C and 225 rpm for 16 h. The bacterial culture was centrifuged, and the pellet was resuspended in 20 ml of phosphate-buffered saline (PBS, Sigma) pH 7.5. Cells were lysed by a French Pressure cell and centrifuged to collect the supernatant. GST-AID was then column-purified from the supernatant of lysed cells using Glutathione Sepharose high-performance beads (Amersham) as per manufacturer's

recommendations. Briefly, the supernatant was applied twice to a purification column, and washed with 50 ml of PBS, pH 7.5. GST-AID was eluted with elution buffer (50 mM Tris [pH 8.0] and 10 mM L-Glutathione reduced) into 0.5-ml fractions. The quantity of protein in each fraction was measured using NanoDrop spectrophotometry (ND-1000) and between four to five fractions containing > 0.5 mg/ml total protein were dialyzed overnight at 4 °C into the final storage buffer (20 mM Tris pH 7.5, 100 mM NaCl, and 1 mM dithiothreitol). Purified GST-AID was aliquoted into 50- to 100-µl aliquots, flash frozen, and stored at -80 °C. Moreover, eukaryotic expression of Gm-AID in HEK293T cells was also carried out (Abdouni et al., 2018). Briefly, GST-AID fragment was inserted into pcDNA3.1-V5-6xHis-Topo vector and 5 µg of plasmid per plate was transfected into 10-cm plates of HEK 293T cells (seeded with 5×10^5 cells) using Polyjet transfection reagent (FroggaBio). Fifty plates were transiently transfected per GST-AID homolog. Following 48 h incubation at 37 °C, cells were resuspended in PBS (pH 7.5) containing 50 µg/ml RNase A (Invitrogen) and 0.2 mM phenylmethylsulfonyl fluoride (PMSF, Sigma). Cells were then lysed using a French Pressure cell. Samples were run through the French Pressure cell three times with a 30-min incubation at room temperature before the last run to allow the RNase A time to act. GST-AID was then purified from supernatant using Glutathione Sepharose high-performance beads (Amersham). Briefly, the supernatant was applied to the purification column twice, and washed with 50 ml of PBS (pH 7.5) containing 0.2 mM PMSF. GST-AID was eluted off the beads using 50 mM Tris (pH 8) and 10 mM L-Glutathione reduced. 0.25-ml fractions were collected and analyzed by SDS-PAGE and stained with Coomassie blue. Fractions containing the band of interest (~ 48

kDa) were combined. Then, 5 % glycerol and 50 µg/ml of bovine serum albumin (BSA, Invitrogen) were added before dialyzing the fractions overnight at 4 °C into the final storage buffer (20 mM Tris pH 7.5, 100 mM NaCl, 5 % glycerol, and 1 mM dithiothreitol). Purified GST-AID was aliquoted into 50- to 100-µl aliquots, flash frozen, and stored at -80 °C. Alternatively, beads with bound GST-AID were washed with PBS (pH 7.5) and stored in AID storage buffer as bead-bound AID. The quality and quantity of the purified prokaryotic and eukaryotic AID preparations were assessed using coomassie staining and western blotting, respectively. In western blot analyses, anti-GST (SantCruz) antibodies and Goat anti-Rabbit IgG (SantaCruz) were used as the primary and secondary antibodies.

3.3.2 Substrate preparation

Partially single-stranded bubble substrates containing a WRC or a non-WRC motif (*i.e.*, WRCbub7 or non-WRCbub7) were used to determine the enzymatic properties of GST-AID homologs (Abdouni et al., 2013; Dancyger et al., 2012; King et al., 2015; Larijani & Martin, 2007; Larijani et al., 2007). 2.5 pmol of the target strand (synthesized by IDT) was 5'-radiolabeled with [γ-^{32}P] dATP using polynucleotide kinase enzyme (PNK, New England BioLabs) at 37 °C for one hour. To remove the excess free [γ-^{32}P] dATP, reactions were purified through mini-Quick spin DNA columns (Roche, Indianapolis, IN, USA). To generate partially single-stranded bubble substrate, the radiolabeled oligo was then annealed to three-fold excess of its partially complementary strand in the presence of 100 mM KCl. Samples were subjected to slow cooling (*i.e.*, 1 °C/min) starting from 96 °C to 4 °C.

3.3.3 pH buffer preparation

100 mM Phosphate buffer with different pH ranging from 5.8 to 8 with 0.1 intervals were prepared using 0.2 M sodium phosphate monobasic (Sigma) and 0.2 M sodium phosphate dibasic (Sigma) solutions. All solutions were made in RNase/DNase free water (Gibco) and filter-sterilized (0.2 μm) afterward. To determine the effective pH in the final alkaline cleavage reaction assay, phosphate buffer, TE buffer (used in substrate preparation), and AID storage buffer (used in GST-AID purification) were mixed to their final ratio of 6:1:3 and final pH was measured. Table 3-1 illustrates the pH solutions used in this thesis.

Table 3-1: pH solutions used in this thesis

	In 50 ml final solution		pH		
	Sodium phosphate monobasic (ml)	sodium phosphate dibasic (ml)	Aim	Measured stock	Measured effective
1	23.375	1.625	5.7	5.65	5.94
2	23	2	5.8	5.81	6.02
3	22.5	2.5	5.9	5.90	6.10
4	21.925	3.075	6.0	6.00	6.13
5	21.25	3.75	6.1	6.10	6.25
6	20.375	4.625	6.2	6.20	6.33
7	19.375	5.625	6.3	6.31	6.42
8	18.375	6.625	6.4	6.40	6.50
9	17.125	7.875	6.5	6.50	6.59
10	15.625	9.375	6.6	6.61	6.70
11	14.125	10.875	6.7	6.70	6.79
12	12.75	12.25	6.8	6.81	6.89
13	11.25	13.75	6.9	6.92	6.99
14	9.75	15.25	7.0	7.03	7.10
15	8.25	16.75	7.1	7.13	7.21
16	7	18	7.2	7.25	7.33
17	5.75	19.25	7.3	7.36	7.45
18	4.75	20.25	7.4	7.47	7.56
19	4	21	7.5	7.56	7.66
20	3.25	21.75	7.6	7.66	7.77
21	2.625	22.625	7.7	7.77	7.89
22	2.125	22.875	7.8	7.89	7.99
23	1.75	23.25	7.9	7.97	8.08
24	1.325	23.675	8.0	8.12	8.20

3.3.4 Biochemical analysis of purified GST-AID

To investigate the full spectrum of the biochemical properties of purified wild type and mutant Gm-AID, optimal temperature, optimal pH, time course, substrate specificity, enzyme kinetics, global ssDNA binding, and activity on 5-methylated cytidine (5-mC) were explored using established assays (Abdouni et al., 2013; Dancyger et al., 2012; Larijani et al., 2007). Experiments were done using standard alkaline cleavage assay where between three to four independent protein preparations of GST-AID homologs and mutants were tested in one to four replicates. This assay is an effective tool to examine various biochemical properties of wildtype and mutant AIDs in a time-efficient manner.

In the standard alkaline cleavage assay (Figure 3-1), the radiolabeled substrate was incubated with purified GST-AID protein in phosphate buffer (Abdouni et al., 2013; Abdouni et al., 2018; Dancyger et al., 2012; Emma M. Quinlan, 2017; King et al., 2015; Larijani & Martin, 2007). Reactions were incubated at the AID's optimal temperature and pH for different time length depending on the activity of each GST-AID homologs. To halt the GST-AID activity, samples were incubated at 85 °C for 20 min. To remove AID-generated uracil from substrate, Uracil-DNA glycosylase enzyme (UDG, NEB) and its corresponding buffer were added to each reaction followed by incubation at 37 °C. The remaining abasic site was cleaved by incubating the reactions at 96 °C for 10 min in the presence of 100 mM NaOH. To separate the substrate from product, reactions were electrophoresed on a 14% denaturing acrylamide gel. To visualize the result, gels were exposed to a Kodak Storage Phosphor Screen GP (Bio-Rad Laboratories, Inc.) and imaged using a PhosphorImager (Bio-Rad, Hercules, CA, USA).

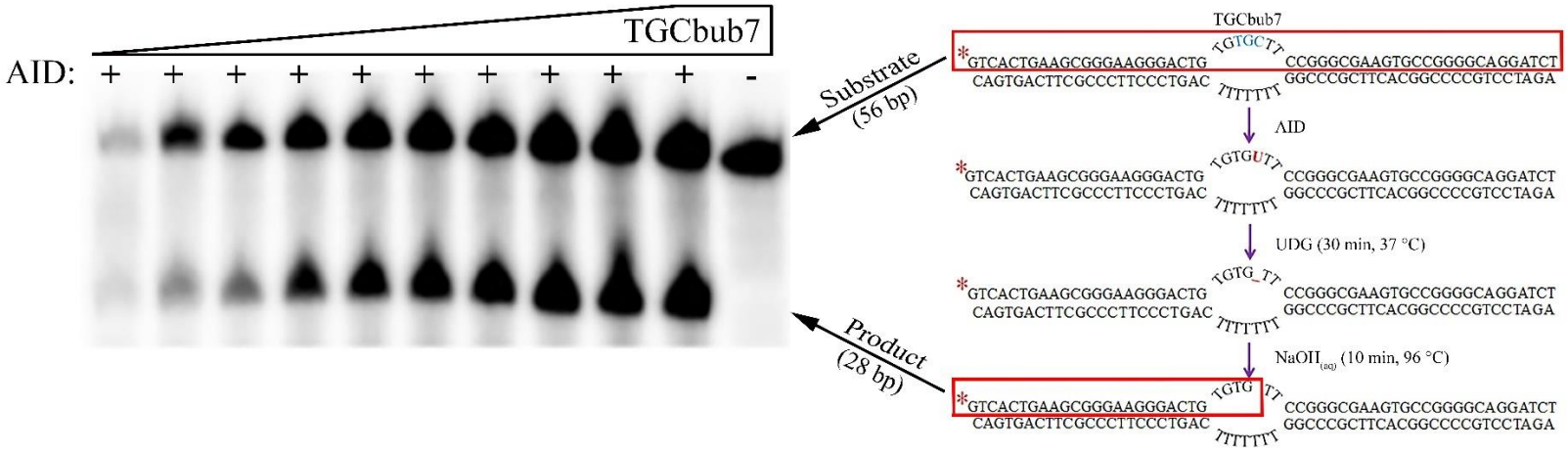


Figure 3-1: Experimental scheme for standard alkaline cleavage assay. TGCbub7 denotes a substrate bearing the WRC motif TGC located in a seven-nucleotide-long bubble region. The right panel shows the scheme for the standard alkaline cleavage assay. The left panel shows a representative denaturing acrylamide gel. The AID activity is reported as the percentage of initial substrate which was converted into product.

To determine the optimal temperature of Gm-AID, 3 µl of purified GST-AID was incubated with 25 fmol of ^{32}P-labelled TGCbub7 substrate at various temperature points (4 °C to 40 °C) in phosphate buffer with effective pH of 7.3. Ip-AID, Dr-AID and Hs-AID were tested alongside as controls. A time course experiment was also performed at three temperatures (low, optimal, high) for Gm-AID (4 °C, 8 °C and 18 °C; 12 time points; 30 min to 73 h), Hs-AID (8 °C, 31 °C and 40 °C; 19 time points; 1 min to 70 h), and Dr-AID (4 °C, 25 °C and 37 °C; 19 time points, 30 sec to 48 h) to confirm the results.

To assess the optimal pH of each GST-AID homolog and mutants, 25 fmol of ^{32}P-labelled TGCbub7 substrate was incubated with 3 µl of GST-AID preparation in 6 µl of phosphate buffer with effective pH ranging from 5.9 to 8.2 (24 or 12 pH points) in final volume of 10 µl. Depending on the activity of each GST-AID, reactions were incubated at optimal temperature for different time length ranging from 40 min to 96 h.

To investigate substrate sequence specificity of Gm-AID, ^{32}P-labelled WRCbub7 (TGC, TAC, and AGC) or ^{32}P-labelled non-WRCbub7 substrates (GGC, GTC, and GAC) were incubated with 3 µl of GST-AID homologs at their optimal temperature and pH. Gm-AID, Ip-AID, Dr-AID, and Hs-AID were incubated for 96 h, 10 h, 20 min, and 3 h, respectively. To investigate the effect of temperature on substrate specificity, Gm-AID and Hs-AID were incubated at a lower and a higher temperature than their optimal point as well. To explore any possible enzymatic role of Gm-AID in epigenetics, the activity of Gm-AID on the substate containing 5-methylcytosine (5-mC) was compared to that of other AID homologs. Deamination activity of GST-AID on 5-mC was studied using ^{32}P-labelled TG(mC)bub7, AG(mC)bub7, and GG(mC)bub7 which are substrates that contain

a target 5-mC rather than dC (Abdouni et al., 2013; Larijani & Martin, 2007; Sohail et al., 2003). In the case of substrate containing 5-mC, AID activity would generate thymidine (dT). Briefly, 50 fmol of substrate was incubated with 3 µl of GST-AID in phosphate buffer in the final volume of 10 µl at their optimal temperature and pH for different times depending on the activity of each AID homologs. Reactions were then incubated at 85 °C for 20 min to halt AID activity. To create a G:T mismatch double-stranded substrate, 40-fold excess of a fully complementary strand was added to each reaction in the presence of 50 mM KCl. Samples were then annealed by incubation at 96 °C for 5 min followed by slow cooling (*i.e.*, 1 °C/min) starting from 96 °C to 4 °C. To excise the dT from the G:T mismatch, one unit (U) of Thymine-DNA glycosylase (TDG, Trevigen, UK) and its corresponding buffer was added to each reaction followed by overnight incubation at 65 °C. The incubation of the reactions at 96 °C in the presence of 100 mM NaOH was done to cleave the abasic site. Samples were electrophoresed on a 14% denaturing gel and the results was visualized as described for the standard alkaline cleavage assay. AID activity on the corresponding standard substrates (*i.e*, containing dC) was carried out alongside as controls. The results were reported as the ratio of AID activity on the standard substrate compared with methylated ones.

To compare the catalytic rate of AID homologs and mutants through Michaelis-Menten kinetics, 3 µl of purified GST-AID were incubated with a 0.03125-600 fmol range (18 points) of ^{32}P-labelled TGCbub7 substrate at their optimal temperature and pH. The results of the time course experiments were used to estimate the proper incubation time for each AID homolog and mutant to ensure that the AID activity was within its initial velocity.

Enzymatic velocity (fmol of deaminated product/min of incubation/µg of AID) were plotted against substrate concentration (nM). To estimate K_{cat}, K_m and V_{max} parameters, the data was fitted into $Y = Et \times K_{cat} \times X/(K_m+X)$ equation. This equation is a modified version of Michaelis-Menten kinetics where the K_{cat} can be calculated as well. In this equation, Y is the enzyme velocity, X is the substrate concentration, Et is the concentration of enzyme catalytic sites, K_{cat} is the number of times each enzyme site converts substrate to product per unit time (*i.*e, the turnover number), and K_m (*i.e.*, the Michaelis-Menten constant) is the substrate concentration needed to achieve a half-maximum enzyme velocity (*i.e.*, V_{max}). Since AID has one catalytic pocket, its Et is equal to the concentration of enzyme used in the experiment. To estimate the Et, the molecular weight of the GST-AID homologs and mutants were calculated using Protein Molecular Weight web-based application (https://www.bioinformatics.org/sms/prot_mw.html).

Global ssDNA binding affinity of Gm-AID isoforms were compared to other AID homologs using electrophoretic mobility shift assay (EMSA) (Larijani et al., 2007). A 0.025-2.5 nM range of ^{32}P-labelled TGCbub7 was incubated with 0.9 µg of purified GST-AID in binding buffer (50 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT in 100 mM Phosphate buffer pH 7.21) for 1 h at their optimal temperature. Samples were then UV cross-linked on ice and electrophoresed on an 8% acrylamide native gel at 4 °C. Results were plotted as fmol bound substrate against nM of free substrate. To estimate half-saturation values, data was fitted into $Y=B_{max} \times X/(K_d+X)$ equation where Y is the concentration of bound fraction, X is the concentration of free fraction, B_{max} is the maximum concentration of bound fraction and K_d is the binding affinity of GST-AID for the substrate.

3.3.5 Data collection and quantification

Quantification was done using Image Lab software (version 6.0.1 build 34, Standard Edition, Bio-Rad Laboratories, Inc.) to perform densitometry. Data were plotted as the arithmetic mean using GraphPad Prism 5 software (version 5.00, GraphPad Software, Inc., USA) and error bars were set to represent standard error (SEM). Each point on an enzyme assay plot corresponds to the arithmetic mean of 4 to 12 data points. Where appropriate, maximum percentage of deamination activity was calculated by dividing each data point by the maximum absolute value for each data set to simplify the comparison between AID homologs with different deamination activity. The statistical significance of the qPCR results was analyzed using one-way ANOVA (IBM SPSS Statistics 20, IBM Corp.). For WRC specificity, the statistical significance of the results was analyzed using nonparametric independent samples test (IBM SPSS Statistics 20, IBM Corp.).

3.3.6 PCR-based AID activity assay

To compare the deamination activity of AID homologs on various DNA sequence and secondary structure, a deamination-specific PCR-based assay was conducted (Emma M. Quinlan, 2017; Larijani, Frieder, Basit, et al., 2005). Briefly, 200 ng of the substrate plasmid was denatured at 98 °C for 10 min in 100 mM phosphate buffer. Four microliters of purified AID and 1^{-3} U of UDG inhibitor (UGI, New England Biolabs) were added to each reaction after snap-cooling in an ice bath (final volume of 10 µl). Samples were incubated for various time-points ranging from 1 to 16 h at optimal conditions of each AID homolog. To detect AID-mutated plasmids, nested-PCR using deamination-specific primers (Table 3-2) was performed on serially diluted reactions (1/2 to 1/1000000). One µl

of each dilution was amplified under an initial denaturation step for 3 min at 96 °C followed by 30 cycles of [30 sec at 96 °C; 30 sec at 58 °C; and 1 min at 72 °C] and 10 min at 72 °C. One µl of primary PCR product was then amplified under the same cycling conditions except for using 57 °C for the annealing step. PCR products were analyzed on a 1.2% agarose gel.

Table 3-2: Deamination-specific primers used in this chapter

Gene		Direction	Primer sequence (5' to 3')	Amplicon size (bp)
Deamination-specific primers[i]	First PCR	Forward	GGGATATAGGGGTTTTTTGAGGTTTGGTATTATTTAAAT	548
		Reverse	ACACAACCAACTTTCATTCCAACCACAAACTTTCAATA	
	Nested PCR	Forward	CTTATCTTGGTTCTGTGGCAACCGACTGCCTGCTAACAGG	442
		Reverse	CCAACTTTCATTCCAACCACAAACTTTCAATAAATT	

[i]: The primer sequences for this gene are modified to specifically amplify heavily-C-to-U-mutated sequence

3.3.7 Structure prediction and AID-DNA binding simulations

We employed a similar structure prediction approach, as described in section 2.3.9 (King & Larijani, 2017; King et al., 2015; Zhu et al., 2015). The recently published human AID crystal structure was chosen as template for homology modeling: MBP fused AID (PDB: 5W0Z), MBP fused AID in complex with cytidine (PDB: 5W0C), MBP fused AID in complex with dCMP (PDB: 5W0U), and MBP fused AID in complex with cacodylic acid (PDB: 5W0R) (Qiao et al., 2017). The template AID structures were obtained from the protein databank (http://www.rcsb.org) and visualized using PyMOL v1.7.6 (http://www.pymol.org/). Using the default parameters of I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/), 200 models were constructed for AID homologs of which the best open conformations (refer to section 1.5.6) were chosen (Roy et al., 2010; Yang et al., 2015; Zhang, 2008). Ramachandran plots were created using Rampage and used to evaluate the quality of the proteins on an individual residue basis based on their stereochemical angles (Lovell et al., 2003).

The catalytic pocket was defined by the indented space containing the Zn-coordinating and catalytic residues (Hs-AID: H56, E58, C87 and C90; Dr-AID: H60, E62, C99 and C102; Ip-AID: H59, E61, C98 and C101; Gm-AID: H60, E62, C100, C103). The catalytically accessible models were defined by accessibility of catalytic glutamate to the surface of the protein. To simulate AID-DNA binding, DNA substrate was docked to each AID model using AutoDock Vina (Trott & Olson, 2010). The substrate was constructed using ChemDraw Prime v.16.0 (http://www.cambridgesoft.com/software/overview.aspx) and Marvin Sketch v.5.11.5 (http://www.chemaxon.com/products/marvin/marvinsketch/),

while surface topology and docking parameters were generated using Swiss-Param (http://swissparam.ch) (Zoete et al., 2011). 5'-TTTGCTT-3' ssDNA was chosen as our substrate, since it has been shown to be the preferred substrate of human and bony fish AID (Emma M. Quinlan, 2017). For each AID homolgs, five models with open conformation were selected for DNA docking. For each model, 20 docking trials were conducted, producing 8 conformations in each trial. In docking trials, we restricted the ssDNA binding within 30×30×30 Å (x, y, z coordinates) from the Tryptophan of the Loop 6. Each model was docked with a substrate. UCSF chimera v.1.11.2 (https://www.cgl.ucsf.edu/chimera) was used to view the conformations of substrate, and its interactions with AID models (Pettersen et al., 2004). Deamination-conducive AID-DNA complexes were defined by the accessibility of the NH_2-group of dC to the catalytic Zn-coordinating and glutamic acid residues. To analyze the interaction of each nucleotide with AID model, PyMol was used to measure amino acid residues within 4Å of the nitrogenous base and the 1st carbon of the deoxyribose sugar.

3.3.8 Characterization of the Atlantic cod *IgV_H* region and

A partial immunoglobulin heavy chain locus of the Atlantic cod has previously been characterized (GenBank identifier: AJ871288.1). This sequence was aligned with BLAST against the improved version of the Atlantic cod genome (gadMor2) using default parameters of blastn task in BLAST+ program (Torresen et al., 2017). Complete protein sequences for IgM, IgD, and IgZ from GenBank were extracted to perform tblastn against the gadMor2 genome (Appendix 1). Possible constant regions were identified manually from blast results, extracted from the genomic sequence, and a reciprocal blast was

performed towards GenBank (blastx) to verify annotation. All sequences extracted from AJ871288.1 and gadMor2 genome were compared, where the annotation from AJ871288.1 was preferred.

3.3.9 WRC/GYW and WGCW motif analysis

For WRC motif analysis, Japanese puffer fish IgV_H (*Tr-*IgV_H), and nurse shark IgV_H (*Gc-*IgV_H) sequences were obtained from NCBI (Appendix 6). The nurse shark complementarity-determining regions (CDRs) were mapped from *Tr-Ig* gene variable regions (Fu et al., 2017; Fu et al., 2015). *Hs-*IgV_H, mouse IgV_H (*Mm-*IgV_H), chicken IgV_H (*Gg-*IgV_H), South African toad IgV_H (*Xl-*IgV_H), *Ip-*IgV_H, salmon IgV_H (*Ss-*IgV_H), and *Dr-*IgV_H sequences were obtained from IMGT (the international ImMunoGeneTics information system) database (http://www.imgt.org/) (Giudicelli et al., 2005; Lefranc, 2001, 2003; Lefranc, Clement, et al., 2005; Lefranc et al., 2015; Lefranc et al., 1999; Lefranc et al., 2009; Lefranc, Giudicelli, et al., 2005; Ruiz et al., 2000). For these sequences, the CDRs and framework regions (FRs) were identified using IMGT database. In these analyses, the number of motifs were counted in each region using Python (Version 3.8) (Van Rossum & Drake, 2009). For WRC/GYW motifs TGC, TAC, AGC, AAC, GCA, GTA, GCT, and GTT and for WGCW motifs AGCA, AGCT, TGCA, and TGCT were counted. Then, the sum of WRC/GYW or WGCW motifs for each region was divided to the number of nucleotides analyzed for that given region to normalize for the variation in the length of each region. The average of these normalized WRC/GYW or WGCW indexes were calculated for CDRs and FRs. The enrichment of the motifs in CDRs was estimated by dividing the average index of CDR 1 and 2 by the average index of FR 1, 2, and 3. Also,

the GC content of the coding sequences was retrieved from Codon and Codon-Pair Usage Tables (CoCoPUTs) server (Alexaki et al., 2019). This database is available on https://hive.biochemistry.gwu.edu/review/codon2.

3.4 Results

3.4.1 Atlantic cod AID extreme cold adaptation and lethargic activity

To investigate the functional properties of Gm-AIDs, we expressed and purified Gm-AID and T-Gm-AID as N-terminally tagged GST fusion proteins (Figure 3-2) (Abdouni et al., 2013; Dancyger et al., 2012; Emma M. Quinlan, 2017; Larijani et al., 2007). We first sought to determine whether Gm-AID is an active cytidine deaminase using the standard alkaline cleavage assay. The seven nucleotide long partially single-stranded oligonucleotide bubble substrate TGCbub7 is the most favored substrate for all studied bony fish AIDs thus far (Emma M. Quinlan, 2017). Therefore, we tested Gm-AID activity on TGCbub7 substrate in the alkaline cleavage assay, which is the gold standard assay to measure AID/APOBEC cytidine deamination activity (Abdouni et al., 2013; Dancyger et al., 2012; Emma M. Quinlan, 2017; Larijani et al., 2007). Following initial experiments with overnight incubation of Gm-AID with TGCbub7 at 18, 25, and 37 °C, Gm-AID appeared to lack enzymatic activity (Figure 3-3 A). Considering that bony fish AIDs vary in their optimal temperature, we incubated Gm-AID at a wider temperature range (10 to 40 °C) for longer incubation periods (Figure 3-3 B). Since the temperature profile of Hs-AID and Ip-AID have been studied before, we also tested these AID homologs alongside of Gm-AID as controls (Figure 3-3 C). Since incubation time used in these experiments were longer than optimal for Hs-AID and Ip-AID, their temperature profiles obtained here are not consistent with their true temperature profile. Overall, these experiments revealed that Gm-AID was an active, but weak cytidine deaminase.

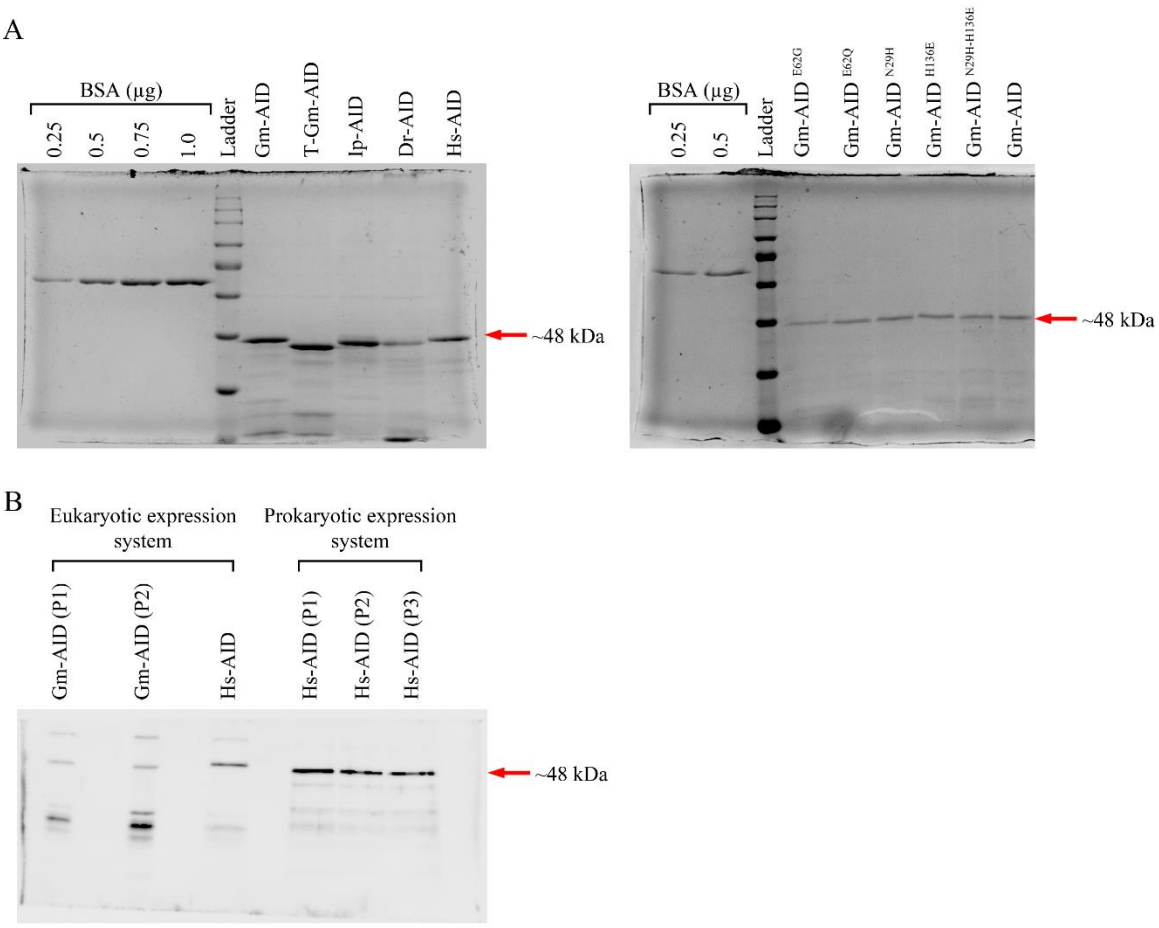


Figure 3-2: AID purification in prokaryotic and eukaryotic expression systems. A) Representative coomassie-stained SDS protein electrophoresis gels. After expression of GST-AID in bacteria (E.coli) and purification of GST-AID protein through GST affinity column, purity and yield of GST-AID were assessed by coomassie-stained SDS protein electrophoresis in comparison to BSA standards. B) A representative western blot. After expression of GST-AID in human embryonic kidney cells 293 (HEK239T) and purification of GST-AID protein through GST affinity batch binding, purity and yield of GST-AID were assessed using western blotting. AID was probed with anti-GST (SantCruz) antibodies, followed by the secondary detection by Goat anti-Rabbit IgG (SantaCruz).

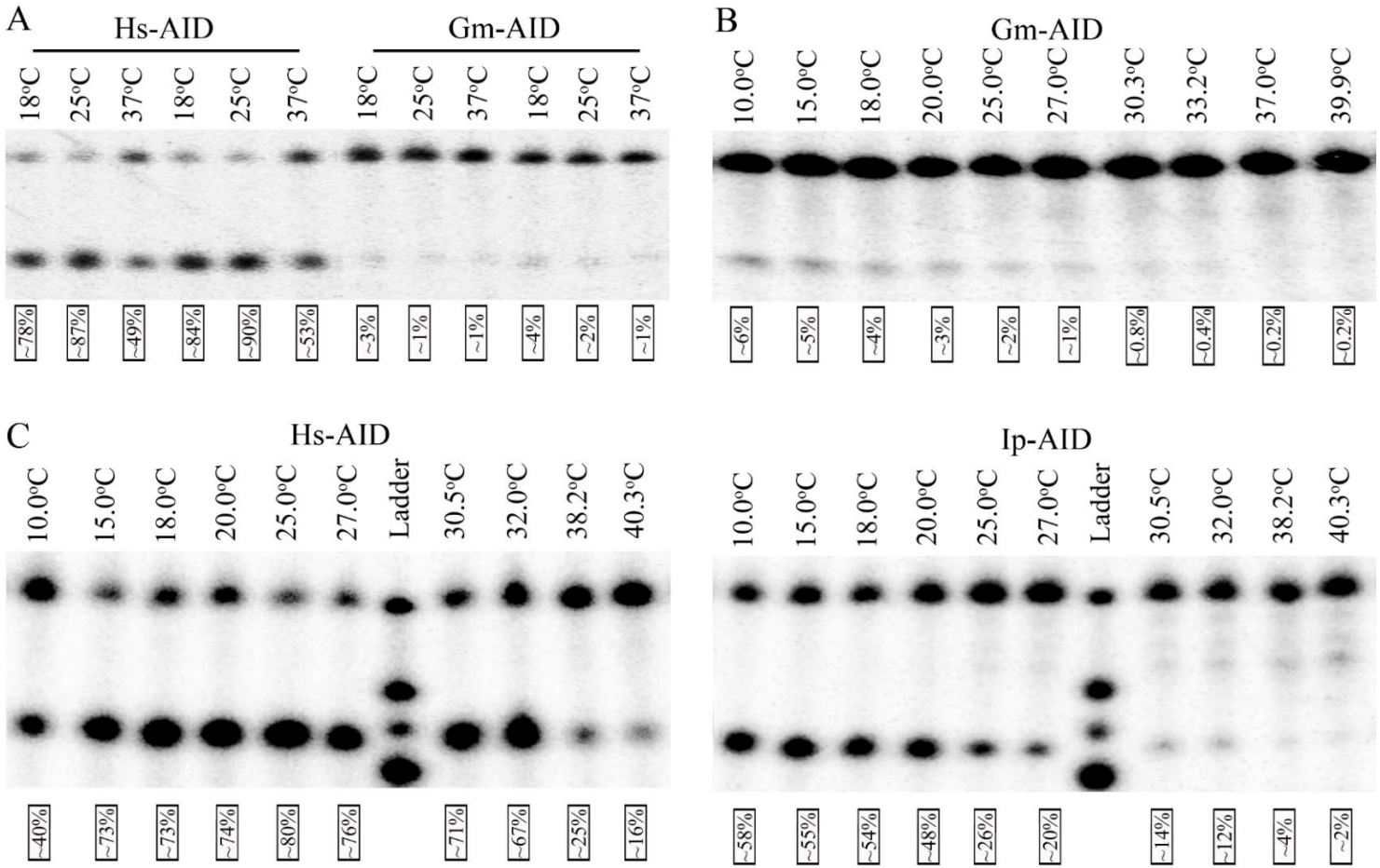


Figure 3-3: Functional analysis of purified Atlantic cod AID. Gm-AID was expressed and purified alongside other AID homologs as a GST-fusion protein and tested for cytidine deamination activity using the standard alkaline cleavage assay. Deamination activity (%)(%) is presented below each lane. All experiments were done using ^{32}P-labelled TGCbub7 substrate in duplicate. A) Purified Gm-AID and Hs-AID were incubated with substrate at 18, 25 and 37°C for 16 h showing barely detectible deamination activity for Gm-AID. B) Sixteen-hour prolonged incubation of purified Gm-AID with substrate revealed a preference for lower temperatures. C) Hs-AID and Ip-AID activity on ^{32}P-labelled TGCbub7 substrate at various incubation temperature points were tested alongside of Gm-AID as controls.

We then examined the temperature and pH profile of Gm-AID. To determine the exact optimal temperature and pH of Gm-AID, it was incubated with TGCbub7 at fine temperature increments (4 to 40 °C) in phosphate buffer with pH of 7.3. As controls, we tested Dr-AID, Ip-AID, and Hs-AID, whose temperature sensitivity profiles are well established (Dancyger et al., 2012; Emma M. Quinlan, 2017). As expected, Ip-AID, Dr-AID, and Hs-AID showed optimal temperature of 14, 25 and 31 °C, respectively; however, Gm-AID was most active at 8 °C. As expected, at 4 °C Hs-AID is completely inactive, and the activity of Dr-AID and Ip-AID are significantly reduced, whilst Gm-AID strikingly maintains a near optimal activity level (Figure 3-4 A).

We then studied the optimal pH of AID homologs in buffer with effective pH ranging from 5.9 to 8.2. We found the optimal pH of about 7.3, 7.6, 7.9, and 8.1 for Hs-AID, Dr-AID, Ip-AID, and Gm-AID, respectively (Figure 3-4 B). These measured optimal temperature and pH were used in all the experiments hereafter. To further confirm the cold adaptation of Gm-AID, time course enzyme kinetics were carried at optimal, higher, and lower than optimal temperatures (Figure 3-4 C). Gm-AID activity continued to increase at 8 °C even after 72 hours, confirming 8 °C as the optimal temperature of this AID. The faster increase in deamination activity of Gm-AID in the beginning (the first 20 hours) at 18 °C compared to 8 °C and the plateau of the activity at 18 °C after initial 24 hours suggest that Gm-AID is less structurally stable at 18 °C than at 8 °C. The continuous increase in AID activity at 8 °C is consistent with this being the optimal temperature of Gm-AID. These results indicate that Gm-AID is a cold-adapted enzyme, exhibiting the coldest optimal temperature reported for a vertebrate DNA/RNA-editing enzyme.

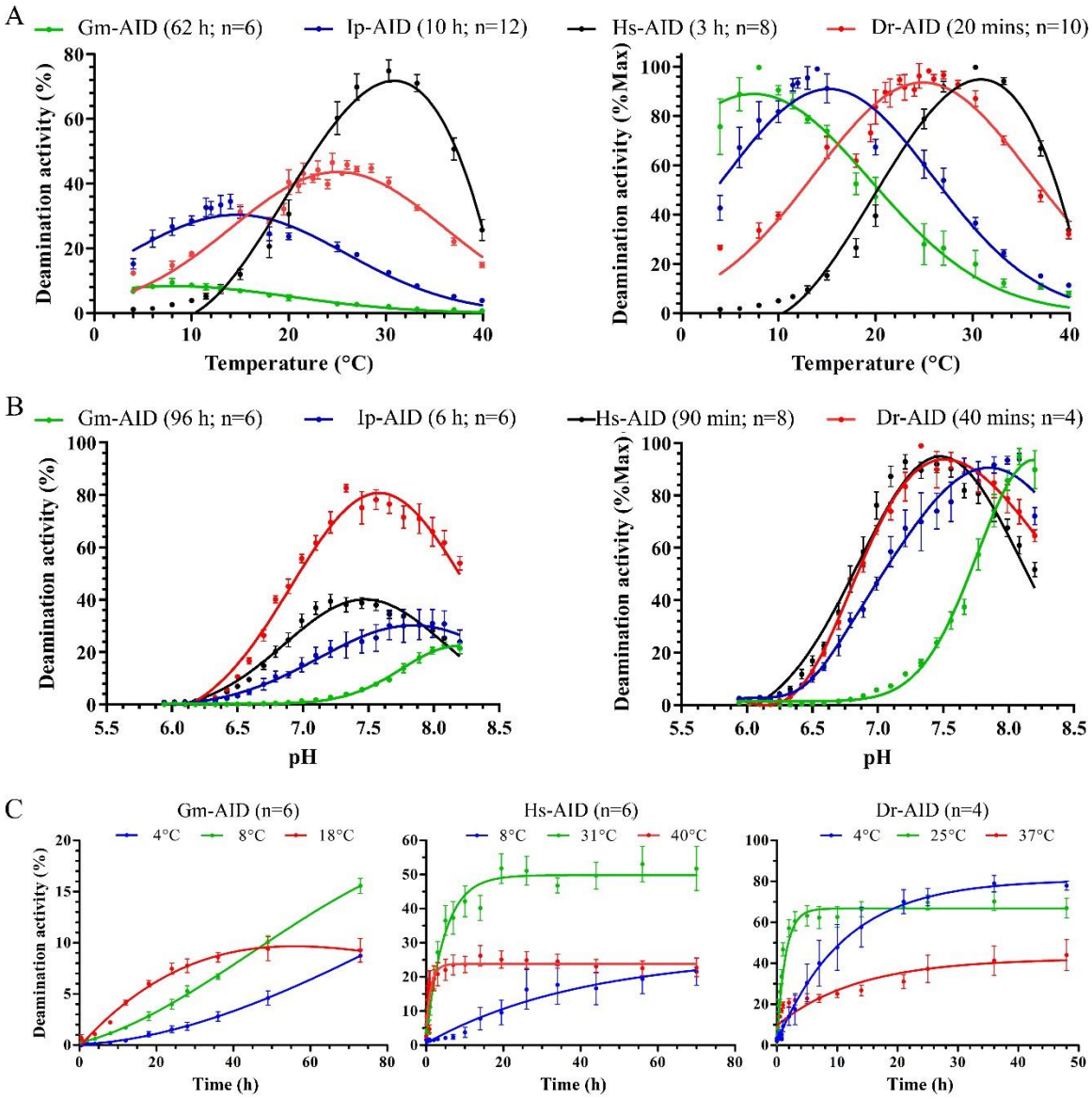


Figure 3-4: Atlantic cod AID Optimal temperature and pH. A) Optimal temperature of Gm-AID was compared to that of other AID homologs at fine temperature increments (4 to 40°C). Three to four independent protein preparations of each AID homolog were tested. Results are plotted as deamination activity percentage (left panel) and percentage of maximum deamination activity (right panel), revealing optimal temperature of 8, 14, 25 and 31 °C for Gm-AID, Ip-AID, Dr-AID, and Hs-AID, respectively. B) Optimal pH of Gm-AID was examined in phosphate buffer with effective pH ranging from 5.9 to 8.2. Results are plotted as deamination activity percentage (left panel) and percentage of maximum deamination activity (right panel). Results indicated that Gm-AID is the most basic-adapted AID reported in this chapter with the optimal pH of 8.1. The optimal pH of Hs-AID, Dr-AID, and Ip-AID were reported as about 7.3, 7.6, and 7.9. C) Time course enzyme kinetic assay was conducted at three temperature points (optimal, below, and above optimal) and corresponding optimal pH of each AID homolog. Three independent preparations of Gm-AID (30 min to 73 h), Hs- AID (1 min to 70 h), and Dr-AID (30 sec to 48 h) were tested in duplicate (n=6). Data is represented as mean ± SEM

Since Gm-AID exhibited extremely weak cytidine deaminase activity only after unusually long incubation periods, we sought to verify that the weak activity was indeed *bona fide* cytidine deaminase catalytic activity. To this end, we generated two independent Gm-AID mutants lacking the catalytic glutamate (E62). Comparison of deamination activity of wildtype Gm-AID to that of Gm-AID^{E62G} and Gm-AID^{E62Q} showed that these mutations indeed abolished deamination activity of Gm-AID (Figure 3-5 A). We also tested T-Gm-AID using our standard alkaline cleavage (Figure 3-5 B) and the PCR-based deamination assays (Figure 3-6 B) and we did not observe any consistent cytidine deamination activity. This result was expected due to truncation of substantial portion of the enzyme (21 amino acids) from its N-terminus in T-Gm-AID.

These data were obtained with bacterially expressed and purified GST-AID. To verify that the obtained result is not due to bacterial expression system, we expressed Gm-AID in a eukaryotic expression system (293T cells), along with Dr-AID as a positive control (Figure 3-2 B). We confirmed that in eukaryotic expression system, Gm-AID exhibits no detectable cytidine deamination activity (Figure 3-5 C). In the eukaryotic expression system, even Dr-AID showed much less cytidine deaminase activity (< 20 % within 16 h incubation at 18 °C) compared with the bacterially expressed Dr-AID (~ 40% within 40 min of incubation at 18 °C). Taken together, these data indicate that Gm-AID, even in optimal conditions, is an ineffective cytidine deaminase.

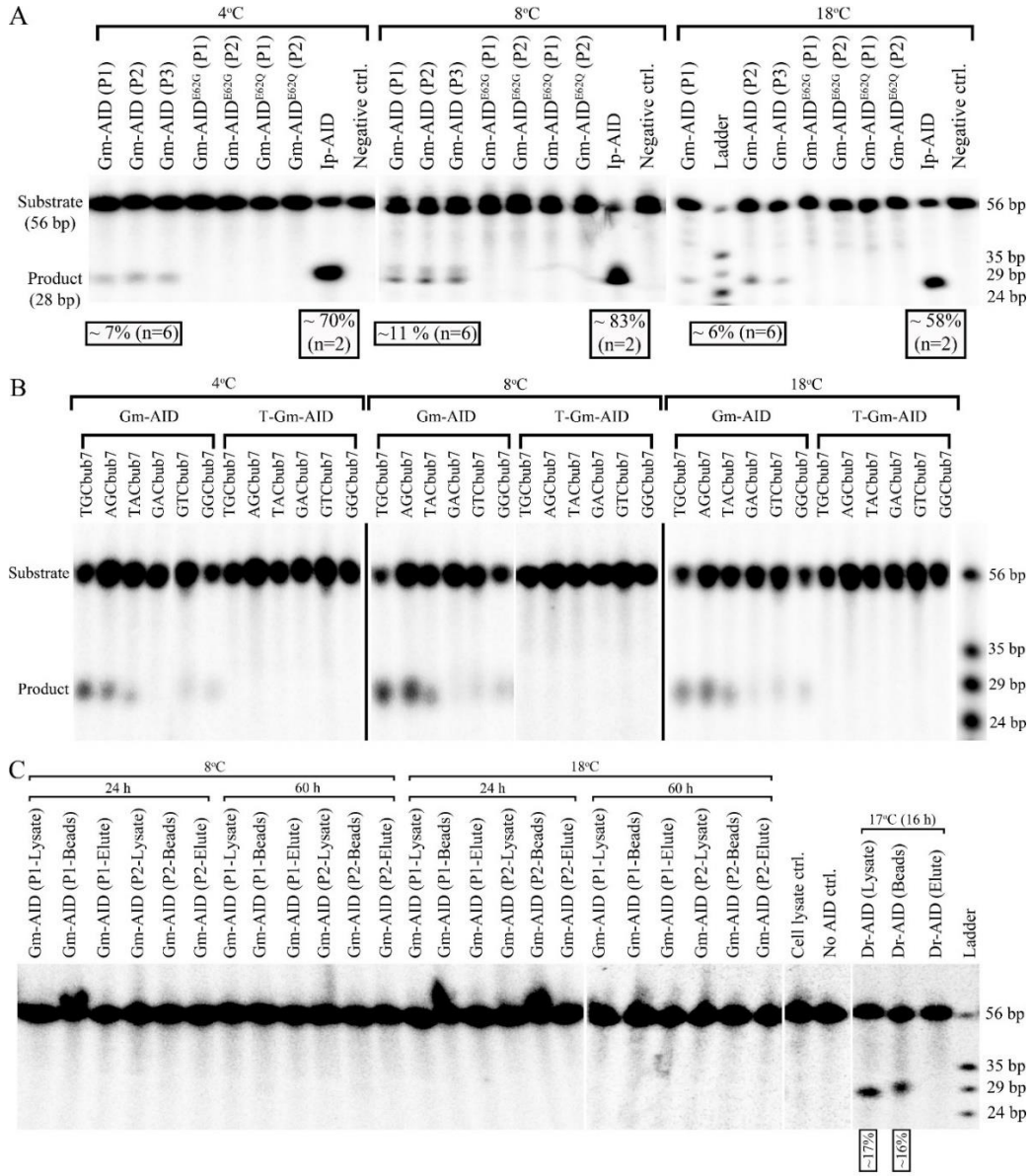


Figure 3-5: Bona fide cytidine deaminase activity of Atlantic cod AID. A) To confirm that the unusually low activity of Gm-AID is bona fide cytidine deaminase activity, wild type Gm-AID catalytic activity was compared to that of two mutants targeting essential catalytic pocket cytidine deamination residues (E62G and E62Q). Two independent protein preparations of each mutant were tested at 4, 8, and 18 °C for 63 h. Ip-AID was tested as a positive control. B) To assess activity of T-Gm-AID, it was incubated with various substrates containing WRC or non-WRC motifs for 96h. Gm-AID was used as control. Two protein preparation of each AID homolog were tested in duplicate at three different temperature point (optimal, below, and above optimal; n=6). C) To exclude the effect of expression system in our analysis, we expressed AID in HEK 293T cells. The cytidine deaminase activity of GST-AID expressed in this system was studied using alkaline cleavage assay. GST-AID was analyzed in the form of cell lysate, purified on GST beads, or eluted from GST beads. AIDs were incubated with TGCbub7 substrate for various time point at 8 or 18°C.

Next, we sought to qualitatively compare the cytidine deaminase capability of Gm-AID to that of other AID homologs through two independent approaches. First, we used our standard alkaline cleavage assay for measuring cytidine deamination by AID/APOBECs to conduct standard Michaelis-Menten (MM) kinetics to compare the catalytic parameters of AID homologs. The MM kinetics describes the enzymatic reaction rate as a function of substrate concentration using the catalytic rate constant (K_{cat}) and the Michaelis-Menten constant (K_m) (Berg et al., 2002; Choi et al., 2017; Roskoski, 2015). K_{cat} is the turnover number of an enzyme and is defined at the number of substrate molecules converted into product by an enzyme molecule in a unit time when the enzyme is fully saturated with substrate (Berg et al., 2002; Choi et al., 2017; Roskoski, 2015). The K_{cat} value for most enzymes is between 1 to 10^4 S^{-1} (Berg et al., 2002). K_m, is an important characteristic of an enzyme and is equal to the substrate concentration at which the reaction rate is half of the maximum rate (V_{max}). V_{max} is reached when the enzyme's catalytic site(s) is saturated with substate (Berg et al., 2002; Choi et al., 2017; Roskoski, 2015). For most enzymes, K_m is between 10^{-1} to 10^{-7} M (Berg et al., 2002). K_m value is determined for a given pair of enzyme and substrate and depends on the environmental conditions such as pH, temperature, and ionic strength. K_m provides a measure of the binding affinity of the enzyme for its substrate, and in the case of AID, because the enzyme has many non-catalytic bindings, the measure of the enzyme's catalytic pocket affinity for the substrate dC (Berg et al., 2002). The K_{cat}/K_m ratio is a measure of catalytic efficiency of an enzyme where a perfect enzyme has a K_{cat}/K_m of $10^8 – 10^9$ $s^{-1}M^{-1}$ (Berg et al., 2002; Newton et al., 2015; Roskoski, 2015). It is important to note that using K_{cat}/K_m to compare the catalytic

efficiency of two enzymes has limitations such as two enzymes with different K_{cat} and K_m values could have the same K_{cat}/K_m (Newton et al., 2015). Therefore, all three values of K_m, K_{cat}, and K_{cat}/K_m should be considered when comparing different enzymes.

As a second independent method for assaying the relative enzymatic activity of Gm-AID, we used a PCR-based deamination assay which is a sensitive method for quantifying AID-mediated mutation levels on multi-kb-long DNA substrates (Emma M. Quinlan, 2017; Larijani, Frieder, Sonbuchner, et al., 2005). In this assay, a plasmid substrate was incubated with each purified GST-AID homolog/mutant at its corresponding optimal pH and temperature. The AID-treated plasmid substrate was then PCR amplified using deamination-specific primers to detect highly mutated plasmid substrate. To determine the relative amount of highly mutated DNA in each reaction, the AID activity assay reactions were serially diluted prior to being subject to deamination-specific PCR.

Consistent with previous findings (Barreto et al., 2005; Dancyger et al., 2012; King et al., 2015; Wakae et al., 2006), we observed that, under our experimental conditions, Dr-AID exhibited the highest catalytic rate, ~9-fold higher that Hs-AID, while the catalytic rate of Ip-AID showed ~13-fold lower activity than Hs-AID. The catalytic rate of Gm-AID, however, was orders of magnitude lower than all three: ~3100, 350, and 25-fold lower than Dr-AID, Hs-AID, and Ip-AID, respectively (Figure 3-6 A and Table 3-3).

As expected, no cytidine deamination activity was detected for T-Gm-AID and the catalytically-inactive mutants (*i.e.,* Hs-AID^{C90F} and Gm-AID^{E62G}), confirming the result of alkaline cleavage assay (King et al., 2015). Serial dilution analysis revealed Dr-AID to be 10-100-fold more active than Hs-AID, whilst Gm-AID supported 100- and 10,000-fold

less mutation levels than Hs-AID and Dr-AID, respectively (Figure 3-6 B). These data provide independent confirmation of relative activity levels obtained in the alkaline cleavage assay.

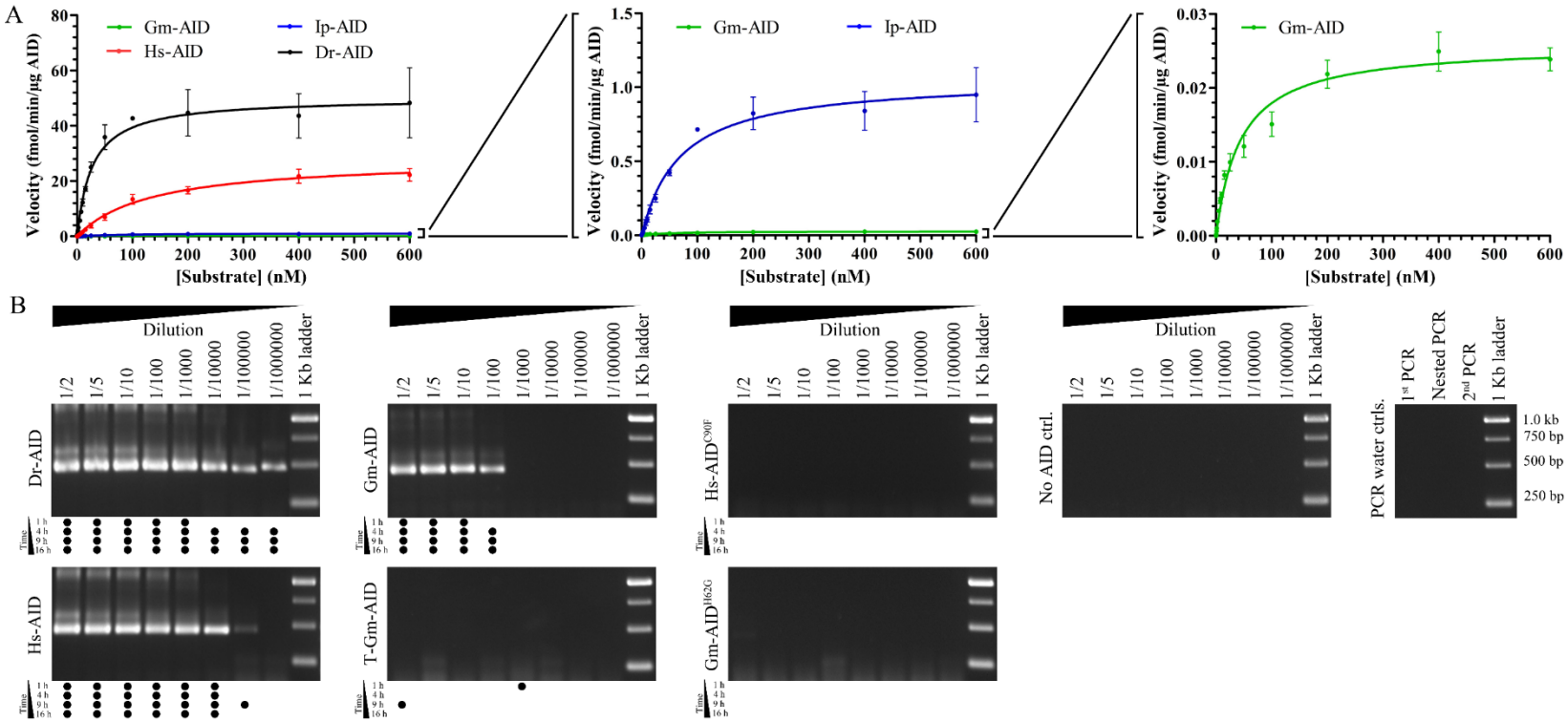


Figure 3-6: Comparison of the catalytic rate of Atlantic cod AID with other AID homologs. A) The catalytic rate of Gm-AID was compared to that of other AID homologs through Michaelis-Menten kinetics. Three independent protein preparations of each of the AID homolog were incubated at their optimal temperature with 0.03125-600 fmol range of TGCbub7 substrate. Each reaction was carried out in duplicate. Results revealed that Gm-AID's catalytic rate was ~3100, 350, and 25-fold lower than Dr-AID, Hs-AID, and Ip-AID, respectively. Data is represented as mean ± SEM (n=6). B) The relative catalytic activity of Gm-AID was confirmed through a PCR-based deamination assay using a single-stranded plasmid as the substrate. To assess AID activity on various ssDNA sequences and topologies, purified AID was incubated between 1 to 16 h with heat-denatured substrate plasmid. Each reaction was diluted up to 1/1000000. PCR was performed using deamination-specific primers that only amplify AID-mutated plasmids. The experiment was repeated 4 times, and the presence of a PCR band in each independent experiment was recorded as a black dot below each lane in the representative gel. Consistent with lack of cytidine deaminase activity in the alkaline cleavage assay, no activity was detected for the catalytically dead AIDs (Hs-AID^{C90F}, Gm-AID^{E62G} and T-Gm-AID). Comparison of the highest dilutions of the AID reaction in which a PCR band was detected showed that Gm-AID is approximately 100 and 10000-fold less active than Hs-AID and Dr-AID, respectively. No PCR band was detected in negative control reactions.

Table 3-3: Michaelis-Menten kinetics parameters measured for each AID homolog

	Temp (°C)	pH	K_{cat} (min^{-1})	K_m (nM)	V_{max} (fmol/min/µg)	Std. Error		R^2	K_{cat}/Km ($min^{-1}nM^{-1}$)	Activity ratio
						K_{cat} (min^{-1})	K_m (nM)			
Gm-AID	8	8.08	1.36E-06	44.05	0.02585	3.05E-08	3.421	0.9735	3.09E-08	1.00
Ip-AID	14	7.89	5.50E-05	68.77	1.058	1.62E-06	6.52	0.9675	8E-07	25.87
Hs-AID	31	7.31	0.001448	133.8	28.13	3.72E-05	9.465	0.9815	1.08E-05	350.01
Dr-AID	25	7.56	0.002612	27.16	50.08	8.31E-05	3.104	0.9543	9.62E-05	3110.37

Abbreviations: Gm-AID: Atlantic cod AID; Ip-AID: channel catfish AID; Hs-AID: human AID; Dr-AID: zebrafish AID.

3.4.2 Atlantic cod AID activity on methylated cytidine

A controversial role for AID in epigenetics reprograming has been suggested through deamination of 5-methylated cytidine (5-mC) leading to CpG motif demethylation (Bhutani et al., 2013; Dominguez et al., 2015; Moon et al., 2016; Popp et al., 2010; Rai et al., 2008). Studied AID homologs do not deaminate 5-mC with significant efficiency with the exception of Dr-AID (Abdouni et al., 2013; Larijani, Frieder, Sonbuchner, et al., 2005; Nabel et al., 2012; Wijesinghe & Bhagwat, 2012). We examined Gm-AID activity on substrate containing 5-mC instead of dC (Abdouni et al., 2013). As expected, TG(mC)bub7 was the most favorable methylated substrate for all studied AIDs (Figure 3-7). We found that unlike Dr-AID and Hs-AID, which could deaminate 5-mC situated in different sequence motifs, Ip-AID and Gm-AID demonstrated activity only on TG(mC)bub7 (Figure 3-7 A). Exploring dC/5-mC deamination efficiency over time showed that for all tested AIDs, the ratio of deamination activity on dC *vs*. 5-mC improves over time (Figure 3-7 B). This phenomenon is due to the higher catalytic efficiency of AID on dC *vs.* 5-mC (Abdouni et al., 2013). Although Gm-AID showed increased activity on 5-mC over time (1.5% at 72 h *vs*. 3.9% at 120 h), its activity on 5-mC was 5-fold and 2-fold lower than on dC at 72 h and 120 h, respectively.

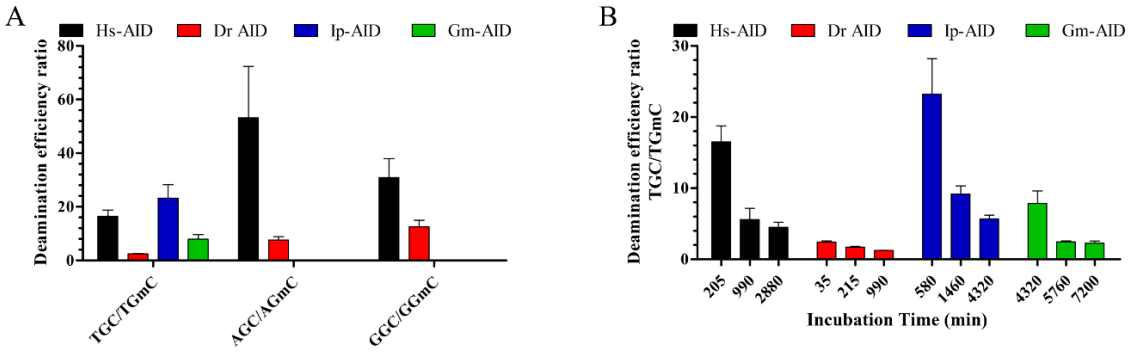


Figure 3-7: Atlantic cod AID activity on 5-mC. A) To examine AID deamination activity on methylated cytidine, activity of AID homologs on TGCbub7, AGCbub7, and GGCbub7 were compared to that of TGmCbub7, AGmCbub7, and GGmCbub7. Two independent protein preparations of each AID homolog were tested in duplicate at their corresponding optimal temperature (i.e., 31, 25, 14, and 8 °C for Hs-Aid, Dr-AID, Ip-AID, and Gm-AID, respectively). The incubation time was set as 30 min, 3 h, 10 h, and 96 h for Dr-AID, Hs-AID, Ip-AID, and Gm-AID. Consistent with previous publications, Dr-AID showed the highest efficiency in deaminating 5-mC; while, Ip-AID and Gm-AID showed low activity on 5-mC, only on TGmCbub7 substrate. B) To confirm the result of the previous experiment, AID activity on TGmCbub7 substrate was studied over 3 various time points for each AID homolog corresponding to their catalytic activity robustness. As expected, the ratio of deamination activity on dC vs. 5-mC improved over time, consistent with the higher catalytic efficiency of AID enzyme on dC. Results confirmed that Gm-AID does not have significant activity on 5-mC. Data is represented as mean ± SEM (n=4).

3.4.3 The basis of Atlantic cod AID lethargy

As mentioned in the previous chapter (section 2.4.5), comparison of Gm-AID primary sequence and computational models with those of other AID homologs revealed similar overall structure and the presence of a viable catalytic pocket. In this chapter, we examined the surface of Gm-AID and its ssDNA binding affinity. Gm-AID has a charge of +10.41 (at pH 7) which, like Hs-AID (charge of +10.25 at pH 7), ought to enable it to efficiently bind negatively charged ssDNA (Figure 3-8 A). We previously showed that all studied jawed vertebrate AID homologs bind ssDNA substrate with nM-range affinity (Dancyger et al., 2012; Emma M. Quinlan, 2017). We evaluated Gm-AID and T-Gm-AID ssDNA binding affinity using electrophoretic mobility shift assay (EMSA) and observed that both bind ssDNA with the same high nM-range affinity (Figure 3-8 B). Therefore, the extremely low catalytic rate of Gm-AID is not due to global ssDNA binding impairment.

EMSA provides a measure of global surface ssDNA binding by AID but only a minor fraction of ssDNA bound on AID's surface passes over its catalytic pocket and can be deaminated (King & Larijani, 2020; King et al., 2015). In other words, the majority of AID:ssDNA interactions result in catalytically non-productive enzyme:substrate complexes (King & Larijani, 2020; King et al., 2015). To evaluate specific ssDNA binding over the catalytic pocket, we performed docking simulations as used previously to discern AID binding to ssDNA target (King & Larijani, 2017; King et al., 2015). As has been shown for Hs-AID, here we observed two distinct ssDNA binding grooves on the surface of Gm-AID, Dr-AID, and Ip-AID (Figure 3-8 C and Figure 3-9) (Abdouni et al., 2018; Emma M. Quinlan, 2017; King et al., 2015; Qiao et al., 2017). However, we also noticed

alternative ssDNA:AID interactions in which substrate was highly solvent exposed. These alternative ssDNA binding modes resulted from involvement of α4 in interacting with ssDNA (Figure 3-8 C and Figure 3-9). Interestingly, we noted that the contribution of α4 in interaction with ssDNA was significantly increased in Gm-AID relative to other homologs: 21%, 6%, 6%, and 8% for Gm-AID, Ip-AID, Dr-AID, and Hs-AID, respectively (Table 3-4).

Docking simulations suggested two potential residues responsible for this phenomenon. First, we noticed a positive residue on ℓ2 in Hs-AID^{R25} and Dr-AID^{H29} which was replaced with a polar uncharged residue in Gm-AID^{N29} (Figure 3-8 D and E). This residue, located at the immediate surface perimeter or the "mouth" of the catalytic pocket, is important for efficient arching and positioning of dC into the catalytic pocket (Harjes et al., 2013; King & Larijani, 2017; King et al., 2015; Shi et al., 2017). Interestingly, it was confirmed that, in the crystal structure of human AID bound to dCMP, R25 interacts with 5′ phosphate of dC (Qiao et al., 2017). Second, we previously demonstrated that AID's catalytic pocket accessibility is determined by secondary catalytic residues that function as a supporting network to stabilize the target dC (King et al., 2015). In the previous chapter, we noted that although most secondary catalytic residues are conserved amongst AID homologs, Hs-AID^{E122} is uniquely different in Gm-AID^{H136} (Figure 3-8 A, D and E). This glutamic acid to histidine change in Gm-AID may favor the aforementioned interaction of α4 with ssDNA. Docking simulations revealed a 3- to 8-fold increase in interactions between Gm-AID^{H136} and the -1 position nucleotide upstream of the target dC, relative to the conserved glutamic acid of other species (Table 3-5). It was previously suggested that

the interactions between ℓ8 and the bases at the -1 and -2 positions (with respect to the dC) define the substrate specificity in the AID/APOBEC family (Gajula et al., 2014; Iyer et al., 2011; Kohli et al., 2009). Therefore, increased interactions between the Gm-AID^{H136} residue, which reside in α4, indirectly indicates that the interactions between Gm-AID ℓ8 and the -1 position nucleotide may be disrupted. Notably, perturbation of Hs-AID^{E122} results in a drastic reduction in activity consistent with its important role as a secondary catalytic residue and its conservation in AIDs (Gajula et al., 2014; King et al., 2015).

In the previous chapter, we also showed that Gm-AID^{H136} could cause protrusion of Gm-AID^{Y127} into the catalytic pocket, leading to its closure. Interestingly, our computational modeling revealed that Gm-AID^{H136E} could prevent the Y127 protrusion into the active site (refer to chapter two; section 2.4.5). Previously, our lab and others have suggested a significant role of Hs-AID^{Y114} in shaping the catalytic pocket and defining the substrate specificity of AID (Gajula et al., 2014; King et al., 2015). More recently, the crystal structure of human AID revealed that Y114 (equivalent to Gm-AID^{Y127}) interacts with the O5′of dC and is involved in holding dC in the catalytic pocket (Qiao et al., 2017). Therefore, any amino acid change in this position may significantly hamper AID activity, consistent with its conservation amongst AID homologs. Close to Gm-AID^{H136}, another amino acid position also showed a noteworthy change of charge. Position 137 in Gm-AID is occupied by the non-polar amino acid of valine (Gm-AID^{V137}) while the corresponding position in Dr-AID and Ip-AID is occupied by the positively-charged amino acid of arginine (Dr-AID^{R136} and Ip-AID^{R135}).

Taken together, the structural prediction and ssDNA docking analysis suggested that the lack of a critical positive residue on ℓ2 and the substitution of Hs-AID^{E122} with Gm-AID^{H136} have created conditions where α4 involvement in substrate binding has increased 3- to 4-fold in Gm-AID. To test this hypothesis, we generated several Gm-AID mutants. We observed that all the mutants, except for Gm-AID$^{H136E\text{-}V137R}$, showed a low to moderate increase in catalytic activity (*i.e.,* K_{cat}/K_m) with Gm-AID$^{N29R\text{-}H136E\text{-}V137R}$ exhibiting the highest improvement in catalytic activity (*i.e.*, ~10-fold; Table 3-6). Therefore, we concluded these residues were partially responsible for the lethargic activity of Gm-AID.

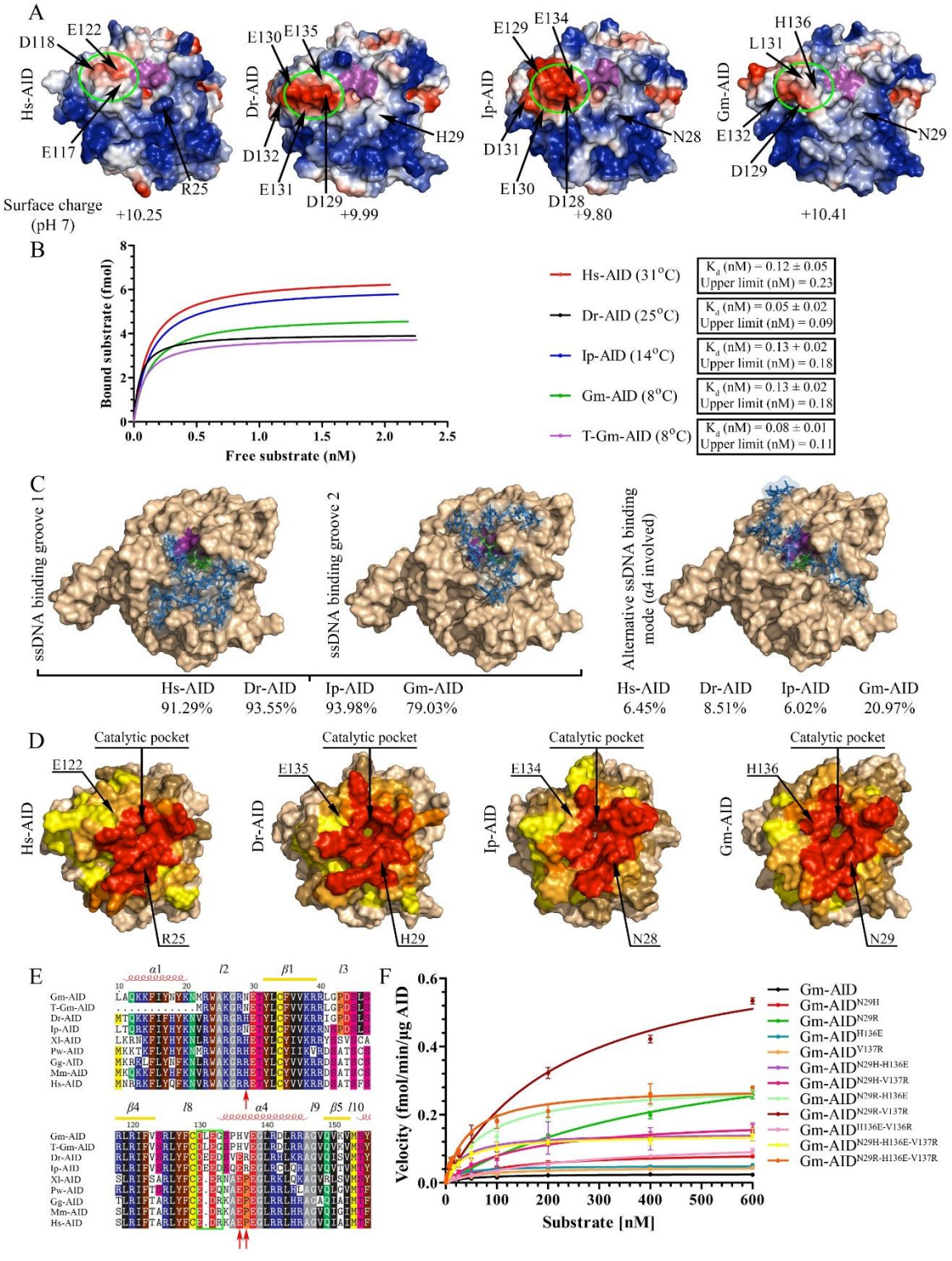


Figure 3-8: Basis of Atlantic cod AID lethargy. A) Predicted surface topology of Gm-AID was compared to that of other AID homologs. Positive, neutral, and negative residues are colored blue, white, and red,

respectively. The putative catalytic pocket is colored in purple. Surface charge (at pH 7.00) is shown below each model. The end of ℓ8 and the beginning of α4 which are different in Gm-AID compared to other AID homologs are labelled with a green circle and residue names. B) EMSA was conducted to compare global ssDNA binding affinity of AID homologs. Purified AIDs were incubated with a 0.025 to 2.5 nM range of [substrate] for 1 h. Results were plotted as fmol bound substrate against nM free substrate. For each of the AID homolog, 2 to 3 protein preparations were tested in duplicate. Estimated K_d and upper limits show no significant difference amongst AID homologs. C) Docking of ssDNA on the surface of the Gm-AID model revealed the presence of the two main ssDNA binding groove 1 and 2 previously identified in Hs-AID, as well as alternative ssDNA binding mode which involve the α4 region. The contribution of different binding modes is shown for each of the AID homolog. D) Interactions between AID residues and ssDNA are shown as heatmaps. Amino acid residues interacting with substrate in 50-100%, 30-50%, 15-30%, 5-15%, 0-5%, and 0% of docking events are shown in red, dark orange, light orange, yellow, sand and wheat colors, respectively. Shown with arrows are two potential amino acids that contribute to increasing the involvement of Gm-AID α4 and their counterparts in other AID homologs. E) Partial alignment of the AID homologs surrounding Gm-AID^{N29}, Gm-AID^{H136}, and Gm-AID^{V137} residues. These residues were later altered to their Hs-AID or Dr-AID counterparts. Green box shows the end of ℓ8 which is different in bony fish AIDs compared to other AID homologs. F) The catalytic rate of Gm-AID mutants was compared to that of wildtype Gm-AID through Michaelis-Menten kinetics. Data is represented as mean ± SEM (n=4).

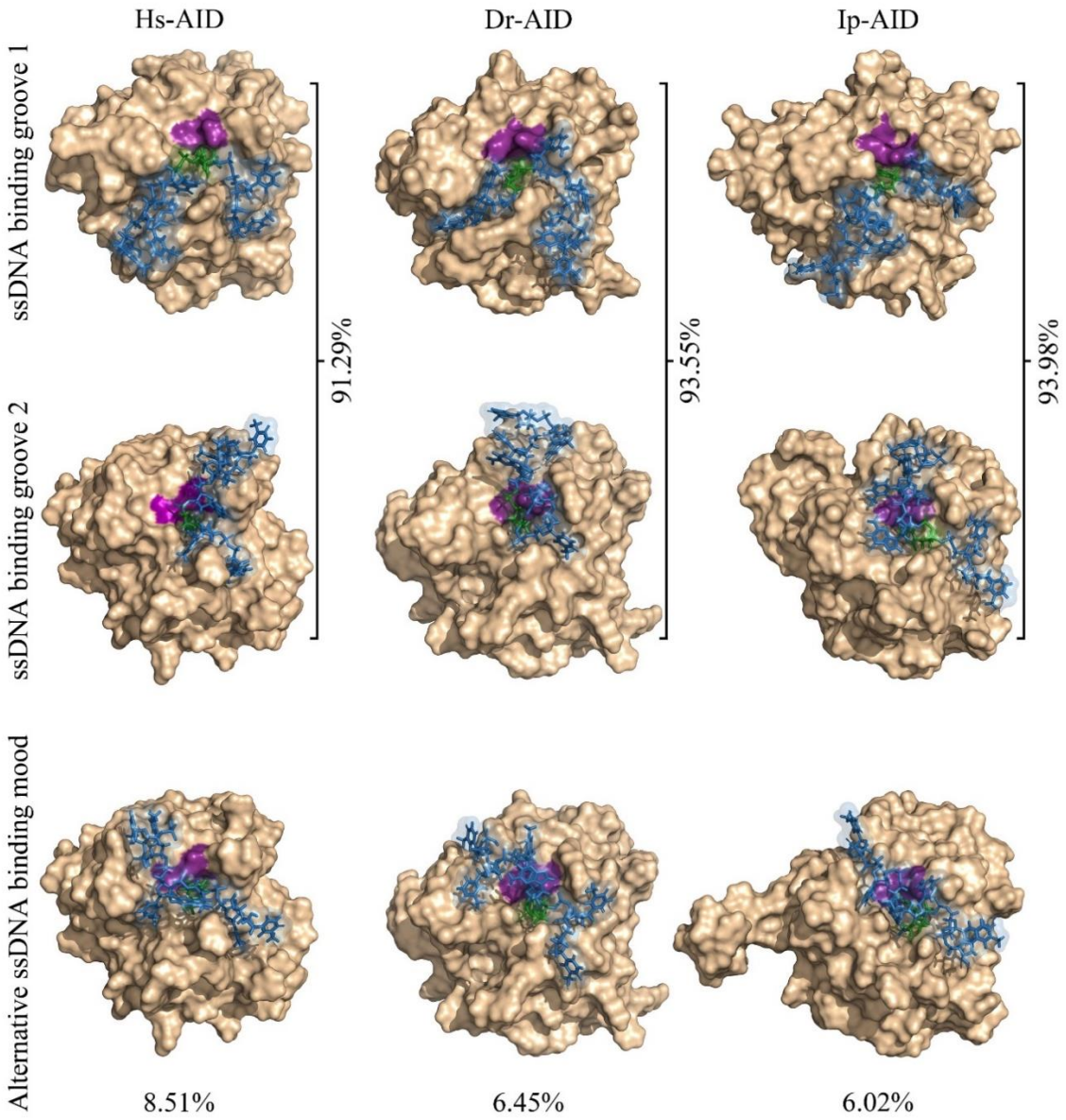


Figure 3-9: AID ssDNA binding modes. Docking of ssDNA on the surface of the Hs-AID, Dr-AID, and Ip-AID revealed the presence of the main ssDNA binding groove 1 and 2, as well as alternative binding mode which involves AID α4 region. The contribution of different binding modes is shown for each AID homolog. Abbreviations: Dr-AID: zebrafish AID; Ip-AID: channel catfish AID; Hs-AID: human AID.

Table 3-4: Comparison of DNA interaction with substrate binding grooves on the surface of AID homologs

	Hs-AID (%)	Dr-AID (%)	Ip-AID (%)	Gm-AID (%)
ssDNA binding groove 1	75.53	38.71	54.22	48.39
ssDNA binding groove 2	7.45	29.03	30.12	16.13
ssDNA binding groove 1 and 2	7.45	19.35	8.43	14.52
Direct involvement of α4	8.51	6.45	6.02	20.97

Abbreviations: Gm-AID: Atlantic cod AID; Dr-AID: zebrafish AID; Ip-AID: channel catfish AID; Hs-AID: human AID.

Table 3-5: Comparison of Gm-AID^{H136} residue in interaction with -1 position nucleotide upstream of the target dC and total interactions with substrate to its equivalent residue in other AID homologs

	Interaction with G in TGC motif (%)	Total interactions with substrate (%)
Hs-AID^{E122}	5.319%	18.685%
Dr-AID^{E135}	3.226%	9.677%
Ip-AID^{E134}	2.410%	28.916%
Gm-AID^{H136}	16.129%	53.226%

Abbreviations: Gm-AID: Atlantic cod AID; Dr-AID: zebrafish AID; Ip-AID: channel catfish AID; Hs-AID: human AID.

Table 3-6: Michaelis-Menten kinetics parameters measured for Gm-AID mutants

	Temp (°C)	pH	K_{cat} (min^{-1})	K_m (nM)	V_{max} (fmol/min/µg)	Std. Error		R^2	K_{cat}/Km ($min^{-1}nM^{-1}$)	Activity ratio
						K_{cat} (min^{-1})	K_m (nM)			
Gm-AID	8	8.08	1.36E-06	44.05	0.02585	3.05E-08	3.421	0.9735	3.09E-08	1.00
Gm-AID^{N29H}	8	8.08	4.87E-06	119.8	0.09242	1.47E-07	10.31	0.9883	4.07E-08	1.32
Gm-AID^{N29R}	8	7.89	2.44E-05	491.7	0.4619	1.23E-06	44.7	0.9944	4.95E-08	1.60
Gm-AID^{H136E}	8	7.89	2.64E-06	26.13	0.05013	8.80E-08	3.206	0.9654	1.01E-07	3.27
Gm-AID^{V137R}	8	7.99	2.43E-06	59.46	0.04616	8.58E-08	6.942	0.973	4.09E-08	1.32
Gm-AID$^{N29H-H136E}$	8	7.99	7.68E-06	32.75	0.1458	4.02E-07	6.161	0.9241	2.35E-07	7.59
Gm-AID$^{N29H-V137R}$	8	7.89	1.01E-05	143.7	0.1916	4.09E-07	15.7	0.9828	7.04E-08	2.28
Gm-AID$^{N29R-H136E}$	8	7.77	1.54E-05	80.84	0.2921	6.06E-07	9.94	0.9718	1.9E-07	6.16
Gm-AID$^{N29R-V137R}$	8	7.56	3.78E-05	241	0.7154	1.83E-06	26.76	0.9856	1.57E-07	5.07
Gm-AID$^{H136E-V137R}$	8	7.89	6.73E-06	254.9	0.1277	4.56E-07	38.89	0.9733	2.64E-08	0.85
Gm-AID$^{N29H-H136E-V137R}$	8	7.89	7.42E-06	36.38	0.1407	2.28E-07	3.979	0.9741	2.04E-07	6.60
Gm-AID$^{N29R-H136E-V137R}$	8	7.77	1.48E-05	44.23	0.2812	4.19E-07	4.334	0.9789	3.36E-07	10.85

Abbreviation: Gm-AID: Atlantic cod AID.

3.4.4 Potentially different substrate binding strategy in bony fish AIDs

We noticed two important structural differences between bony fish and tetrapod AIDs studied in this chapter. First, the end of ℓ8 contains four negatively charged amino acids in Dr-AID^{129DEED132} and Ip-AID^{128DEED131} (bony fish), creating a highly negative region close to α4. This region in Xl-AID^{120EERN123} (*Xenopus laevis*, the South African clawed toad, amphibian), Pw-AID^{117EEQN120} (*Pleurodeles waltl*, Iberian ribbed newt, reptile), Gg-AID^{117EDRK120} (*Gallus gallus domesticus*, chicken, bird), Mm-AID^{117EDRK120} (*Mus musculus*, mouse, rodent), and Hs-AID^{117EDRK120} (*Homo sapiens*, human, primate) contains only two negatively charged amino acids. Additionally, except for Pw-AID, other tetrapod AIDs studied in this report contain a positively charged amino acid (*i.e.*, arginine) in this region (Figure 3-8 A [green circle] and E [green box]). Secondly, the amino acid position 25 in Hs-AID (R25) and its corresponding position in other tetrapod AIDs contains a positively charged amino acid (*i.e.*, Xl-AID^{H27}, Pw-AID^{H25}, Gg-AID^{R25}, and Mm-AID^{H25}).

The crystal structure of APOBEC3A (A3A; AID's close relative) has shown that the same position in A3A (*i.e.*, H29) hydrogen bonds with the phosphate backbone of ssDNA substrate and stabilizes the substrate binding by contributing to the formation of the U-shaped DNA conformation that fits into the DNA binding groove (Shi et al., 2017). This amino acid is located at the mouth of the catalytic pocket in the AID/APOBEC enzymes and seems to act as an anchor for the substrate (Harjes et al., 2013; King & Larijani, 2017; King et al., 2015; Pham et al., 2013; Shi et al., 2017). Interestingly, since the protonation state of histidine, and therefore the number of hydrogen bonds, varies at

different pH points, the acidic pH preference of the A3A^{H29} and A3G^{H216} was attributed to this residue (Harjes et al., 2013; Pham et al., 2013; Shi et al., 2017). The same position in Hs-AID is occupied with an arginine (*i.e.*, R25) and we have previously shown that this position is a part of secondary catalytic residues that stabilize the dC in the catalytic pocket (King & Larijani, 2017; King et al., 2015). We also previously showed that arginine to alanine mutation in Hs-AID at this position reduced the conformations with dC docked in catalytic pocket by 40% compared with the wild type Hs-AID (King & Larijani, 2017).

Based on the above-mentioned difference, we propose that bony fish AIDs might have evolved to utilize a different strategy to direct ssDNA into the substrate binding groove compared to tetrapods. It is possible that in bony fish AIDs, the repellent forces originated from negatively charged region of ℓ8 is an important contributing factor to substrate binding, especially in the bony fish AIDs lacking the highly positive amino acid at the mouth of catalytic pocket (*e.g.*, Ip-AID^{N28} and Gm-AID^{N29}). While, in tetrapods, it seems that the arching of ssDNA around the positively charged amino acid at the mouth of the catalytic pocket (*i.e.,* Hs-AID^{R25} and its equivalent in other species) might be the main strategy to position dC in the AID's catalytic pocket (Harjes et al., 2013; King & Larijani, 2017; King et al., 2015; Shi et al., 2017).

We have compared three bony fish AIDs in this chapter. Dr-AID and Ip-AID both possess four negative residues in the end of their ℓ8 region (Dr-AID^{129DEED132}, Ip-AID^{128DEED131}) while Gm-AID contains only two negative residues in this region (Gm-AID^{130DLEG133}). However, Dr-AID is the only studied bony fish that contains a positively charged amino acid at the mouth of its catalytic pocket (Dr-AID^{H29}). Ip-AID and Gm-AID

both possess an N in this position (Ip-AID^{N28} and Gm-AID^{N29}). Based on our hypothesis presented here, the presence of the positively charged amino acid at the mouth of Dr-AID catalytic pocket (Dr-AID^{H29}) and the lack of two negatively charged amino acids at the end of Gm-AID ℓ8 could contribute to the higher and lower catalytic activity of Dr-AID and Gm-AID, respectively. To test the role of the positively-charged amino acid at the mouth of the catalytic pocket in AID's enzymatic activity, we generated Hs-AID^{R25H}, Hs-AID^{R25N}, Hs-AID^{R25A}, Hs-AID^{R25del}, Dr-AID^{H29N}, Ip-AID^{N28H}, and Gm-AID^{N29H} mutants and compared their catalytic activity with that of their corresponding wildtype AIDs. Since previous reports have suggested an involvement of this position in regulating the pH preference of APOBECs, we examined the activity of these mutants in different pH points (Harjes et al., 2013; Pham et al., 2013; Shi et al., 2017). We observed that the histidine/arginine to asparagine/alanine/deletion mutations drastically decrease the catalytic activity while asparagine to histidine/arginine mutations can significantly improve the catalytic activity of AIDs (Figure 3-10). These results demonstrate that the presence of a positively charged amino acid at the "mouth" of AID's catalytic pocket could enhance its catalytic activity, and that the lack of this residue in Gm-AID could be a contributor to its lower activity. We suggest that this phenomenon could be due to improvement positioning of dC in the catalytic pocket. Further AID:ssDNA docking simulations are required to confirm this hypothesis.

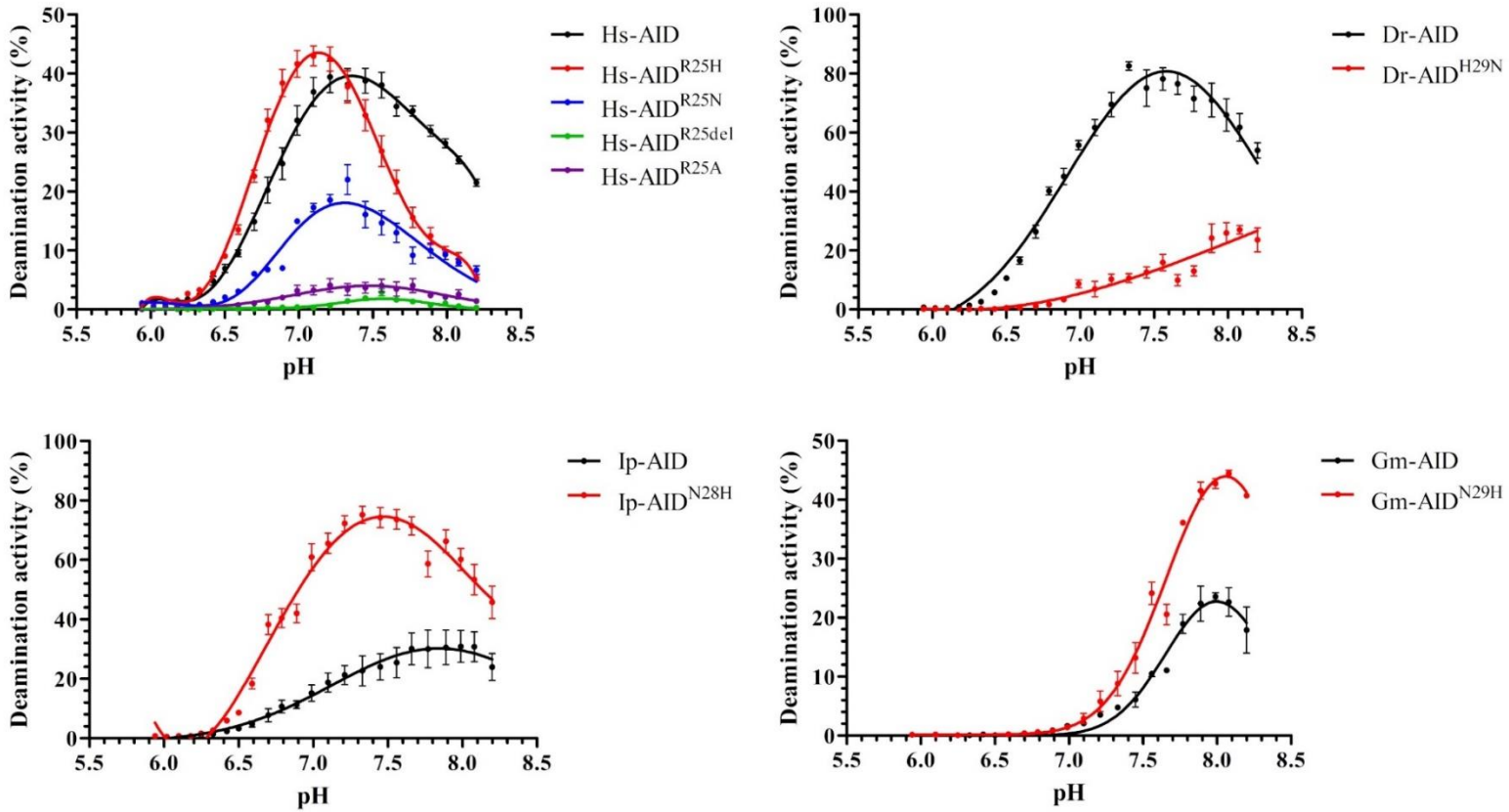


Figure 3-10:The role of positively-charged amino acid at the mouth of AID's catalytic pocket in its activity. Hs-AID^{R25} and its corresponding amino acid in other AID homologs was mutated to assess the impact of this positively-charged amino acid in AID activity. The activity of the purified wildtype and mutants was tested on TGCbub7 substrate using our standard alkaline cleavage assay. Dr-AID (n=4) and its mutant (n=4) were incubated at 25 °C for 20 min. Hs-AID (n=8) and its mutants (n=4) were incubated at 31°C for 3 h. Ip-AID (n=4) and its mutant (n=4) were incubated at 14 °C for 10 h. Gm-AID (n=2) and its mutant (n=2) were incubated at 8 °C for 96 h. Data is represented as mean ± SEM. Abbreviations: Gm-AID: Atlantic cod AID; Dr-AID: zebrafish AID; Ip-AID: channel catfish AID; Hs-AID: human AID.

3.4.5 Atlantic cod AID sequence specificity and co-evolution with *Ig* genes

Analysis of AID from different species have defined the WRC motif as the AID favored target motif (Dancyger et al., 2012; Gajula et al., 2014; Larijani, Frieder, Basit, et al., 2005; Larijani & Martin, 2007; Malecek et al., 2005; Marianes & Zimmerman, 2011; Yang et al., 2006). However, more-distant homologs such as cartilaginous fish and lamprey AID, exhibit divergent patterns of sequence specificity (Emma M. Quinlan, 2017). To examine substrate specificity of Gm-AID, we compared its deamination activity on WRC and non-WRC motifs (Figure 3-11 A). Given its extreme cold adaptation, we also examined the dependence of substrate specificity on incubation temperature. As expected for Hs-AID, we observed clear WRC specificity which was not dependent on incubation temperature (Figure 3-11 B). Dr-AID and Ip-AID exhibited WRC preference, as did Gm-AID, favoring the two WRC motifs, TGC and AGC. The statistical analyses revealed that WRC specificity was more strict in Hs-AID compared to fish AIDs (Figure 3-12 and Appendix 2 to Appendix 5), consistent with previous findings (Dancyger et al., 2012; Emma M. Quinlan, 2017). Specifically, the distribution of each substrate was compared to that of all WRC and all non-WRC motifs. The distribution of all WRC and all non-WRC motifs were significantly different for all AID homologs tested here, indicating the AID specificity for WRC motifs. In the case of Hs-AID, the distribution of each WRC or non-WRC motif was significantly different from average of all non-WRC or WRC motifs, respectively. However, the distribution of TAC *vs.* all non-WRC motifs and GAC *vs.* all WRC motifs were not statistically different for any of the bony fish AID homologs studied here. This suggests that the specificity of bony fish AIDs is slightly different from that of

Hs-AID where GAC seems to be a better substrate than TAC. These results are consistent with high but not absolute conservation in the substrate specificity loop (ℓ8) amongst the four studied homologs (Carpenter et al., 2010; Gajula et al., 2014; Wang et al., 2010).

Co-evolution of *Ig* variable (V) genes with AID WRC specificity has been observed in mammals, birds, amphibians, bony and cartilaginous fish (Conticello et al., 2005; Detanico et al., 2016; Golub & Charlemagne, 1998; Jolly et al., 1996; Oreste & Coscia, 2002; Wagner et al., 1995). In these studies, serine codons were divided into AGY and TCN and a clear preference for AGY (WRC) over TCN (non-WRC) was observed in *IgV* CDRs *vs.* framework regions (FRs). Also, the WGCW motifs, which contain AID hotspots on both strands, have been suggested as an evolutionary feature of human IgV_H genes (Tang et al., 2020) that attract AID to these loci (Hwang et al., 2017; Ohm-Laursen & Barington, 2007; Wei et al., 2015; Yeap et al., 2015). We reasoned that if lack of a robust humoral immune response in Atlantic cod is indeed related to a severely compromised AID enzyme, there ought to have been a lower degree of evolutionary pressure to maintain enrichment of WRC motifs in CDR regions of Atlantic cod *IgV* genes.

We annotated the *IgH* loci in the Atlantic cod genome and analyzed the pattern of WRC enrichment in its IgV_H region (Figure 3-13 A). To characterize *IgH* loci in the improved version of the Atlantic cod genome (gadMor2), previously published partial Atlantic cod *IgH* chain region (with GenBank identifier AJ871288.1) along with complete protein sequences for other bony fish IgM, IgD, and IgZ, extracted from GenBank, were searched against gadMor2 genome sequence. However, since this region was not fully assembled in gadMor2, we were not able to completely verify the J segment. Although two

potential regions for *IgH* loci were characterized, we found that *IgH* variable regions were more centralized in larger clusters in Linkage Group (LG) 02. Interestingly, our BLAST results revealed no evidence of *IgZ* heavy chain in garMor2. Also, a scaffold containing various *IgV* regions was identified (Figure 3-13 A).

The AID hotspot enrichment was calculated excluding CDR3 and FR4 since the VDJ recombination is responsible for forming CDR3 and we could not fully characterize the FR4 of all species (Table 3-7). Strikingly, amongst analyzed species, we found that Atlantic cod CDR 1 and 2 exhibited the lowest level of WRC/GYW and WGCW enrichment compared to FR 1, 2, and 3 (Figure 3-13 B and Table 3-7). To confirm the lack of enrichment is not due to a lower overall usage of WRC motifs in Atlantic cod IgV_H genes, the abundance of WRC in the entire IgV_H fragment was also compared (Table 3-8). Results showed that the abundance of WRC in the IgV_H region of all examined species is comparable despite the higher GC content of the Atlantic cod CDSs (Table 3-8). Thus, the Atlantic cod IgV_H CDR regions exhibit a specific lack of AID hotspot motif enrichment, in accordance with relieved evolutionary pressure from its near inactive AID enzyme.

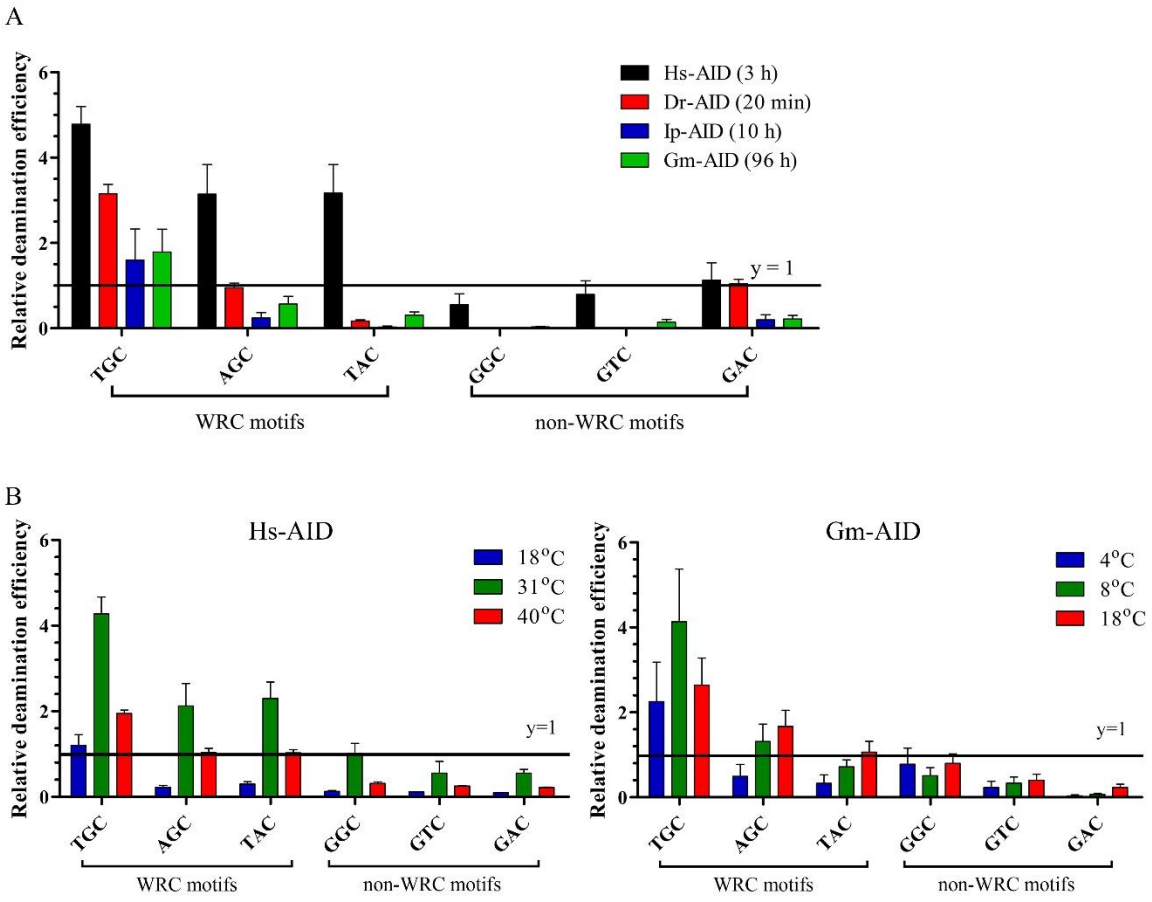


Figure 3-11: Atlantic cod AID sequence specificity. In these experiments, three independent protein preparations were tested for each AID homologs in duplicate (n=6). Incubation time was selected based on catalytic robustness of each AID homolog. All studied AIDs revealed preference for WRC motifs. Since the absolute activity level on each substrate varies amongst AID homologs, relative deamination efficiency was used to enable comparison between AID homologs. Relative deamination efficiency was calculated by dividing the activity on each substrate to that of the average activity for all 6 studied substrates. Data is represented as mean ±SEM. A) Gm-AID substrate preference was compared to that of other AID homologs. AID was incubated with various substrates containing WRC (TGC, AGC, and TAC) or non-WRC (GGC, GTC, and GAC) motifs at their corresponding optimal temperature. B) Given Gm-AID extreme cold adaptation, Gm-AID substrate specificity was studied in different incubation temperature (optimal, below, and above optimal; right panel). Hs-AID was used as control (left panel). Abbreviations: Gm-AID: Atlantic cod AID; Dr-AID: zebrafish AID; Ip-AID: channel catfish AID; Hs-AID: human AID.

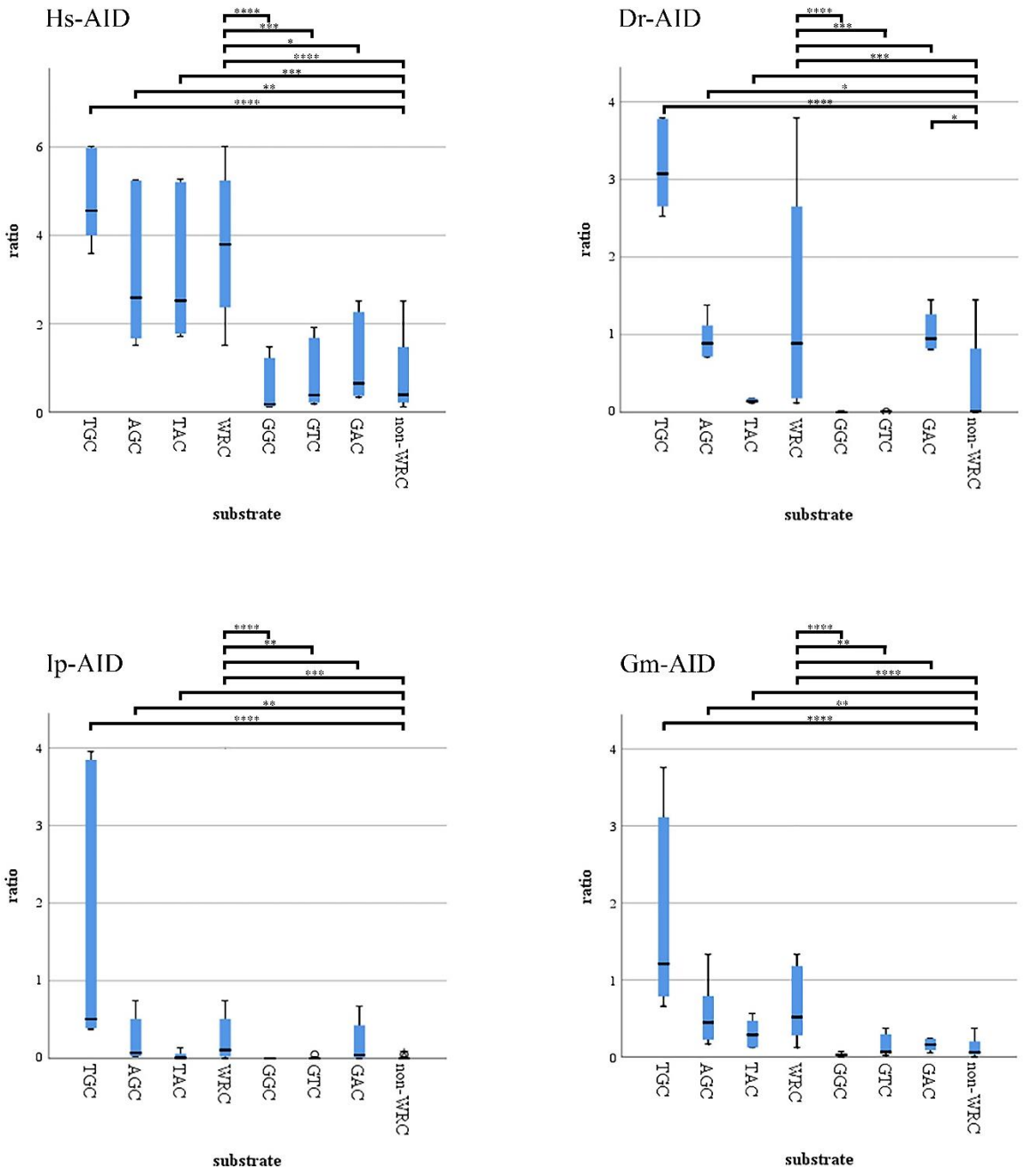


*Figure 3-12: The statistical analyses of the difference observed between substrate relative deamination efficiency of various AID homologs. The statistical difference between AID homologs were calculated using the independent samples Rruskal-Wallis test. The null hypothesis was considered as the distribution is the same between each pair of samples (n=6; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.005$; ****: $p < 0.001$). Abbreviations: Gm-AID: Atlantic cod AID; Dr-AID: zebrafish AID; Ip-AID: channel catfish AID; Hs-AID: human AID.*

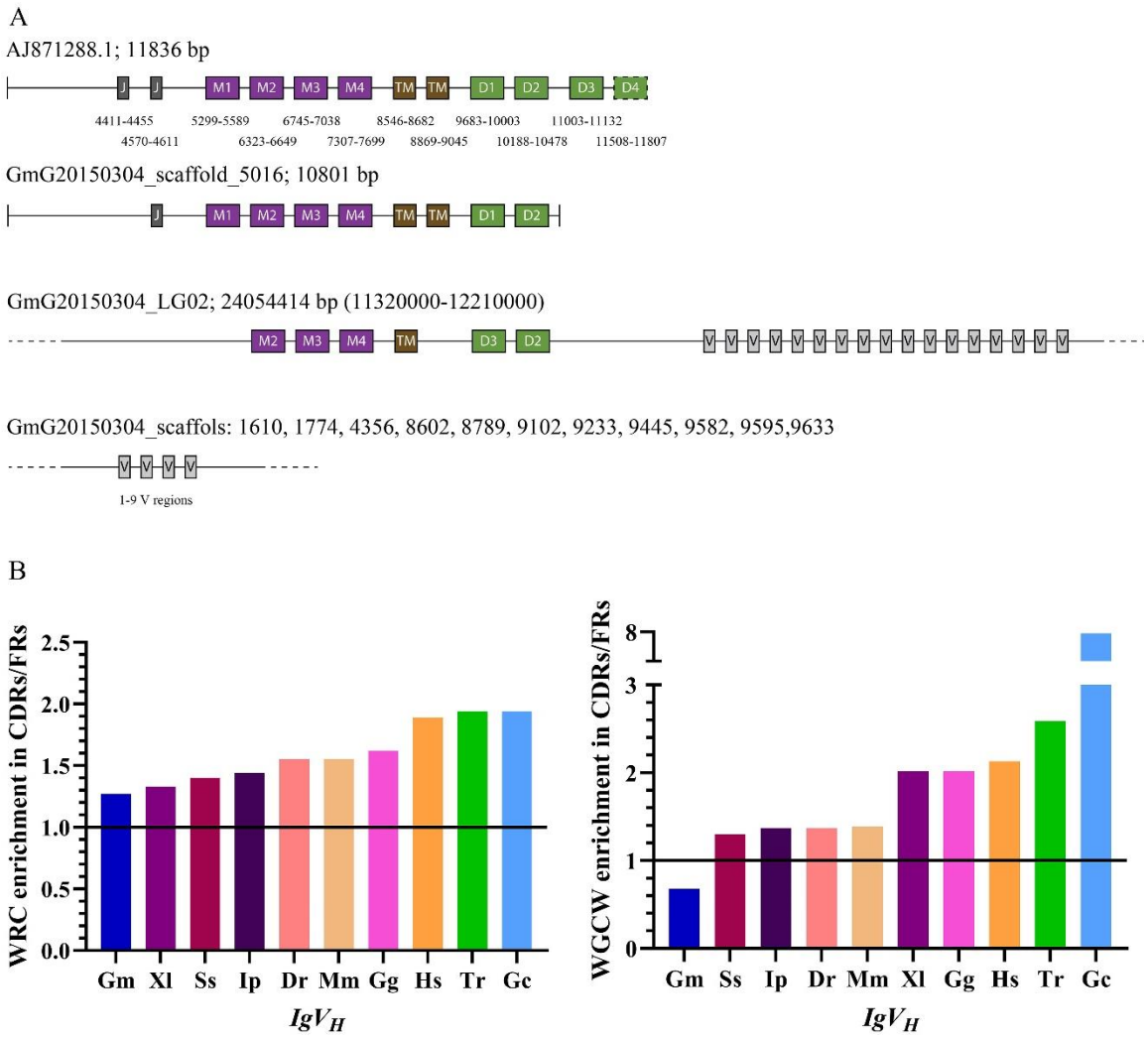


Figure 3-13: Co-evolution of Atlantic cod AID substrate specificity with Atlantic cod Ig genes. A) To characterize IgH loci in the improved version of the Atlantic cod genome (gadMor2), previously published partial Atlantic cod IgH chain region (with GenBank identifier AJ871288.1; top panel) along with complete protein sequences for other bony fish IgM, IgD, and IgZ, were searched against gadMor2 genome sequence using default parameters of BLAST+ program. Two potential regions (second and third panel) and a centralized larger cluster in Linkage Group (LG) 02 were characterized as IgH loci. Also, a scaffold containing various IgV regions was identified (fourth panel). B) Since AID's WRC specificity has been suggested as an evolutionary pressure in elevating AID hotspot motifs in vertebrates IgV genes, AID hotspot enrichment in CDRs vs. FRs was studied in IgV_H genes of Atlantic cod and several other vertebrate species. Abbreviations: Gm: Atlantic cod; Dr: zebrafish; Ss: Atlantic salmon; Ip: channel catfish; Tr: Japanese puffer fish; Gc: nurse shark; Xl: South African clawed toad; Gg: chicken; Mm: mouse; Hs: human.

Table 3-7: AID hotspot enrichment in IgV_H genes of various vertebrate species

	FR1			CDR1			FR2			CDR2			FR3					
	# AID hotspots	# nt. analyzed	index	# AID hotspots	# nt. analyzed	index	# AID hotspots	# nt. analyzed	index	# AID hotspots	# nt. analyzed	index	# AID hotspots	# nt. analyzed	index	Ave. FRs	Ave. CDRs	CDRs/FRs
Gm-IgV_H	790	7837	0.10	435	2373	0.18	363	5610	0.06	174	1989	0.09	1955	12765	0.15	0.11	0.14	1.27
Ip-IgV_H	652	7199	0.09	482	2709	0.18	662	5498	0.12	379	2360	0.16	1753	12381	0.14	0.12	0.17	1.44
Tr-IgV_H	309	3675	0.08	219	1245	0.18	183	2361	0.08	268	1215	0.22	803	5517	0.15	0.10	0.20	1.94
Dr-IgV_H	410	5234	0.08	307	1786	0.17	396	3774	0.10	220	1510	0.15	1127	9143	0.12	0.10	0.16	1.55
Ss-IgV_H	2509	28445	0.09	1571	9215	0.17	2201	19629	0.11	1196	8333	0.14	6042	44363	0.14	0.11	0.16	1.40
Gc-IgV_H	727	7407	0.10	578	3102	0.19	664	6579	0.10	569	3027	0.19	1284	14250	0.09	0.10	0.19	1.94
Xl-IgV_H	88	902	0.10	50	292	0.17	67	611	0.11	33	252	0.13	192	1449	0.13	0.11	0.15	1.33
Gg-IgV_H	1218	15455	0.08	995	5010	0.20	1391	10627	0.13	1011	5031	0.20	3903	24359	0.16	0.12	0.20	1.62
Mm-IgV_H	3112	20493	0.15	1054	4209	0.25	689	11341	0.06	1730	13394	0.13	3907	25318	0.15	0.12	0.19	1.55
Hs-IgV_H	3322	27855	0.12	1452	5900	0.25	932	15503	0.06	2424	19590	0.12	4328	38075	0.11	0.10	0.18	1.89

Abbreviations: Gm: Atlantic cod; Dr: zebrafish; Ss: Atlantic salmon; Ip: channel catfish; Tr: Japanese puffer fish; Gc: nurse shark; Xl: South African clawed toad; Gg: chicken; Mm: mouse; Hs: human.

Table 3-8: AID hotspot enrichment in the entire IgV_H genes and GC content of annotated complete protein coding genes (CDSs) of various vertebrate species

	IgV_H gene analysis				Genomic analysis	
	# AID hotspot	# nt. analyzed	AID hotspots/nt. analyzed	# transcripts	# CDSs	GC%
Gm-IgV_H	3717	30574	0.1216	112	44330	59.53
Ip-IgV_H	3928	30147	0.1303	109	47956	51.46
Tr-IgV_H	1782	14013	0.1272	49	46294	54.11
Dr-IgV_H	2460	21447	0.1147	76	57060	49.85
Ss-IgV_H	13519	109985	0.1229	405	97576	55.12
Gc-IgV_H	3822	34365	0.1112	129	1507	47.97
Xl-IgV_H	430	3506	0.1226	44	49356	45.62
Gg-IgV_H	8518	60482	0.1408	239	56680	50.23
Mm-IgV_H	10492	74755	0.1404	420	88579	51.96
Hs-IgV_H	12458	106923	0.1165	727	120426	51.02

Abbreviations: Gm: Atlantic cod; Dr: zebrafish; Ss: Atlantic salmon; Ip: channel catfish; Tr: Japanese puffer fish; Gc: nurse shark; Xl: South African clawed toad; Gg: chicken; Mm: mouse; Hs: human.

3.5 Discussion

Previous studies in various vertebrate species have pinpointed AID (encoded by *aicda* gene) as the enzyme responsible for introducing mutations in *Ig* genes, initiating the processes of antibody affinity maturation and class switch recombination (Bransteitter et al., 2003; Bransteitter et al., 2006; Frieder et al., 2006; Kolar et al., 2007; Larijani, Frieder, Basit, et al., 2005; Meffre et al., 2001; Muramatsu et al., 2000; Muramatsu et al., 1999; Muto et al., 2000; Nagaoka et al., 2002; Sernandez et al., 2008; Wang et al., 2009). Therefore, the emergence of ancestral AID at the base of vertebrate evolution has coincided with the presence of antibody maturation (Bromage et al., 2006; Cain et al., 2002; Dooley & Flajnik, 2005; Dooley et al., 2006; Hsu, 2016; Jenne et al., 2003; Kaattari et al., 2002; Malecek et al., 2005; Marianes & Zimmerman, 2011; Mehr et al., 2004; Wilson et al., 1992; Yang et al., 2006). However, unlike other studied vertebrates, functional analyses of the Atlantic cod humoral responses revealed the absence of antigen-specific affinity-matured antibodies in this species (Arnesen et al., 2002; Lund et al., 2008; Lund et al., 2006; Magnadottir et al., 2001; Schroder et al., 2009; Solem & Stenvik, 2006). In the previous chapter, we showed that two *aicda* transcripts are expressed mainly in Atlantic cod immune-related tissues. We also showed that Atlantic cod *aicda* transcript expression increases in response to immune stimulation. In this chapter, we sought to explore the functional properties of purified Atlantic cod AID proteins (Gm-AID). Remarkably, we found that the full-length Atlantic cod AID protein is a lethargic enzyme exhibiting the lowest optimal temperature reported for AIDs thus far. Intriguingly, the evolutionary ramification of this drastic decline in the activity of Gm-AID is evident in the sequence of

its *Ig* genes where we observed the lowest AID hotspot enrichment compared to other studied vertebrate species in this chapter.

The level of AID expression and activity are crucial determinants of SHM and CSR level (Sernandez et al., 2008; Takizawa et al., 2008; Wang et al., 2009). Previous studies unveiled significant variation in the biochemical properties of AID in bony and cartilaginous fish. Dr-AID is the most catalytically robust AID studied to date, possibly due to its requirement for epigenetic functions (Abdouni et al., 2013). However, other studied AID homologs, had catalytic rates lower or similar to that of Hs-AID, which only performs one deamination in several minutes (Abdouni et al., 2013; Emma M. Quinlan, 2017; Larijani et al., 2007). This rate is orders of magnitude lower than a typical enzyme (Larijani et al., 2007). Here, we showed that the catalytic rate of Gm-AID is orders of magnitude lower than that of other AID homologs. Even under its most optimal conditions (*i.e.*, optimal pH, temperature, and substrate), we observed approximately 350- and 3000-fold less activity for Gm-AID compared to Hs-AID and Dr-AID, respectively; thus, making it unlikely to play a functional role as a cytidine deaminase *in vivo.* Also, we showed that T-Gm-AID lacks cytidine deaminase activity in our experimental conditions. The N-terminal sequence, which is missing in T-Gm-AID, is involved in stabilization of the core of the enzyme, stabilization of the surface DNA binding residues, and contains potential DNA binding residues and NLS (Hu et al., 2013; King et al., 2015; Patenaude et al., 2009). Given the importance of the N-terminal amino acids, the lack of cytidine deaminase activity observed here is most likely a *bona fide* property of T-Gm-AID.

It was previously proposed that AID optimal temperature correlates with the ambient temperature of given species (Barreto et al., 2005; Conticello et al., 2005; Dancyger et al., 2012; Ichikawa et al., 2006; Wakae et al., 2006). Accordingly, we and others have previously shown that the activity of mammalian and bird AIDs at 37 °C is higher than amphibian and bony fish counterparts and bony fish AIDs are more active at lower temperatures (16 to 25 °C) (Barreto et al., 2005; Dancyger et al., 2012; Emma M. Quinlan, 2017; Ichikawa et al., 2006; Wakae et al., 2006). Here, we identified 4 to 8 °C as the optimal temperature of Gm-AID, which also makes it not only the most cold-adapted AID/APOBEC family member (Hori et al., 2012; Petersen & Steffensen, 2003), but the most cold-adapted vertebrate DNA/RNA-editing enzyme reported to date to the best of our knowledge. In the future, it will be interesting to investigate the structural basis of this cold adaptation.

Remarkably, we also noticed that bony fish AIDs might be employing a different strategy than tetrapod AIDs to position dC in the catalytic pocket. Here, we propose that the bony fish AIDs most likely utilize a repellent force originated from their highly negatively charged ℓ8 (*i.e.*, ending region close to α4) to repel the ssDNA into DNA binding groove. However, it seems that tetrapod AIDs are more dependent on the arching of ssDNA around the positively charged amino acid positioned at the mouth of the catalytic pocket (*i.e.*, Hs-AID^{R25} and its corresponding amino acid in other tetrapod AIDs in this chapter) to locate the substrate into the DNA binding groove. In the case of Dr-AID, both negatively charged ℓ8 and slightly positively charged amino acid at the mouth of catalytic pocket (Dr-AID^{H29}) could assist with the proper positioning of the ssDNA/dC into the

substrate binding groove/catalytic pocket. Additionally, the lack of two negative amino acids on the Gm-AID ℓ8 (i.e., Gm-AID^{130DLEG133}) compared to the other bony fish AIDs examined in this chapter, might be a contributing factor in its slow catalytic activity. However, more mutational analysis is required to confirm this hypothesis.

Our computational modeling pinpointed three amino acid positions that may have contributed to the lethargic catalytic activity of Gm-AID (N29, H136, and V137). To test this hypothesis, we created Gm-AID mutants with single, double, and triple mutations where these amino acids were changed into their counterpart in Hs-AID or Dr-AID. We found that although Gm-AID$^{N29R\text{-}H136E\text{-}V137R}$ showed a 10-fold increase in catalytic rate, none of the variants rescued the catalytic rate of Gm-AID to levels comparable to the other AID homologs. These results indicate that other global residue changes in Gm-AID are responsible for its diminished catalytic activity (Gajula et al., 2014). This is highly suggestive of the presence of yet-to-be-identified restrictive mutation(s) or a lack of permissive mutation(s) in Gm-AID. Restrictive mutations mask the effect of causative key mutation(s) and permissive mutations are pre-requisite for causative ones to be effective (R. Merkl & R. Sterner, 2016). Therefore, methods such as comparing extant sequences (horizontal approach) could fail in pinpointing causative mutations.

Sequencing of the Atlantic cod genome has revealed a unique loss of genes involved in adaptive immunity, including *mhc II*, invariant chain, and *cd4* genes (Buonocore & Gerdol, 2016; Malmstrom et al., 2016; Solbakken, Rise, et al., 2016; Solbakken, Torresen, et al., 2016; Solbakken et al., 2017; Torresen et al., 2017). It has been suggested that the loss of these genes is correlated with immigration of cod to higher

latitudes where the cost of keeping these genes might have caused their loss (Solbakken, 2016; Solbakken, Rise, et al., 2016; Solbakken et al., 2017). Somatic hypermutation followed by clonal selection of B cells improves the affinity of antibodies. However, the majority of the B cells undergoing somatic hypermutation would be eliminated in the following clonal selection process; making antibody affinity maturation an expensive process (Wiens et al., 2001; Wiens et al., 1998). Remarkably, this has coincided with the expansion of *mhc I* and *tlr* genes, suggesting a re-modeling of Atlantic cod immune system to rely more on innate and cell-mediated immunity (Parham, 2015, 2016; Solbakken, 2016; Star et al., 2011). In addition to targeting *Ig* loci, AID is a genome-wide mutator known as a leading source of tumor-initiating double-strand breaks (Burns et al., 2017; Choudhary et al., 2017; Lindley et al., 2016; Steele, 2016). In this light, our finding of a potential loss of function for Atlantic cod AID is consistent with its lack of reliance on antibody affinity maturation, since in the absence of a critical requirement for a genome-damaging agent like AID, the evolutionary pressure to retain such an agent is absent.

Genomic analyses of Gadiformes species have dated the loss of MHC II pathway to their common ancestor (Malmstrom et al., 2016). Given the importance of MHC II pathway in the B cell activation and subsequent AID expression, it is interesting to know whether the functional impairment of AID is a phenomenon limited to the Atlantic cod or common in the Gadiformes lineage. Methods which take into account the evolutionary trajectory of a protein family such as ancestral sequence reconstruction (ASR, a vertical approach) could identify the evolutionary timepoint of AID functional impairment (Harms & Thornton, 2010; R. Merkl & R. Sterner, 2016). ASR could also assist in finding the

definitive structural basis of Gm-AID's extremely low catalytic activity (Harms & Thornton, 2010; R. Merkl & R. Sterner, 2016). In the later endeavor, computational studies of Gm-AID structure will also prove useful. Hs-AID is notoriously difficult to purify due partially to its highly positive surface charge of +10.25 mediating rampant non-specific protein:protein/DNA/RNA interactions; thus, the only available Hs-AID crystal structures are of heavily mutated and/or truncated versions (Pham et al., 2016; Qiao et al., 2017). To this end, we embarked on an alternative combined computational-evolutionary-biochemical approach to gain insights into functional and native structure of Hs-AID (King & Larijani, 2017; King et al., 2015). Similar approaches may prove useful for Gm-AID since it has a surface charge of +10.41 that may also impede crystallography of the native protein.

In summary, here, we reported that Gm-AID is a lethargic deaminase adapted to cold temperatures. Since the gene synteny and transcript expression of Gm-AID seems to be conserved compared to other studied teleosts (refer to the previous chapter), we propose that the altered functionality of Atlantic cod AID is more likely a result of active selection aimed at some sort of end point, most likely inactivation in this case. It has been suggested that the chronological loss of the immune related genes in ancestor of Atlantic cod is correlated with immigration of the ancestral species to the higher latitude where the cost of keeping some immune genes might have caused the loss of them (Solbakken, Rise, et al., 2016; Solbakken, Torresen, et al., 2016; Solbakken et al., 2017). Remarkably, re-design of the Atlantic cod immune system has not significantly reduced its fitness in its natural habitat. Re-modeling of the Atlantic cod immune system might be an on-going process and

complete functional impairment of Gm-AID might be the next step. Nevertheless, the implication of this study in Atlantic cod vaccine strategies is evident where vaccines targeting cell-mediated immune response might be the more promising approaches in Atlantic cod aquaculture.

Chapter 4:

Evolutionary trajectory of activation induced cytidine deaminase (AID) within Gadiformes lineage

4.1 Abstract

Unlike other jawed vertebrates, the humoral immune response of Atlantic cod does not generate antigen-specific high affinity antibodies. Previous studies revealed that in jawed vertebrates, the enzyme activation-induced cytidine deaminase (AID encoded by *aicda* gene) is responsible for the production of high affinity antibodies by converting deoxycytidine (dC) into deoxyuracil at immunoglobulin (*Ig*) loci. In the previous chapters, we showed that although the *aicda* gene synteny was conserved in Atlantic cod, its purified AID enzyme lacks robust cytidine deaminase activity. Based on these observations, we concluded that the lack of high affinity antibody production in Atlantic cod is likely due to the functional impairment of its AID enzyme. In this chapter, we expanded our enzymatic investigations to 33 AID homologs from extant bony fish species and applied ancestral sequence reconstruction (ASR) to examine the evolution of AID in the phylogenetic branches leading to Atlantic cod (*i.e.,* within the Gadiformes order). We found that the catalytic efficiency of AID enzyme was 15-fold reduced in the ancestor of Gadiformes. Interestingly, we detected a more drastic decrease of 33-fold in the catalytic efficiency of Gadidae ancestor. We pinpointed five potential amino acid mutations involved in catalytic activity reduction of Gadidae ancestor. These observations suggest that the evolution of AID within the Gadiformes species is most likely directed towards its complete functional impairment. These findings are consistent with recent findings of drastic remodeling of other humoral immune genes in the Gadiformes order. In addition, our comprehensive evolutionary comparative approach marks the first application of ancestral reconstruction and functional analyses to an enzyme involved in immunity and cancer.

4.2 Introduction

In vertebrates, B cell activation leads to the expression of activation-induced cytidine deaminase (AID, encoded by the *aicda* gene), which initiates the secondary antibody diversification process (Maul & Gearhart, 2010; Owen, 2019). Introduction of AID-mediated C to U mutations at the *Ig* genes results in production of antibodies with a higher affinity for cognate antigen (Maul & Gearhart, 2010; Owen, 2019). This process is known as antibody affinity maturation (AM) (Maul & Gearhart, 2010; Owen, 2019). Given the crucial role of AID in initiating AM, the rise of AID genes at the beginning of jawed vertebrate radiation was considered as the emergence of AM (Betz et al., 1993; Bromage et al., 2006; Cain et al., 2002; Diaz et al., 1999; Dooley & Flajnik, 2005; Dooley et al., 2006; Hsu, 2016; Jenne et al., 2003; Kaattari et al., 2002; Lee et al., 2002; Malecek et al., 2005; Marianes & Zimmerman, 2011; Mehr et al., 2004; Wilson et al., 1992; Yang et al., 2006). However, functional analyses of the Atlantic cod (*Gadus morhua*) have proved the absence of AM in this species. Specifically, high levels of low affinity IgM and lack of robust antigen-specific antibody responses upon immunization were observed in this species (Arnesen et al., 2002; Magnadottir et al., 1999; Magnadottir et al., 2001; Solem & Stenvik, 2006).

In the previous chapters, we sought to investigate the genetics, synteny, and enzymatic activity of AID in the Atlantic cod to uncover the molecular bases behind the lack of AM in this species. We found that although the gene synteny and transcript expression of *aicda* in Atlantic cod is mainly conserved compared to other teleost species, Atlantic cod AID (Gm-AID) enzyme is a very lethargic cytidine deaminase, exhibiting

350- to 3000-fold less activity than human and zebrafish AIDs, respectively. Therefore, we concluded that functional impairment of Gm-AID would contribute to the lack of affinity-matured antibodies in this species.

The teleost lineage of the ray-finned fishes (class Actinopterygii) is the largest and most diverse group of vertebrates (Ron Fricke; Sallan, 2014; Solbakken et al., 2017). Successful radiation and speciation within the Teleostei infraclass have been attributed to the adaptability of their immune system in response to major environmental changes (Malmstrom et al., 2016; Solbakken et al., 2017). The Atlantic cod is a member of the taxonomic order Gadiformes within Teleostei infraclass (Solbakken et al., 2017). It has been shown that immune gene losses and expansions in the Gadiformes overlap with major paleoclimatic and geological events (Solbakken, Rise, et al., 2016; Solbakken et al., 2017). Specifically, within Gadiformes order, the loss of key genes involved in adaptive humoral immunity (*i.e., mhc II*, *cd4*, and invariant chain [also known as *cd74*]), and expansion of genes involved in innate immunity (*i.e.*, *tlrs*), and cellular immunity (*i.e., mhc I*) indicate the probability of alternative immune strategies. Given the importance of AID in humoral immunity (Sernandez et al., 2008; Takizawa et al., 2008; Wang et al., 2009), our findings on the extremely lethargic activity of Atlantic cod AID, and the cod-like remodeling of immune genes in other Gadiformes, we sought to extend our studies to other Gadiformes species. We asked whether the functional impairment of AID is a phenomenon unique to Atlantic cod, or a wider trend within the Gadiformes group. In addition to extant species, we wished to examine the ancestral AIDs within and leading up to Gadiformes to decipher

the evolutionary points at which AID activity may have been shaped to its present extremely lethargic state in Atlantic cod.

Ancestral sequence reconstruction (ASR) is a tool to infer the sequence of ancestral proteins based on the contemporary gene sequences (Harms & Thornton, 2010; R. Merkl & R. Sterner, 2016; Yang, 2006). By studying the predicted ancestral proteins' biochemical and structural properties, significant novel insights have been gained regarding past environmental conditions (Akanuma, 2017), protein structure and functional evolution (Babkova et al., 2020; Holinski et al., 2017; Qiu et al., 2019; Wheeler et al., 2016; Yang et al., 2020), and the evolutionary history of a protein family (Gumulya & Gillam, 2017; Harms & Thornton, 2010; Laursen et al., 2020). Notable proteins to which this approach has been fruitfully applied include thioredoxins (Ingles-Prieto et al., 2013), 3-isopropylmalate dehydrogenase (Furukawa et al., 2020; Groussin et al., 2015), haloalkane dehalogenases (Babkova et al., 2020), laccases (Gomez-Fernandez et al., 2020), postsynaptic density-95/Discs large/Zonula occludens (PDZ) 3 domain of Discs large (Laursen et al., 2020), cysteine-rich interactor of PDZ3 (Laursen et al., 2020), ribonuclease H (Lim et al., 2018), coagulation factor VIII (Zakas et al., 2017), short wavelength-sensitive type 1 UV pigment (Shi & Yokoyama, 2003), Pax proteins (Sun et al., 2002), elongation factors of the Tu family (Gaucher et al., 2003), steroid receptors (Thornton, 2001; Thornton et al., 2003), and rhodopsins (Chang, 2003; Chang et al., 2002). The power of ASR and the noticeable increase in ancestral gene prediction has inspired the establishment of a database called Revenant (https://revenant.inf.pucp.edu.pe/) (Carletti et al., 2020). The Revenant database contains a hand-curated collection of ancestral genes

annotated with methodological, structural, and biochemical information (Carletti et al., 2020).

In this chapter, we applied the ASR methodology to gain inside into the timepoint when AID became nearly inactivated in the evolutionary branches leading to the Atlantic cod. Here, we report the presence of an unexpected functional plasticity within bony fish AIDs. Our results showed that the functional impairment of Atlantic cod AID, examined in the previous chapter, was not an exception compared to its closely related species. We identified catalytically inactivated AID homologs from two other Gadiform species. We also showed that during the evolution of Gadiformes lineage, two separate events resulted in the cold adaptation and catalytic impairment of ancestral AID. Based on our ASR results, we concluded that the aforementioned evolutionary events have occurred in the common ancestor of Gadiformes and Gadidae species, respectively. This is the first report characterizing completely/near-completely functionally inactivated AID enzymes within vertebrate class. Since AID deficiency causes immunodeficiency in humans and in mouse models, these findings could change our perspective regarding the vertebrates' immune system evolution.

4.3 Methods

4.3.1 Ancestral sequence reconstruction (ASR)

ASR methodology is comprised of five steps:

4.3.1.1 Selecting extant species

To infer the ancestral protein sequences of AID within and outside of the Gadiformes lineage, the homologous *aicda* sequences were retrieved from 66 teleost genomes sequenced previously (Malmstrom et al., 2016). The Atlantic cod *aicda* gene locus (Ensemble gene identifier: ENSGMOG00000004114) was BLAST aligned against the assembled and raw genomic data of each species (the European Nucleotide Archive (ENA) accession number: PRJEB12469 and the Dryad repository: doi:10.5061/dryad.326r8) using default parameters of blastn task in BLAST+ program (Malmstrom et al., 2016). The genomic region was then retrieved as the *aicda* locus. The *aicda* mRNA transcript was then predicted using the AUGUSTUS server (http://bioinf.uni-greifswald.de/augustus/submission.php) (Stanke et al., 2006). The initiation codon, coding sequence (CDS), and the stop codon for identified *aicda* transcripts were confirmed using the ATGpr website (https://atgpr.dbcls.jp/) (Nishikawa et al., 2000). In total, the AID gene sequence from 73 species (74 gene sequences) was used to perform ASR analyses. The basic information and the AID nucleotide sequence of bony fish species selected for ASR analyses can be found in Appendix 6 and Appendix 7, respectively.

4.3.1.2 Creating a multiple sequence alignment

In this thesis, the amino acid multiple alignments were built based on our predicted structure of Gm-AID (Appendix 8) using the PROMALS3D web interface

(http://prodata.swmed.edu/promals3d/promals3d.php) (Pei et al., 2008). The generated amino acid alignment was then used to guide the nucleotide sequences alignment using the TranslatorX server (http://translatorx.co.uk) (Abascal et al., 2010). Since the accuracy of the multiple sequence alignment (MSA) impacts the ASR results (Vialle et al., 2018), the final nucleotide and amino acid alignments were manually inspected to assure the quality of the alignment.

4.3.1.3 Computing a phylogenetic tree

Another important factor contributing to the accuracy of ASR results is the topology of the phylogenetic tree (Groussin et al., 2015; R. Merkl & R. Sterner, 2016; Vialle et al., 2018). We used RAxML package version 8.2.9 to build the gene tree (Malmstrom et al., 2016). Appendix 9 contains all the scripts used for RAxML analyses. First the best substitution model was selected. The GTRCAT substitution model (*i.e.,* the General Time Reversible model with the CAT model of rate heterogeneity) gave the highest ML in the model test runs. Then, the initial rearrangement settings (*i.e.*, -*i*) and the number of categories (*i.e.*, -*c*) were calculated. The best ML tree and bootstrap values were estimated using -*i 10* and -*c 55*. However, our constructed gene tree did not fully agree with the previously published species tree. It was also shown that combining the information on the species phylogeny with the gene phylogenetic tree can improve the ASR results by predicting a more biochemically realistic and kinetically stable ancestral protein (Groussin et al., 2015). In this regard, we decided to use the previously estimated species tree for our dataset as the start tree in ASR calculations (Malmstrom et al., 2016). As an outgroup, *Lampetra tridentata CDA1* cytidine deaminase gene was used. In RAxML

analyses, ASR was performed based on both the AID's gene tree (constructed in this thesis), and the species tree previously published (Malmstrom et al., 2016).

4.3.1.4 Reconstructing ancestral sequences

We applied three approaches to predict the ancestral states (Appendix 9, Appendix 10, and Appendix 11). First, we used RAxML, which is based on the protein alignment and takes advantage of the ML algorithm (Stamatakis, 2014). Second, we used the ProtASR package to infer the ancestral sequences based on protein structure and ML algorithms (Arenas & Bastolla, 2019; Arenas et al., 2017). Finally, we used MrBayes software to predict ancestral states based on the protein alignment and Bayesian statistics (Altekar et al., 2004; Ayres et al., 2012; Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003; Ronquist et al., 2012).

For RAxML package, ancestral sequences were predicted using the GTRCAT substitution model (refer to section 4.3.1.3), *-i 10*, *-c 55*, and the best ML tree obtained in this thesis or the species tree previously published (Malmstrom et al., 2016).

In ASR analyses using MrBayes version 3.2.7, we used the GTR model with Gamma distribution of rate variation. Additionally, the 1st, 2nd, and 3rd nucleotide positions of a codon were unlinked. Each run was continued until the standard deviation of split frequencies of 0.01 or less was achieved, and the potential scale reduction factor (PSRF) for all parameters was reasonably close to 1.0. The previously published species tree (Malmstrom et al., 2016) was used as the start tree for the MyBayes analyses. Proper tree topology constraints were defined to infer the ancestral sequence of the desired node. For

each ancestral node, analyses were run four independent times, summed up, and reported as the results.

In ASR analyses using ProtASR versions 2.0 and 2.2 (Arenas & Bastolla, 2019; Arenas et al., 2017), we used our computationally predicted Gm-AID 3D structure (Appendix 8) and the previously published species tree. Since the length of the alignment was different from the length of the PDB file, we used version B of ProtASR. Unlike other ASR frameworks, ProtASR implements a structurally constrained substitution model of evolution called "Mean-field" (Arenas & Bastolla, 2019; Arenas et al., 2017).

The results of RAxML, ProtASR, and MrBayes were compared. The consensus ancestral sequences were obtained with more weight on the MrBayes results since the previous studies concluded that Bayesian inference with rate variation model might outperform other methods (Appendix 12) (Joy et al., 2016; Randall et al., 2016). For any position with ambiguity above 0.2, any predicted amino acid(s) with probability higher than 0.2 was also considered. For these positions, mutant versions of the predicted AID were made.

4.3.2 AID expression and purification

The abbreviations used for extant and ancestral AIDs are described in Table 4-1. Extant and ancestral AID homologs were expressed in the same pGEX5.3-based GST-fusion bacterial expression system and purified as described before in section 3.3.1 (Abdouni et al., 2013; Dancyger et al., 2012; Emma M. Quinlan, 2017; Larijani et al., 2007). Briefly, the CDS of each extant and ancestral AID homolog was synthesized (Integrated DNA Technologies [IDT], Inc., USA) or built using site-directed mutagenesis.

The ORFs were then inserted into pGEX-5x-3 (GE Healthcare, Waukesha, WI, USA) vector using EcoRI-HF® and NotI-HF® enzymes (New England BioLabs). *E. coli* Bl21(DE3) cells were used as the host cells to express GST-AID proteins (Abdouni et al., 2013; Dancyger et al., 2012; Emma M. Quinlan, 2017; Larijani et al., 2007). As the expression-inducing agent, Isopropyl β-d-1-thiogalactopyranoside (IPTG, 1 mM) was added to the bacterial culture containing a GST-AID expression vector followed by 16 h incubation at 16 °C. The GST-AID protein was purified from the lysed bacterial culture using Glutathione Sepharose high-performance beads (Amersham) and stored in 20 mM Tris-Cl pH 7.5, 100 mM NaCl, 1 mM dithiothreitol. The quality and quantity of the purified GST-AID preparations were measured using the coomassie staining protocol.

Table 4-1: Name and abbreviations of the extant AID homologs studied in this chapter.

Species		AID				
Scientific name	Common name					
Gadus morhua	Atlantic cod	Gm-AID	Gadinae	Gadidae (Gd-ANC)	Gadiformes (Gf-ANC)	(Zeiogadaria (Zg-ANC)
Boreogadus saida	Polar cod	Bs-AID				
Arctogadus glacialis	Arctic cod	Ag-AID				
Merlangius merlangus	Whiting	Mmerla-AID				
Melanogrammus aeglefinus	Haddock	Ma-AID				
Pollachius virens	Saithe	Pv-AID				
Gadiculus argenteus	Silvery pout	Ga-AID				
Trisopterus minutus	Poor cod	Tmi-AID				
Brosme brosme	Cusk	Bb-AID	Lotinae			
Molva molva	Ling	Mmol-AID				
Lota lota	Burbot	Llo-AID				
Phycis phycis	Forkbeard	Pp-AID	Phycinae			

Phycis blennoides	Greater forkbeard	Pb-AID	
Malacocephalus occidentalis	Western softhead grenadier	Mo-AID	Macrourinae
Macrourus berglax	Roughhead grenadier	Mb-AID	
Bathygadus melanobranchus	Vaillant's grenadier	Bm-AID	Bathygadinae
Laemonema laureysi	Guinean codling	Lla-AID	Moridae
Mora mora	Common mora	Mmor-AID	
Trachyrincus murrayi	Roughnose grenadier	Tmu-AID	Trachyrincinae
Trachyrincus scabrus	Roughsnout grenadier	Tsc-AID	
Muraenolepis marmoratus	Marbled moray cod	Mma-AID	Muraenolepididae
Melanonus zugmayeri	Arrowtail	Mz-AID	Melanonidae
Merluccius merluccius	European hake	Mmerlu-AID	Merlucciidae
Stylepnorus chordatus	Tube-eye	Sc-AID	Stylephoriformes
Cyttopsis roseus	Rosy dory	Cr-AID	Zeiformes
Zeus faber	John dory	Zf-AID	
Typhlichthys subterraneus	Southern cavefish	Tsu-AID	Percopsiformes

Gadidae sister group (Gds-ANC)

Percopsis transmontana	Sand roller	Pt-AID	
Polymixia japonica	Silver eye	Pj-AID	Polymixiiformes
Salmo salar paralog 1	Atlantic salmon	Ss-AID-1	Salmoniformes
Salmo salar paralog 2	Atlantic salmon	Ss-AID-2	
Danio rerio	Zebrafish (Zebra danio)	Dr-AID	Cypriniformes
Oryzias latipes	Medaka (Japanese rice fish)	Ol-AID	Beloniformes
Takifugu rubripes	Japanese pufferfish	Tr-AID	Tetraodontiformes
Ictalurus punctatus	Channel catfish	Ip-AID	Siluriformes
Homo sapiens	Human	Hs-AID	Hominidae

4.3.3 Substrate preparation

To assess the enzymatic properties of purified GST-AID proteins, a partially single-stranded bubble substrate containing a TGC (a WRC hotspot) motif (TGC strand) was synthesized by IDT. Previous studies have shown this substrate as the most favorite substrate for most AIDs studied thus far (Abdouni et al., 2013; Dancyger et al., 2012; King et al., 2015; Larijani & Martin, 2007; Larijani et al., 2007). As described in section 3.3.2, the TGC strand was 5'-radiolabeled with [γ-^{32}P] dATP and purified through mini-Quick spin DNA columns (Roche, Indianapolis, IN, USA). Using slow cooling (*i.e.*, 1 °C/min from 96 °C to 4 °C), the purified TGC strand was then annealed to three-fold excess of its partially complementary strand to generate partially single-stranded bubble substrate (TGCbub7).

4.3.4 pH buffer preparation

As described in the previous chapter (section 3.3.3), 100 mM Phosphate buffer with pH ranging from 5.8 to 8 were prepared in RNase/DNase free water (Gibco). The effective pH in the final reaction assay was measured by mixing phosphate buffer, TE buffer (used in substrate preparation), and AID storage buffer (used in GST-AID purification) to the ratio of 6:1:3.

4.3.5 Biochemical analysis of purified GST-AID

In this study, we explored the optimal temperature, pH, time course, substrate specificity, and enzyme kinetics of the purified GST-AID proteins (Abdouni et al., 2013; Dancyger et al., 2012; Larijani et al., 2007). For each experiment, at least two independent

protein preparation of GST-AID were tested in at least duplicate. All the experiments are described in more detail in chapter 3 in the Methods section.

All the enzymatic properties were examined using the previously published standard alkaline cleavage assay (Abdouni et al., 2013; Abdouni et al., 2018; Dancyger et al., 2012; Emma M. Quinlan, 2017; King et al., 2015; Larijani & Martin, 2007). In the standard assay, purified GST-AID protein and the radiolabeled TGCbub7 were incubated in phosphate buffer at the corresponding pH, temperature, and time length for each AID homologs. The reactions were then halted at 85 °C for 20 min. The enzyme Uracil-DNA glycosylase enzyme (UDG, NEB) was added to each reaction to remove the AID-mediated dU and create an abasic site, which was then alkaline cleaved at 96 °C. Using denaturing acrylamide gel electrophoresis, the cleaved TGC strands were separated, and the GST-AID activity was reported as the percentage of TGCbub7 which were cleaved.

The optimal temperature of purified GST-AID proteins was determined in phosphate buffer pH 7.3. In these experiments, 3 µl of AID protein preparation and 25 fmol of ^{32}P-labelled TGCbub7 substrate were incubated at various temperature points (4 °C to 50 °C). In the case of more cold-adapted GST-AIDs, a colder range of temperature was selected (*i.e.*, starting from -4 °C or -10 °C with 2 °C increments). To reach the colder temperatures than 0 °C, the reactions were incubated in different cooling baths containing a slush of aqueous NaCl solution. The freezing point depression formula (*i.e.*, Blagden's Law) was used to calculate the required NaCl amount to create cooling baths with desired melting temperature points (Table 4-2) (Averill, 2011):

$$\Delta T_f = K_f m i$$

where the K_f is the freezing point depression constant (*i.e.,* cryoscopic constant; $K_f water = 1.86$), m is the molal concentration of the solute, and i is the Van't Hoff factor:

$$i = \frac{moles\ of\ particles\ in\ the\ solution}{moles\ of\ formula\ units\ dissolved}$$

Table 4-2: Amount of NaCl added to 1 Kg of water to establish below 0 °C incubation temperatures

Freezing point (°C)	-2	-4	-6	-8	-10
NaCl (g)	31.42	62.84	94.26	125.68	157.1

The optimal pH of each GST-AID proteins was examined at their corresponding optimal temperature in a reaction containing 3 µl of GST-AID preparation, 25 fmol of ^{32}P-labelled TGCbub7, and 6 µl of phosphate buffer with effective pH ranging from 5.9 to 8.2 (8 pH points) in the final volume of 10 µl.

The maximum incubation time to retain the GST-AID activity within the initial velocity was estimated from a time-course experiment. In this experiment, 3 µl of purified GST-AID was incubated with 25 fmol of radiolabeled TGCbub7 substrate at its corresponding optimal temperature and pH for various incubation time points. These results were used to estimate the proper incubation time for the Michaelis-Menten kinetics assay, which is essential to be done within the initial velocity of the enzyme activity.

The catalytic properties (*i.e.,* K_{cat}, K_m and V_{max}) were calculated through Michaelis-Menten kinetics assay at the optimal temperature and pH and within the initial velocity of each GST-AID protein. Specifically, 3 µl of purified GST-AID were incubated with a 0.03125-600 fmol range (18 points) of ^{32}P-labelled TGCbub7 substrate. The results were

plotted as velocity (fmol of deaminated product/min of incubation/µg of AID) against substrate concentration (nM). The Michaelis-Menten parameters were estimated according to $Y = Et \times K_{cat} \times X/(K_m+X)$ equation where Y is the enzyme velocity, X is the substrate concentration, Et is the concentration of enzyme catalytic sites, K_{cat} is the number of times each enzyme site converts the substrate to product per unit time (*i.*e., the turnover number), and K_m (*i.e.*, the Michaelis-Menten constant) is the substrate concentration needed to achieve a half-maximum enzyme velocity (*i.e.*, V_{max}). Since AID has one catalytic pocket, the concentration of enzyme used in the experiment was used as an estimated Et. The molecular weight of the GST-AID proteins was calculated using Protein Molecular Weight web-based application (https://www.bioinformatics.org/sms/prot_mw.html).

4.3.6 Enzyme assay data collection and quantification

As mentioned in the methods section of chapter 3, the alkaline cleavage results were quantified by performing densitometry using Image Lab software (version 6.0.1 build 34, Standard Edition, Bio-Rad Laboratories, Inc.). Data were plotted using GraphPad Prism 5 software (version 5.00, GraphPad Software, Inc., USA), and error bars were set to represent standard error (SEM). In each experiment, two to three independent protein preparations of extant, ancestral, and mutated GST-AIDs were tested in duplicate or triplicate. Therefore, each point on an enzyme assay plot corresponds to the arithmetic mean of four to nine data points.

4.3.7 Correlation analyses of biochemical properties of extant AID homologs

Here we sought to explore the correlation relationship of optimal temperature and/or pH with K_{cat} or K_m. First, correlation coefficient was calculated in Microsoft Excel 365 using the following equation:

$$r_{xy} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

where the $\bar{x}$ and $\bar{y}$ are the average value of each parameter. The correlation coefficient more than 0.9 (positive correlation) or less than -0.9 (negative correlation) were considered significant. Using this correlation coefficient, we only observed correlation relationship between optimal temperature and K_{cat}.

Considering the relationship between optimal temperature and K_{cat}, we sought to verify whether the optimal temperature can be used to cluster our dataset according to catalytic rate. We performed a clustering algorithm called K-means clustering (VanderPlas, 2016). K-means clustering is an unsupervised machine learning algorithm which clusters data points of a multidimensional dataset into n clusters with equal variance (VanderPlas, 2016). To divide a set of N samples X into K disjoint clusters C, the algorithm tries to minimize a criterion known as “inertia” or within-cluster sum-of-square errors which is defined as:

$$\sum_{i=0}^{n} min\left(\left\|x_i - \mu_j\right\|^2\right)$$

where $\mu j \in C$.

To use this algorithm, first, the number of clusters, n, should be specified. Then, the algorithm chooses n random observations from the dataset and assigns them as the initial clusters' centroid. To assign the remaining data points to the nearest cluster, the algorithm calculates the distance between a given data point and the clusters' centroids and chooses the one with the smallest distance. When each data point is assigned to a cluster, then, the centroid of each cluster is updated by averaging the value of all instances in that cluster. Using the new clusters' centroids, the algorithm re-assigned all the datapoints to new clusters and updates the centroids. This process is repeated until the difference between the old and the new centroids do not change significantly (*i.e.*, less than a threshold) (VanderPlas, 2016). Selecting the optimal number of clusters is challenging. Here, we applied a heuristic method known as "Elbow method" to determine the optimal number of clusters for our dataset (Kodinariya & Makwana, 2013). The aim of this mathematical optimization method is to find the "elbow of a curve" where diminishing returns are no longer worth the additional cost. In our case, we first ran the K-means algorithm with default number of clusters (n_clusters = 8) and plotted the "inertia" *vs.* "number of clusters" graph. In this graph, as the number of clusters increases, the inertia decreases, where initially the reduction is significant and slows down as the number of clusters increases. However, at a specific number of clusters, this reduction is not as sharp as before (*i.e.,* there is a sudden change of slope). This point is referred to as the elbow point and specify the optimal number of clusters for our dataset.

To perform this analysis, Python (Version 3.8) (Van Rossum & Drake, 2009) was used. Since K_{cat} and K_m values of our dataset vary within wide ranges, we used $\log K_{cat}$

and $\log K_m$ for simplicity and better visualization of graphs. The K-means clustering was done using Scikit-learn library (Version 0.23.2) (Pedregosa et al., 2012) with default parameters except for the number of clusters, n, which was calculated using Elbow method. The predicted optimal number of clusters was then used to re-run the K-means clustering algorithm. We performed this analysis to divide the dataset based on the optimal temperature alone, optimal pH alone, and both optimal temperature and pH. Using the clustering results, we plotted the $\log K_{cat}$ *vs*. $\log K_m$ to assess the relationship between the enzymatic efficiency and optimal temperature and/or pH of extant AID homologs.

4.3.8 Calculating the predicted protein stability curve of AID homologs

In our dataset, we found two closely related AID homologs (Tsu-AID and Pt-AID) with a significant difference in their optimal temperatures (20 °C). We sought to compare their predicted stability curve using SCooP server (http://babylone.ulb.ac.be/SCooP) (Pucci et al., 2017; Pucci & Rooman, 2014, 2016). SCooP is a fast and accurate tool to estimate the Gibbs-Helmholtz equation of folding process of a protein with known or modeled structure. It predicts the change in enthalpy and in heat capacity upon folding (ΔH_m and ΔC_p, respectively), the melting temperature (T_m), and the folding free energy at room temperature (ΔG_r) (Pucci et al., 2017; Pucci & Rooman, 2014, 2016). T_m measures the thermal stability while ΔG_r can be considered as a descriptor of thermodynamic stability of proteins (Pucci et al., 2014; Pucci et al., 2017; Pucci & Rooman, 2014, 2016). We used 5 predicted computational models of each AID homologs to estimate all the thermodynamic quantities that characterize the folding transition (*i.e.,* ΔH_m, ΔC_p, T_m, and

ΔG_r). The final value for each parameter was reported as the arithmetic mean ± standard error (SEM).

4.3.9 WRC and WGCW motif analyses of other Gadidae species

The Atlantic cod IgV_H sequences obtained in section 3.3.8 were used to extract and annotate the IgV_H regions of other Gadidae species using the raw genomic data of each species (the European Nucleotide Archive (ENA) accession number: PRJEB12469 and the Dryad repository: doi:10.5061/dryad.326r8) (Malmstrom et al., 2016). A similar BLAST protocol to section 3.3.8 was used to obtain the IgV_H regions of other Gadidae species. The WRC and WGCW analyses were exactly done as section 3.3.9 and the GC content of the coding sequences was retrieved from Codon and Codon-Pair Usage Tables (CoCoPUTs) server available at https://hive.biochemistry.gwu.edu/review/codon2 (Alexaki et al., 2019). In these analyses, for each parameter, the average of that parameter for Arctic cod (Ag), Polar cod (Bs), Haddock (Ma), Silvery pout (Ga), and Atlantic cod (Gm) were reported as the "Gadinae" group. Similarly, the average of Arctic cod (Ag), Polar cod (Bs), Haddock (Ma), Silvery pout (Ga), Atlantic cod (Gm), Burbot (Llo), and Forkbeard (Pp) were reported as the "Gadidae" group.

4.4 Results

4.4.1 Biochemical properties of the extant Gadiformes AIDs

4.4.1.1 Selected extant AID homologs for biochemical analyses

To study AID's evolutionary trajectory in the Gadiformes group, we synthesized and characterized the biochemical properties of 36 AID proteins from 35 extant species. Twenty-three of the included species belong to the Gadiformes taxonomic order of bony fish class. Selected non-Gadiformes species were used as comparison points. Figure 4-1 illustrates the protein alignment of these extant AID proteins. To examine the biochemical properties of the extant Gadiformes AIDs, we expressed and purified them as N-terminally tagged GST fusion proteins (Abdouni et al., 2013; Dancyger et al., 2012; Emma M. Quinlan, 2017; Larijani et al., 2007). We compared their optimal temperature, optimal pH, and Michaelis-Menten kinetics parameters to that of non-Gadiformes AIDs using our alkaline cleavage assay (Abdouni et al., 2013; Dancyger et al., 2012; King et al., 2015; Larijani & Martin, 2007; Larijani et al., 2007). In all assays, at least two independent protein preparations of each AID homolog were tested in duplicate. We tested all activity assays on the 7 nucleotide long partially single-stranded bubble substrate containing AID hotspot (TGCbub7). All bony fish AIDs studied thus far favor TGCbub7 as the optimal substrate (Abdouni et al., 2013; Dancyger et al., 2012; Emma M. Quinlan, 2017; Larijani et al., 2007). Interestingly, we did not observe any cytidine deaminase activity for purified AID from *B. saida* (Bs-AID) and *M. zugmayeri* (Mz-AID) in our assays. Both Atlantic cod and *B. saida*, known as polar cod, belongs to Gadinae group of Gadidae (cods) family. However, *M. zugmayeri*, also known as arrowtail cod, belong to Melanonidae (pelagic

cods) family of the Gadidae sister group. There are only three amino acid differences between Gm-AID and Bs-AID: K13, R54, and L143 in Gm-AID *vs.* N13, H54, and P143 in Bs-AID, amongst which L143P seems to be the most drastic amino acid change. This amino acid resides in α4 in Gm-AID and its replacement with a proline in Bs-AID most likely causes a truncated α4.

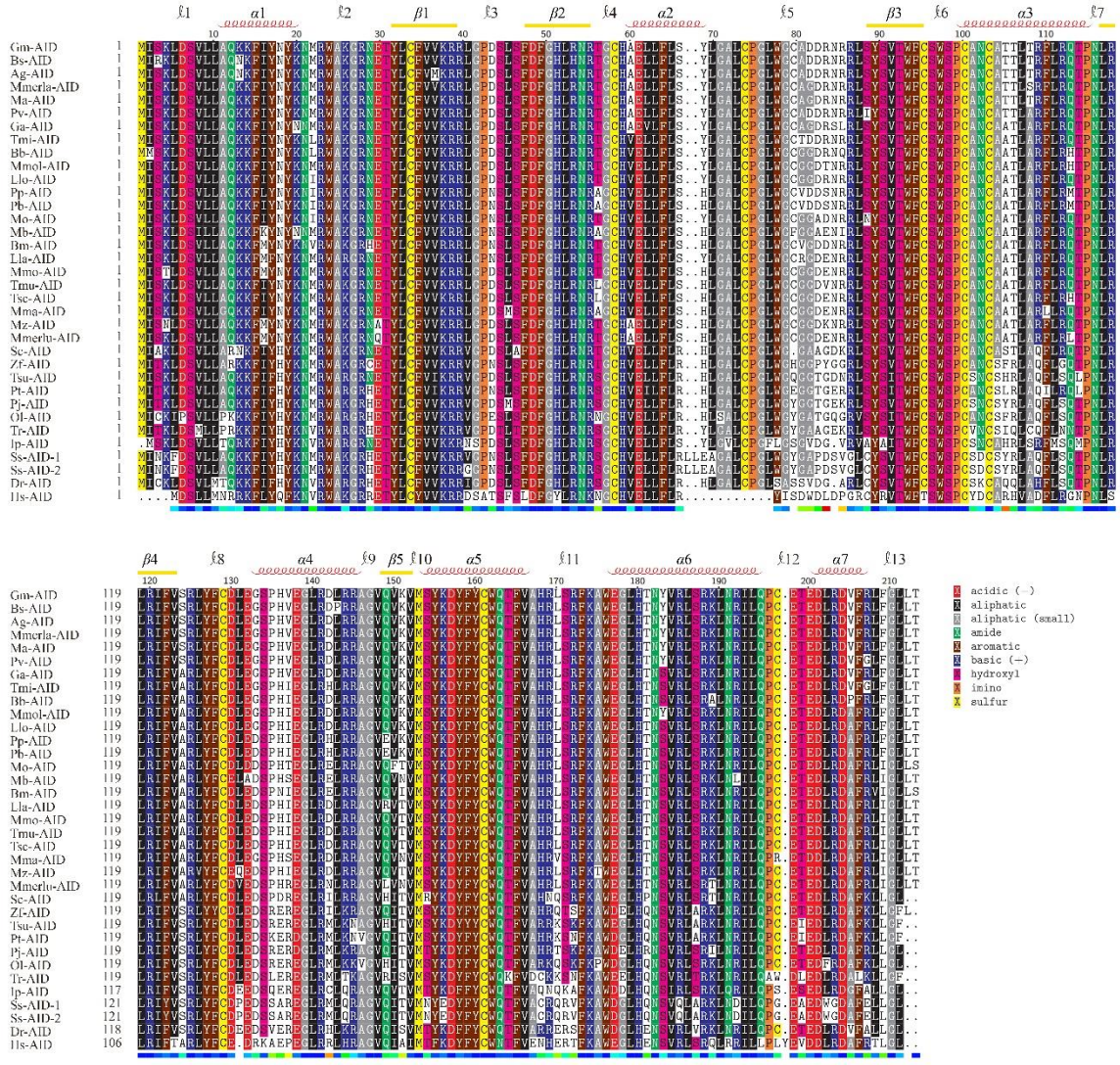


Figure 4-1: Protein alignment of extant AID homologs the enzymatic properties of which were characterized in this chapter. The approximate secondary structure of α-helical (α), β-strand (β), and loop (l) regions are shown. Residues are colored according to chemical properties of the side chain. For abbreviations, refer to Table 4-1.

4.4.1.2 Examining the optimal temperature of extant Gadiformes AIDs

We first examined the optimal temperature of purified AIDs. Since our findings in the previous chapter revealed that Atlantic cod AID is a cold-adapted enzyme, we tested the activity of the purified AIDs in a wide range of -10 to 40 °C. The minimum and maximum incubation temperature as well as the incubation duration were decided based on the preliminary results (not shown). Gm-AID, Ip-AID, Dr-AID and Hs-AID, which have known temperature profile, were also tested as controls (Dancyger et al., 2012; Emma M. Quinlan, 2017).

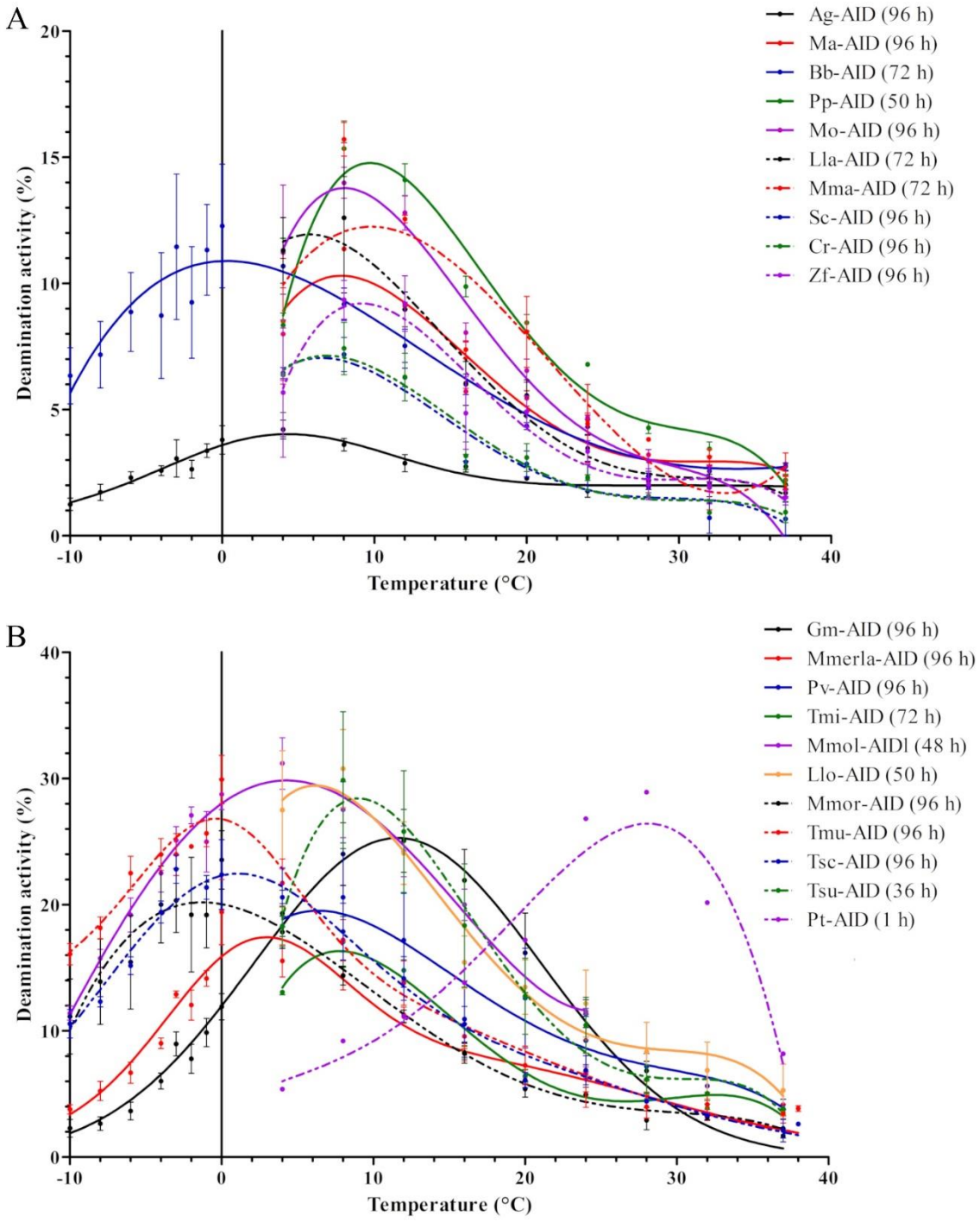
As illustrated in Figure 4-2, the majority of extant Zeiogadaria AIDs are cold-adapted enzymes. Particularly, Mmor-AID, Tmu-AID and Tsc-AID have an optimal temperature of 0 °C (Table 4-4). All the extant species studied in this thesis are marine species except for *T. subterraneus*, *P. transmontane*, *D. rerio*, *O. latipes*, and *I. punctatus* which are freshwater fish (www.fishbase.se; Appendix 6). Among the AID homologs from these species, Tsu-AID has the lowest optimal temperature of about 8 °C, very similar to most of the marine fish in this study. *T. subterraneus*, and *P. transmontane* both belong to the Percopsiformes family; however, their AIDs exhibited a substantial ~ 20 °C difference in their optimal temperature (Pt-AID ~ 28 °C and Tsu-AID ~ 8 °C). These two AIDs have 19 amino acid differences which mostly reside within the α3, α4, and ℓ11 regions (T3S, H29N, N44D, I63L, E79Q, E84D, R85N, A101S, L105H, I110F, R112S, K135R, D138E, V146A, Q149H, F159Y, H168R, N172K, and D177E in Pt-AID *vs.* Tsu-AID; Figure 4-3).

It has been proposed by Nojima and colleagues that proteins may increase their thermoresistance using three main strategies. In the first strategy, the enthalpy change

(ΔH_s) measured at the temperature of maximum stability (T_s) is more negative, causing ΔG for all temperatures to decrease. This strategy can be seen as a curve to be shifted downward (Figure 4-4 A). The second strategy is to increase (less negative) the change in the heat capacity upon folding (ΔC_p) which causes T_m to increase. In this case, the stability curve would broaden (Figure 4-4 B). The third strategy is to increase T_s which shifts the curve to the right (Figure 4-4 C) (Nojima et al., 1978; Pucci & Rooman, 2014; Razvi & Scholtz, 2006). Proteins may apply one, two, or all of these strategies to improve their thermal resistance. For example, it was shown that *Thermus thermophilus* cytochrome c employed the first and third strategies while *T. thermophilus* phosphoglycerate kinase has achieved higher thermo stability by using the second strategy, with some contribution from the first strategy (Nojima et al., 1978; Nojima et al., 1977). In the case of *Thermococcus kodakaraensis* O^6-methyl-guanine-DNAmethylytransferase, all three strategies were used to enhance thermal stability (Shiraki et al., 2001).

To investigate the strategies used by Pt-AID to acquire higher optimal temperature compared with that of Tsu-AID, we used SCooP web interface to predict the stability curves of five computationally predicted models for each AID (Pucci et al., 2017; Pucci & Rooman, 2014, 2016). The predicted thermodynamic parameters of Pt-AID and Tsu-AID are summarized in Table 4-3. Consistent with higher temperature of Pt-AID compared with that of Tsu-AID, the predicted folding free energy value at room temperature (ΔG_r) and the change in enthalpy upon folding (ΔH_m) for Pt-AID were lower than that of Tsu-AID (-6.52 ± 0.409 *vs*. -5.38 ± 0.132 for ΔG_r and -79.32 ± 2.039 *vs.* -74.72 ± 1.602 for ΔH_m), suggesting more thermodynamic stability for Pt-AID. Although, the ΔC_p of Pt-AID was

less negative than that of Tsu-AID (-1.176 ± 0.1264 *vs*. -1.43 ± 0.0498), the predicted T_m of Pt-AID was lower than that of Tsu-AID (64.72 ± 0.991 *vs.* 68.76 ± 0.289). As illustrated in Figure 4-5, the T_s of Pt-AID was also lower than that of Tsu-AID (~ 4 °C *vs.* ~ 22 °C). T_s is the temperature of maximum stability. Based on these observations, it seems that Pt-AID has only applied the first strategy to increase its thermoresistance compared with Tsu-AID. However further studies are required to pinpoint the mutation(s) responsible for Pt-AID thermoresistance.


A
Deamination activity (%)
Temperature (°C)
Ag-AID (96 h)
Ma-AID (96 h)
Bb-AID (72 h)
Pp-AID (50 h)
Mo-AID (96 h)
Lla-AID (72 h)
Mma-AID (72 h)
Sc-AID (96 h)
Cr-AID (96 h)
Zf-AID (96 h)
B
Deamination activity (%)
Temperature (°C)
Gm-AID (96 h)
Mmerla-AID (96 h)
Pv-AID (96 h)
Tmi-AID (72 h)
Mmol-AIDl (48 h)
Llo-AID (50 h)
Mmor-AID (96 h)
Tmu-AID (96 h)
Tsc-AID (96 h)
Tsu-AID (36 h)
Pt-AID (1 h)

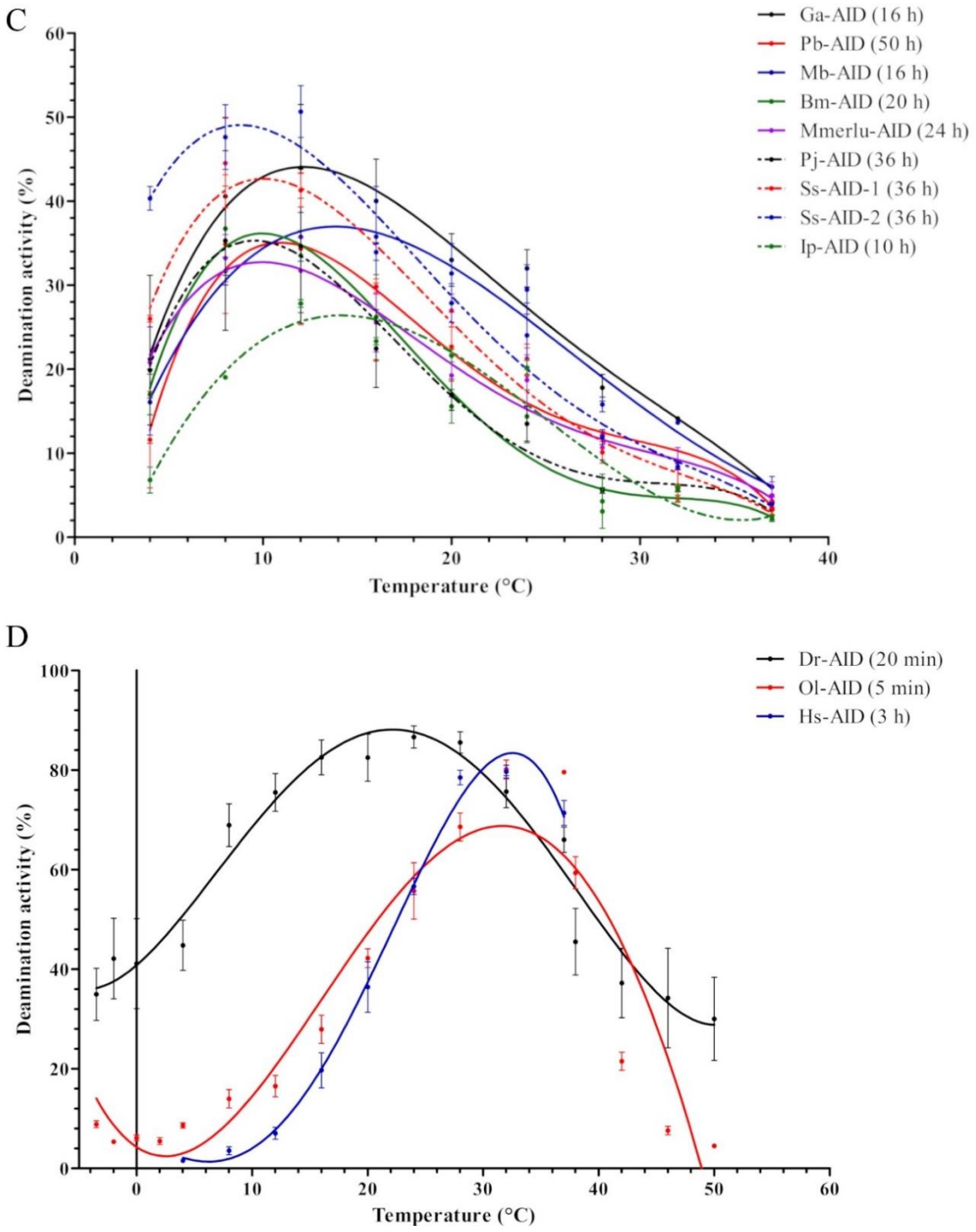


Figure 4-2: Temperature profile of extant AID homologs. The optimal temperature of each AID was assessed using our standard alkaline cleavage assay and bub7TGC substrate. The incubation duration, minimum, and maximum temperature limits were tailored to the activity level of each purified AID obtained in the preliminary results. For better representation, results were graphed based on the AIDs' activity level. A through D show AIDs with low to high activity levels. Data is graphed as mean ± SEM (n=4). For abbreviations, refer to Table 4-1.

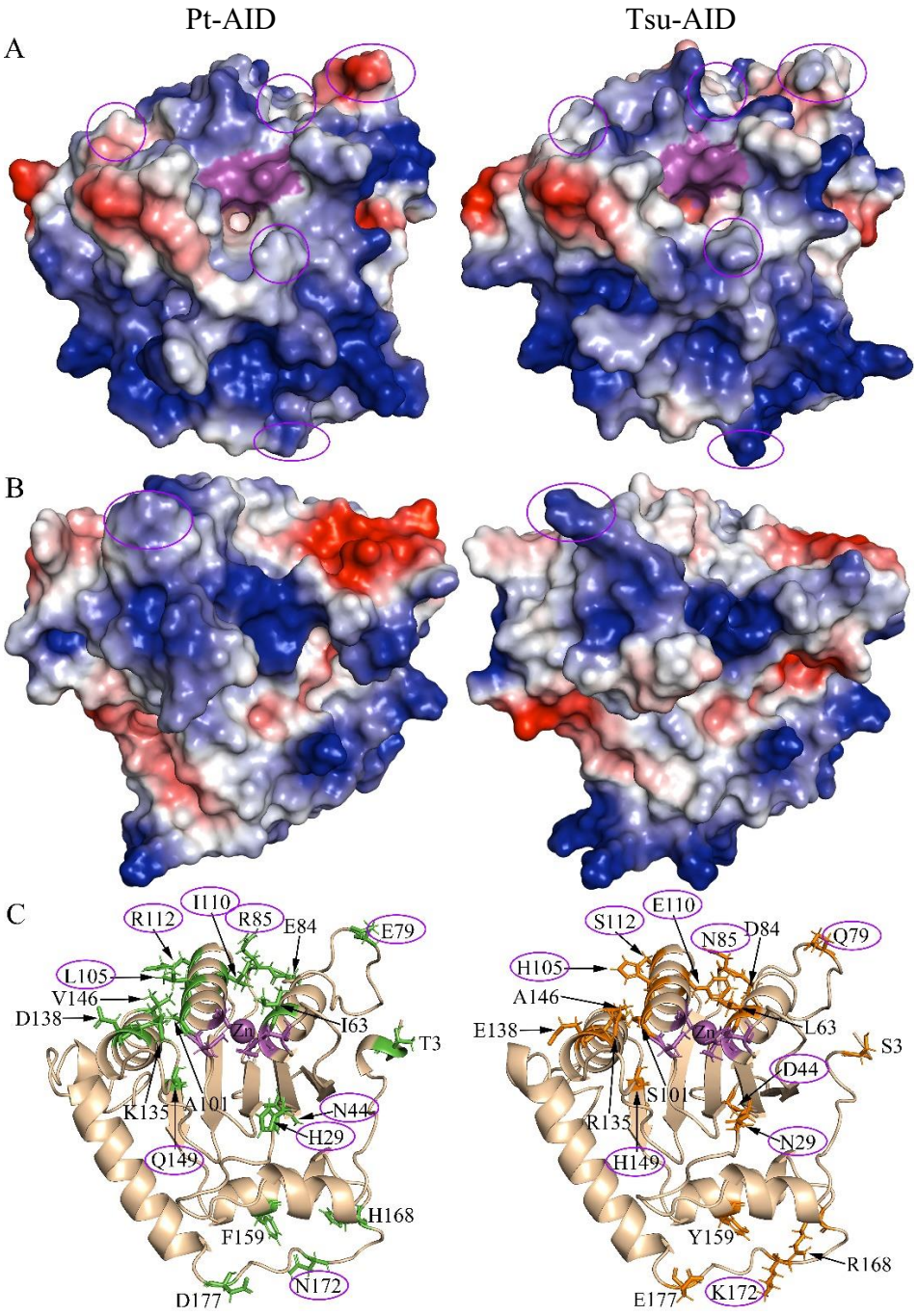


Figure 4-3: Predicted 3D structure of Pt-AID (left) vs. Tsu-AID (right). Predicted surface topology of Pt-AID and Tsu-AID were compared. Panel A and B illustrate the front and the back view of their predicted surface topology, respectively. Positive, neutral, and negative residues are colored blue, white, and red, respectively. The putative catalytic pocket is colored in purple. C) Representative ribbon model of their predicated 3D structures. Positions containing different amino acids between the two AIDs are labeled. The purple circles show positions that are occupied with differently charged amino acids amongst these AIDs.

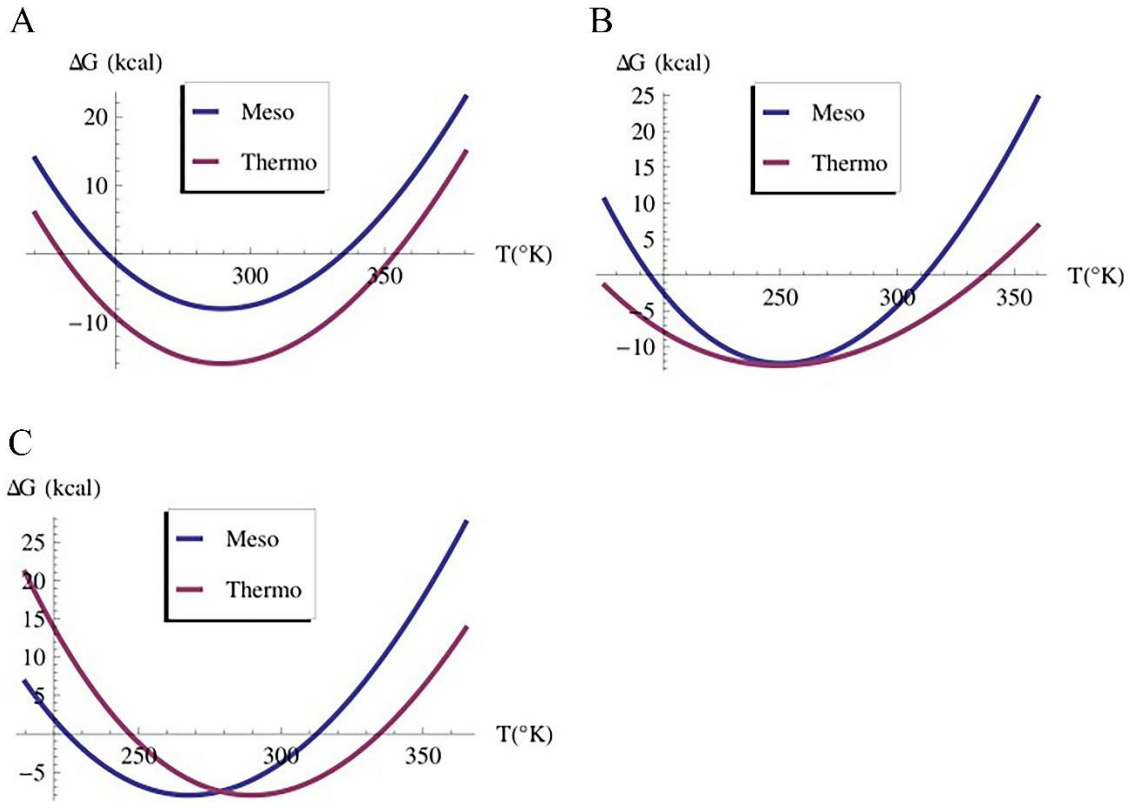


Figure 4-4: Main thermal adaptation strategies employed by proteins. Proteins can modify their thermoresistance through changing A) ΔH_s, B) ΔC_p, and/or C) T_s. Graphs represent the stability curve of hypothetical mesostable (Meso) and thermostable (Thermo) proteins. Adapted from Pucci and Rooman, 2014. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Table 4-3: Predicted thermodynamic quantities of Pt-AID and Tsu-AID using SCooP server

	Optimal temp. (°C)	ΔH_m (kcal/mol)	ΔC_p (kcal/(mol K))	T_m (°C)	ΔG_r (kcal/mol)
Pt-AID	28	-79.32 ± 2.039	-1.176 ± 0.1264	64.72 ± 0.991	-6.52 ± 0.409
Tsu-AID	8	-74.72 ± 1.602	-1.43 ± 0.0498	68.76 ± 0.289	-5.38 ± 0.132

Abbreviations: Pt-AID: sand roller (*P. transmontane*) AID; Tsu-AID: Southern cavefish (*T. subterraneus*) AID.

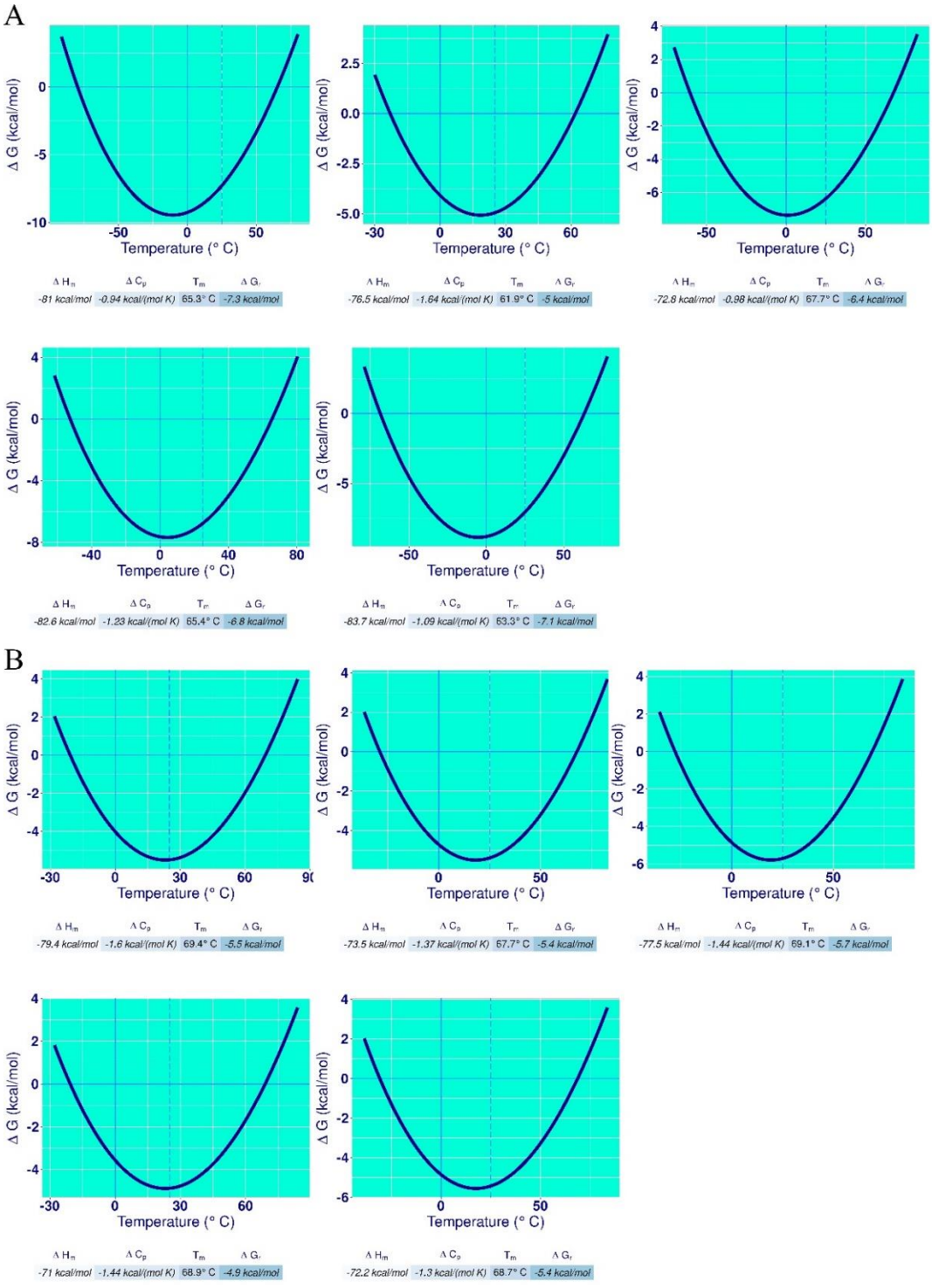
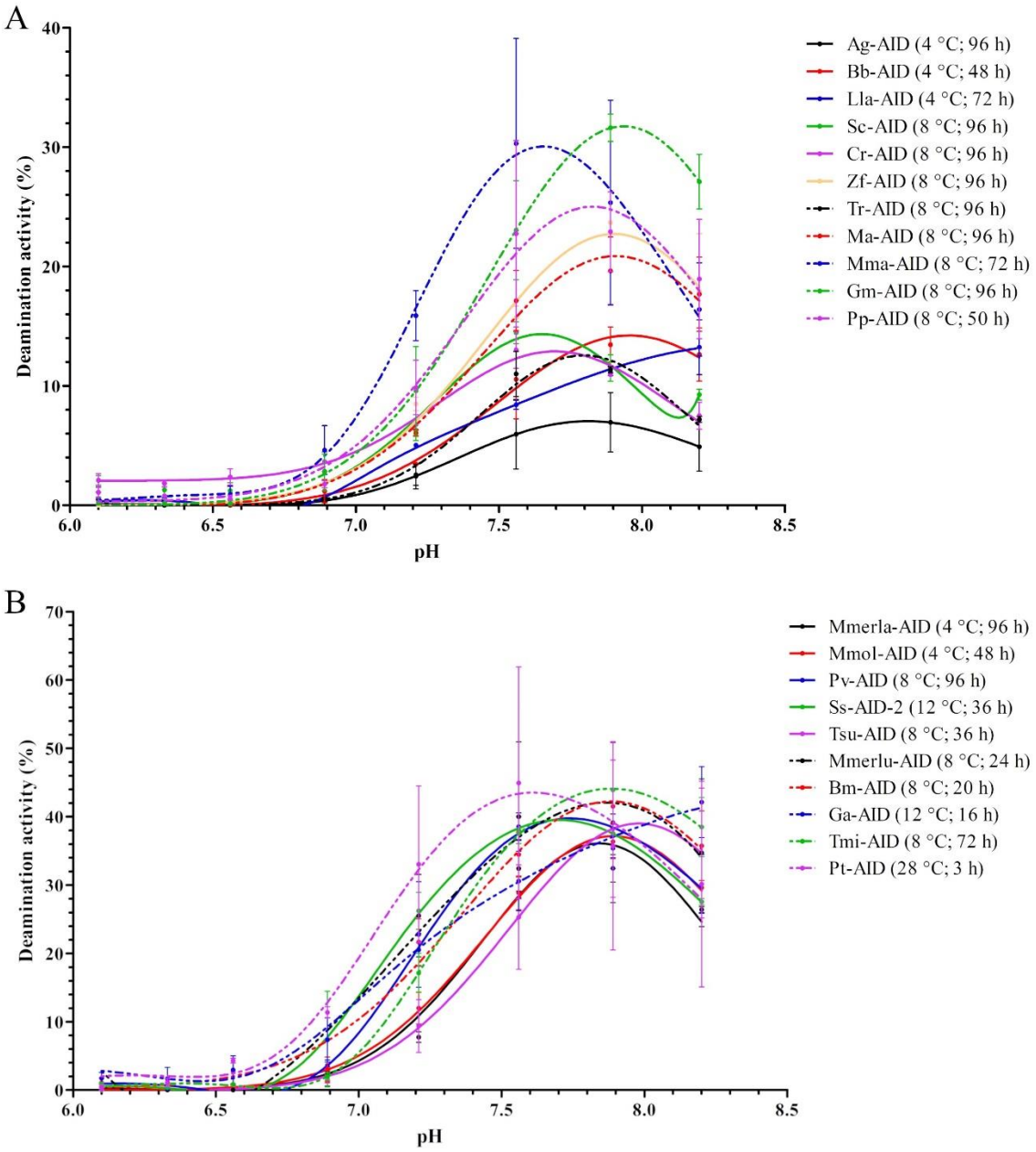


Figure 4-5: Predicted stability curves for A) Pt-AID and B) Tsu-AID. Five predicted models of each AID homolog were submitted to the SCooP server. Please note that the scales on both axes vary between panels. Abbreviations: Pt-AID: sand roller (P. transmontane) AID; Tsu-AID: Southern cavefish (T. subterraneus) AID.

4.4.1.3 Examining the optimal pH of extant Gadiformes AIDs

We then examined the pH profile of purified AIDs at their corresponding optimal temperature using phosphate buffer with effective pH ranging from 6.1 to 8.2 (Figure 4-6). The pH profile of Gm-AID, Ip-AID, Dr-AID, and Hs-AID were also tested as known controls. The optimal pH obtained for the controls here was consistent with the data from the previous chapter (section 3-4-1). Our results showed that AIDs with lower optimal temperature generally tend to have a higher optimal pH and vice versa (Figure 4-7 and Table 4-4). However, this trend is not absolute since we also observed AID homologs with similar optimal temperature but different optimal pH, such as Bb-AID, Ag-AID and Mmol-AID, and Mmerla-AID with optimal temperature of 4 °C but optimal pH of 8.1, 7.9, and 7.8, respectively. Also, amongst the AID homologs with optimal temperature of 8 °C, we found optimal pH of 8.2 (Pj-AID), 8.1 (Gm-AID, Mo-AID, and Llo-AID), 7.9 (Ma-AID, Tmi-AID, Pp-AID, Pb-AID, Bm-AID, Zf-AID, and Tsu-AID), 7.8 (Mmerlu-AID and Tr-AID), 7.7 (Pv-AID), and 7.6 (La-AID, Mma-AID, Sc-AID, and Cr-AID). The optimal temperature and pH of Ga-AID were measure at 12 °C and 8.2, while the pH of Mb-AID and Ss-AID, which exhibited a similar optimal temperature, was measured at 7.9. We also found AID homologs with similar optimal pH exhibiting distinct optimal temperatures. For example, amongst AIDs with optimal pH of 8.1, we found AIDs with optimal temperature of 0 °C (Tsc-AID, Tmu-AID, and Mmor-AID), 4 °C (Bb-AID), and 8 °C (Gm-AID, Mo-AID, and Llo-AID). The AID homologs with optimal pH of 7.9, revealed optimal temperature of 4 °C (Ag-AID and Mmol-AID), 8 °C (Ma-AID, Tmi-AID, Pp-AID, Pb-AID, Bm-AID, Zf-AID, and Tsu-AID), 12 °C (Mb-AID and Ss-AID), and 14 °C (Ip-AID).

Additionally, although the optimal pH of Dr-AID, Pt-AID, and Ol-AID were measured at 7.6, their optimal temperatures were estimated as 25 °C, 28 °C, and 32 °C, respectively.


A
Deamination activity (%)
pH
Ag-AID (4 °C; 96 h)
Bb-AID (4 °C; 48 h)
Lla-AID (4 °C; 72 h)
Sc-AID (8 °C; 96 h)
Cr-AID (8 °C; 96 h)
Zf-AID (8 °C; 96 h)
Tr-AID (8 °C; 96 h)
Ma-AID (8 °C; 96 h)
Mma-AID (8 °C; 72 h)
Gm-AID (8 °C; 96 h)
Pp-AID (8 °C; 50 h)
B
Deamination activity (%)
pH
Mmerla-AID (4 °C; 96 h)
Mmol-AID (4 °C; 48 h)
Pv-AID (8 °C; 96 h)
Ss-AID-2 (12 °C; 36 h)
Tsu-AID (8 °C; 36 h)
Mmerlu-AID (8 °C; 24 h)
Bm-AID (8 °C; 20 h)
Ga-AID (12 °C; 16 h)
Tmi-AID (8 °C; 72 h)
Pt-AID (28 °C; 3 h)

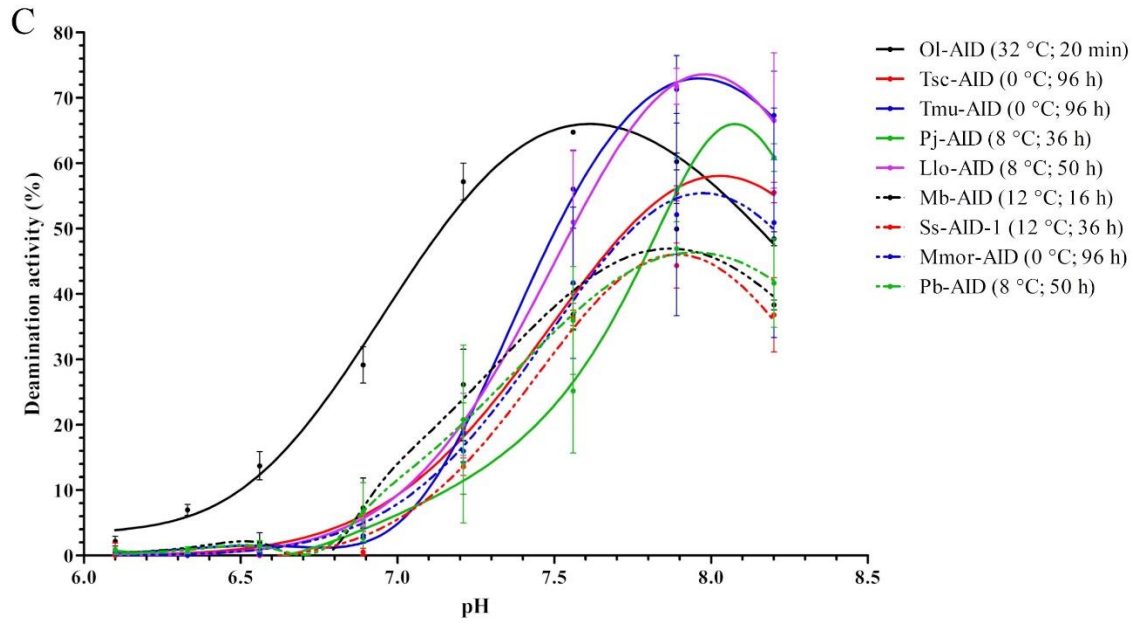


Figure 4-6: pH profile of extant AID homologs. The optimal pH of each AID was assessed using our standard alkaline cleavage assay and bub7TGC substrate in their corresponding optimal temperature. The incubation time for each AID homolog was decided based on its activity level. For better representation, results were graphed based on the AIDs' activity level. A through C show AIDs with low to high activity levels. Data is graphed as mean ± SEM (n=4). For abbreviations, refer to Table 4-1.

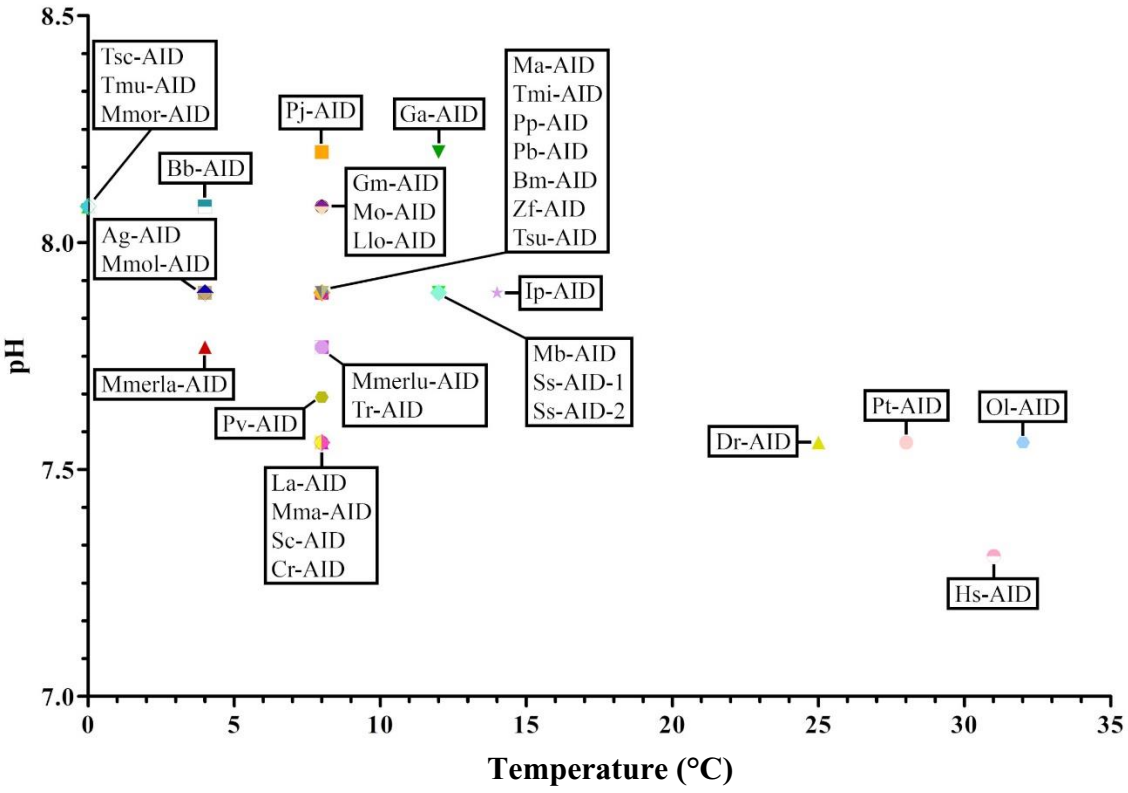
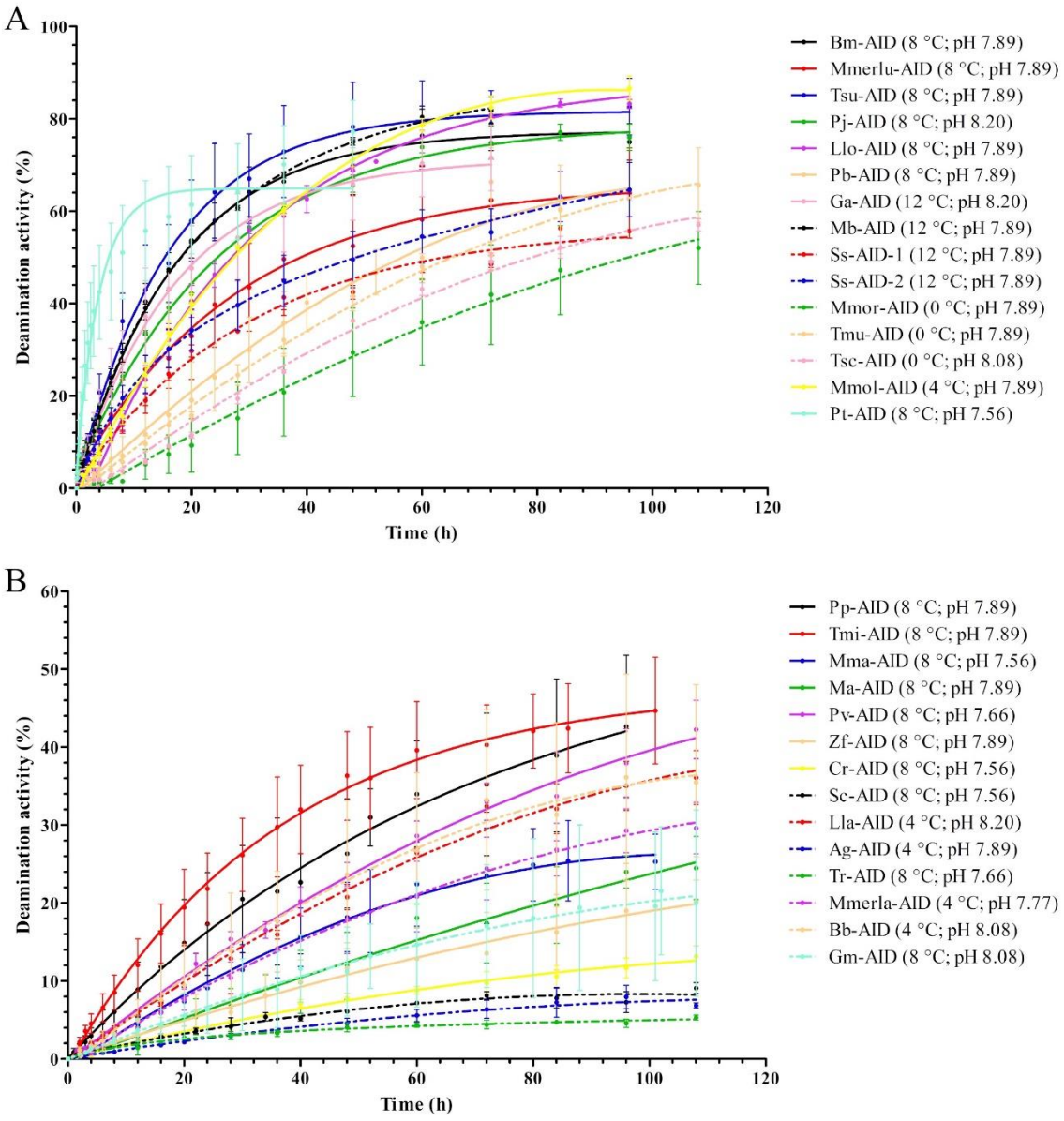


Figure 4-7: Optimal pH vs. optimal temperature of extant AID homologs. For abbreviations, refer to Table 4-1.

4.4.1.4 Examining the catalytic properties of extant Gadiformes AIDs

To compare the catalytic activity of the Gadiformes AIDs to that of other extant species, we conducted standard Michaelis-Menten kinetics. In preparation for Michaelis-Menten kinetics, we performed a time-course experiment to estimate the proper incubation time when the AID activity falls within the initial velocity. Gm-AID was tested alongside other extant AID homologs as a control. The time-course experiment was done in the corresponding optimal pH and temperature of each AID homolog (Figure 4-8).


A
Deamination activity (%)
Time (h)
Bm-AID (8 °C; pH 7.89)
Mmerlu-AID (8 °C; pH 7.89)
Tsu-AID (8 °C; pH 7.89)
Pj-AID (8 °C; pH 8.20)
Llo-AID (8 °C; pH 7.89)
Pb-AID (8 °C; pH 7.89)
Ga-AID (12 °C; pH 8.20)
Mb-AID (12 °C; pH 7.89)
Ss-AID-1 (12 °C; pH 7.89)
Ss-AID-2 (12 °C; pH 7.89)
Mmor-AID (0 °C; pH 7.89)
Tmu-AID (0 °C; pH 7.89)
Tsc-AID (0 °C; pH 8.08)
Mmol-AID (4 °C; pH 7.89)
Pt-AID (8 °C; pH 7.56)
B
Deamination activity (%)
Time (h)
Pp-AID (8 °C; pH 7.89)
Tmi-AID (8 °C; pH 7.89)
Mma-AID (8 °C; pH 7.56)
Ma-AID (8 °C; pH 7.89)
Pv-AID (8 °C; pH 7.66)
Zf-AID (8 °C; pH 7.89)
Cr-AID (8 °C; pH 7.56)
Sc-AID (8 °C; pH 7.56)
Lla-AID (4 °C; pH 8.20)
Ag-AID (4 °C; pH 7.89)
Tr-AID (8 °C; pH 7.66)
Mmerla-AID (4 °C; pH 7.77)
Bb-AID (4 °C; pH 8.08)
Gm-AID (8 °C; pH 8.08)

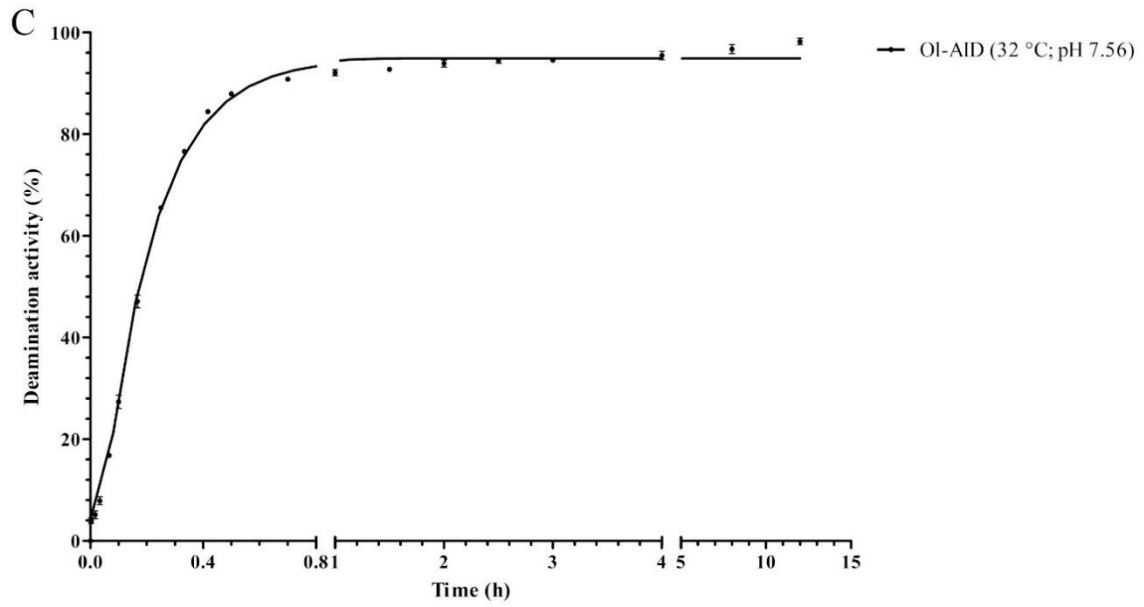
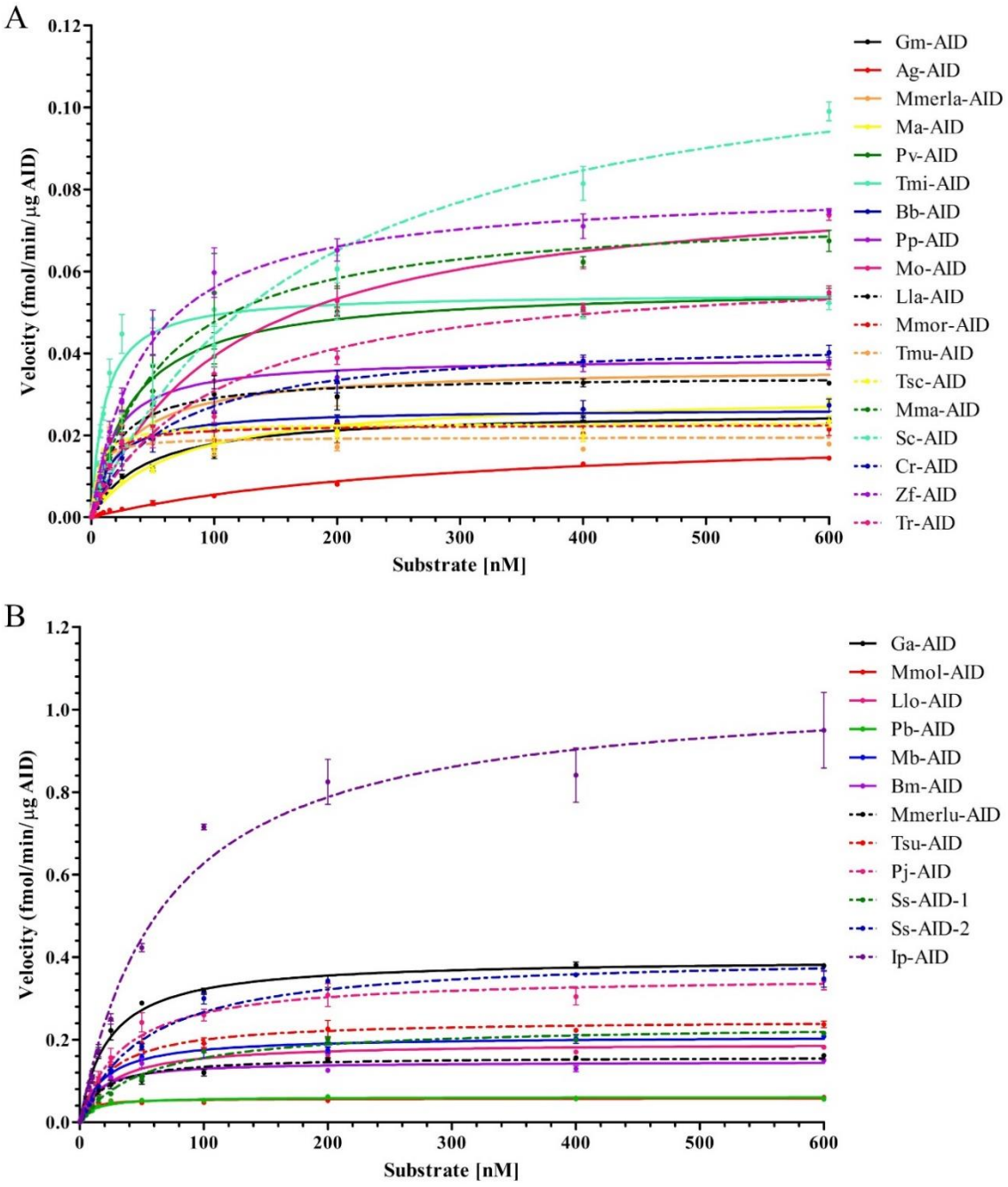


Figure 4-8: Time-course experiment. Catalytic activity of each AID homolog over time was measured at its corresponding optimal pH and temperature. These results were used to estimate the incubation duration of the following Michaelis-Menten kinetics assay for each AID homolog. For better visualization, data is graphed based on the AIDs activity level. A through C correspond to AIDs with low, medium, and high activity levels, respectively. The error bars represent SEM (n=4). For abbreviations, refer to Table 4-1.

Using the time-course results, we conducted a standard Michaelis-Menten kinetics to quantitatively compare the enzymatic activity of extant Gadiformes AIDs (Figure 4-9 and Figure 4-10). At least, two independent protein preparations of each AID homolog were tested in duplicate on bub7TGC substrate. The biochemical properties of extant AID proteins examined in this thesis are summarized in Table 4-4. We measured the maximum velocity (*i.e.*, maximum reaction rate that was achieved in reaction [V_{max}]), the affinity of enzyme for its substrate (*i.e.*, the Michaelis constant which is the substrate concentration at which the enzyme operates at one half of its maximum velocity [K_m]), the turnover number (*i.e.*, the catalytic constant which is the number of catalytic cycles that each active site undergoes per unit time [K_{cat}]), and the catalytic efficiency (*i.e.*, the enzyme's overall ability to converts substrate to product [K_{cat}/K_m]). It should be noted that in the context of AID, K_m could be considered as a measure of target dC positioning in the catalytic pocket.

We found that, on average, the catalytic efficiency of the Gadinae species (1.77e-07) is slightly less than the rest of Gadiformes lineage (2.71e-07). We also observed a strong positive correlation ($r_{Temp.,log(K_{cat})}$ = 0.95) between the optimal temperature and the $\log K_{cat}$ of the extant AIDs analyzed here (Figure 4-11). To confirm this correlation, we also applied K-means clustering, an unsupervised machine learning clustering algorithm, to divide the dataset into discrete groups based on their optimal temperature. We tested the scenarios where the dataset was divided into two to eight clusters. The Elbow methods revealed that three is the optimal number of the clusters for our dataset (Figure 4-12 A). We then categorized our dataset into three groups based on optimal temperature

using K-means clustering model (Figure 4-12 B). This clustering was used to group AID species in the $\log K_{cat}$ *vs.* $\log K_m$ plot (Figure 4-12 C). The analyses revealed that clustering based on optimal temperature was mostly successful in clustering AID homologs based on their enzymatic efficiency. We then examined whether considering the optimal pH of extant AIDs could improve the clustering results. Including the optimal pH in the clustering analyses did not affect the accuracy of $\log \mathrm{K_{cat}}$ vs. $\log \mathrm{K_m}$ graph obtained when only optimal temperature was considered (Figure 4-13). Moreover, considering optimal pH as the clustering parameter failed to properly divide the extant AIDs according to their catalytic efficiency (Figure 4-14). Therefore, we concluded that optimal pH is not a determining factor in catalytic efficiency of AIDs studied herein. These results, indirectly, further confirmed our previously observed strong positive correlation between optimal temperature and K_{cat}. These findings suggest that the two biochemical characteristics of low temperature adaptation and low catalytic rate might be associated in Gadiformes AIDs, and that perhaps while Gadifomes AIDs adapted to lower temperature, they experienced a significant reduction in their enzymatic efficiency.


A
Velocity (fmol/min/µg AID)
Substrate [nM]
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0
100
200
300
400
500
600
Gm-AID
Ag-AID
Mmerla-AID
Ma-AID
Pv-AID
Tmi-AID
Bb-AID
Pp-AID
Mo-AID
Lla-AID
Mmor-AID
Tmu-AID
Tsc-AID
Mma-AID
Sc-AID
Cr-AID
Zf-AID
Tr-AID
B
Velocity (fmol/min/µg AID)
Substrate [nM]
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
100
200
300
400
500
600
Ga-AID
Mmol-AID
Llo-AID
Pb-AID
Mb-AID
Bm-AID
Mmerlu-AID
Tsu-AID
Pj-AID
Ss-AID-1
Ss-AID-2
Ip-AID

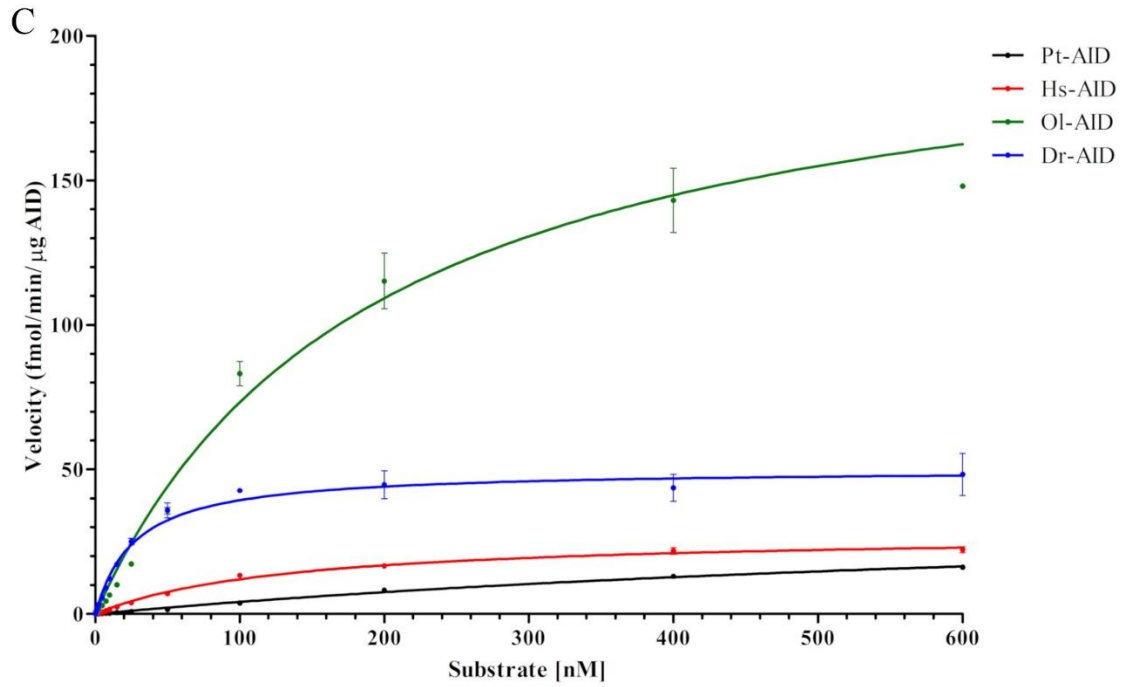


Figure 4-9: Comparison of the catalytic rate of Gadiformes AIDs with other AID homologs. A) The catalytic rate of Gadiformes AIDs was compared to that of other AID homologs through Michaelis-Menten kinetics. At least two independent protein preparations of each AID homolog were incubated at their optimal pH and temperature with 0.03125-600 fmol range of TGCbub7 substrate. Each reaction was carried out in duplicate. For better visual representation, the data is graphed based on the AIDs' activity level. A through C show AIDs with low to high activity levels. Data is represented as mean ± SEM ($n \geq 4$). For abbreviations, refer to Table 4-1.

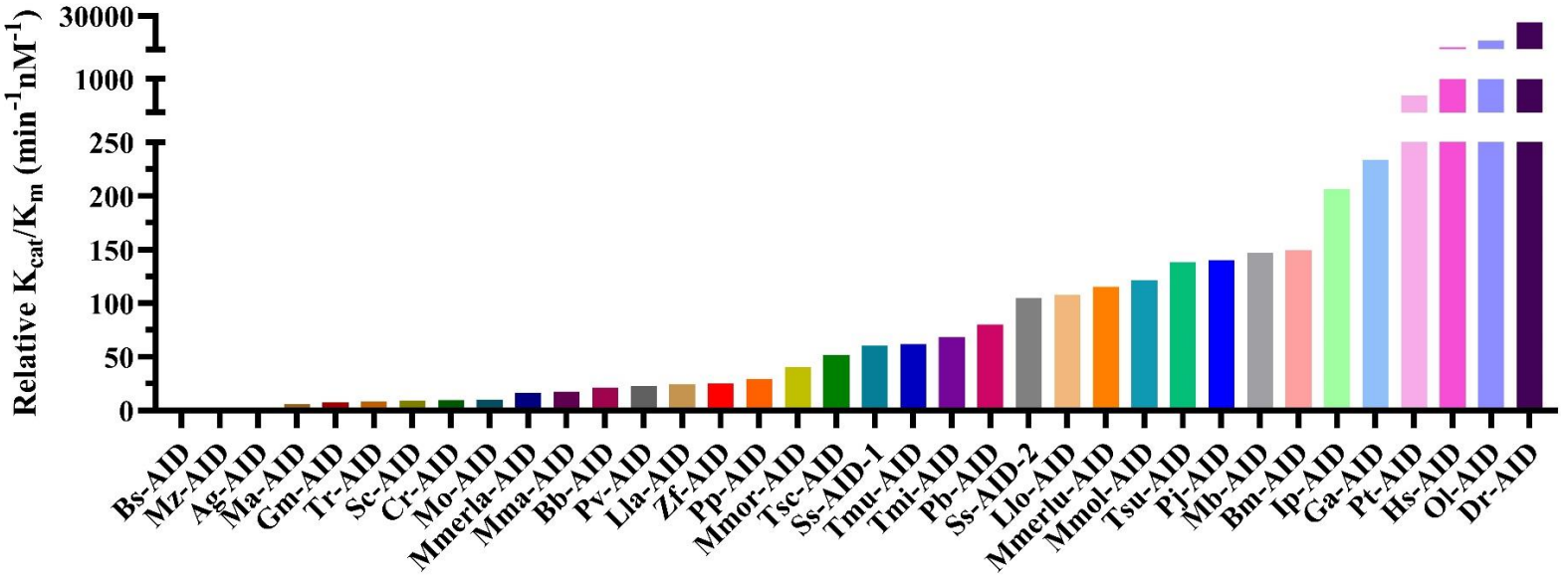


Figure 4-10: Relative catalytic efficiency of all AID homologs examined here. For a better comparison, the catalytic efficiency (Kcat/Km) of AID homologs were reported relative to the value of this parameter for Ag-AID (AID with lowest non-zero catalytic efficiency). Therefore, the relative catalytic efficiency of Ag-AID is set to 1. We did not detect ant cytidine deaminase activity for Bs-AID and Mz-AID. For the list of abbreviations, please refer to Table 4-1.

Table 4-4: The enzymatic parameters measured for extant AID homologs examined in this thesis

	Temp. (°C)	pH	K_{cat} (min^{-1})	K_m (nM)	V_{max} (fmol/min/µg)	Std. Error		R^2	K_{cat}/K_m ($min^{-1}nM^{-1}$)			
						K_{cat} (min^{-1})	K_m (nM)					
Gm-AID	8	8.08	1.36E-06	44.05	0.026	3.05E-08	3.421	0.97	3.09E-08	Gadidae	Gadiformes	Zeiogadaria
Bs-AID	No detectable cytidine deaminase activity was observed in our assays.											
Ag-AID	4	7.89	1.14E-06	295.1	0.022	4.99E-08	27.56	0.98	3.88E-09			
Mmerla-AID	4	7.77	1.92E-06	29.47	0.036	4.74E-08	2.653	0.96	6.50E-08			
Ma-AID	8	7.89	1.57E-06	67.47	0.030	3.99E-08	5.549	0.97	2.33E-08			
Pv-AID	8	7.66	2.96E-06	33.15	0.056	4.60E-08	1.849	0.99	8.93E-08			
Ga-AID	12	8.2	2.07E-05	22.86	0.396	3.27E-07	1.345	0.99	9.07E-07			
Tmi-AID	8	7.89	2.87E-06	10.83	0.055	6.44E-08	0.9843	0.96	2.65E-07			
Bb-AID	4	8.08	1.39E-06	16.73	0.026	3.72E-08	1.725	0.95	8.29E-08			
Mmol-AID	4	7.89	3.03E-06	6.438	0.058	7.39E-08	0.6859	0.94	4.71E-07			
Llo-AID	8	8.08	1.01E-05	24.12	0.192	1.92E-07	1.7	0.98	4.18E-07			

Pp-AID	8	7.89	2.05E-06	17.77	0.039	4.15E-08	1.379	0.97	1.15E-07			
Pb-AID	8	7.89	3.21E-06	10.31	0.061	7.02E-08	0.9191	0.96	3.11E-07			
Mo-AID	8	8.08	4.34E-06	110.9	0.083	1.23E-07	9.595	0.98	3.91E-08	Gadidae sister group		
Mb-AID	12	7.89	1.09E-05	19.12	0.209	1.96E-07	1.307	0.98	5.71E-07			
Bm-AID	8	7.89	7.70E-06	13.28	0.147	1.72E-07	1.169	0.94	5.80E-07			
Lla-AID	8	7.56	1.81E-06	18.92	0.034	5.15E-08	2.041	0.94	9.59E-08			
Mmor-AID	0	8.08	1.19E-06	7.585	0.023	4.94E-08	1.344	0.86	1.56E-07			
Tmu-AID	0	8.08	1.02E-06	4.274	0.020	3.64E-08	0.7125	0.88	2.40E-07			
Tsc-AID	0	8.08	1.20E-06	5.972	0.023	4.21E-08	0.9249	0.89	2.02E-07			
Mma-AID	8	7.56	3.93E-06	57.69	0.075	1.33E-07	6.495	0.95	6.81E-08			
Mz-AID	No detectable cytidine deaminase activity was observed in our assays.											
Mmerlu-AID	8	7.77	8.33E-06	18.6	0.159	1.79E-07	1.521	0.97	4.48E-07			
Sc-AID	8	7.56	6.30E-06	172.6	0.121	3.17E-07	22.18	0.96	3.65E-08			
Cr-AID	8	7.56	2.30E-06	61.05	0.044	7.60E-08	6.658	0.96	3.76E-08			
Zf-AID	8	7.89	4.22E-06	43.6	0.080	9.69E-08	3.475	0.97	9.68E-08			

Tsu-AID	8	7.89	1.30E-05	24.25	0.248	3.18E-07	2.202	0.96	5.36E-07
Pt-AID	28	7.56	0.002082	848.4	39.74	0.0001989	122.4	0.98	2.45E-06
Pj-AID	8	8.2	1.86E-05	34.27	0.355	5.40E-07	3.554	0.95	5.44E-07
Dr-AID	25	7.56	0.002612	27.16	50.08	8.31E-05	3.104	0.95	9.62E-05
Ss-AID-1	12	7.89	1.24E-05	52.73	0.238	1.87E-07	2.673	0.98	2.36E-07
Ss-AID-2	12	7.89	2.12E-05	51.92	0.405	3.91E-07	3.253	0.98	4.07E-07
Ol-AID	32	7.56	0.03874	1169	737.7	0.006819	285.2	0.97	3.31E-05
Ip-AID	14	7.89	5.50E-05	68.77	1.058	1.62E-06	6.52	0.97	8.00E-07
Tr-AID	8	7.77	3.27E-06	101.8	0.062	1.07E-07	9.85	0.97	3.21E-08
Hs-AID	31	7.31	0.001448	133.8	28.130	3.72E-05	9.465	0.98	1.08E-05

For abbreviations, refer to Table 4-1.

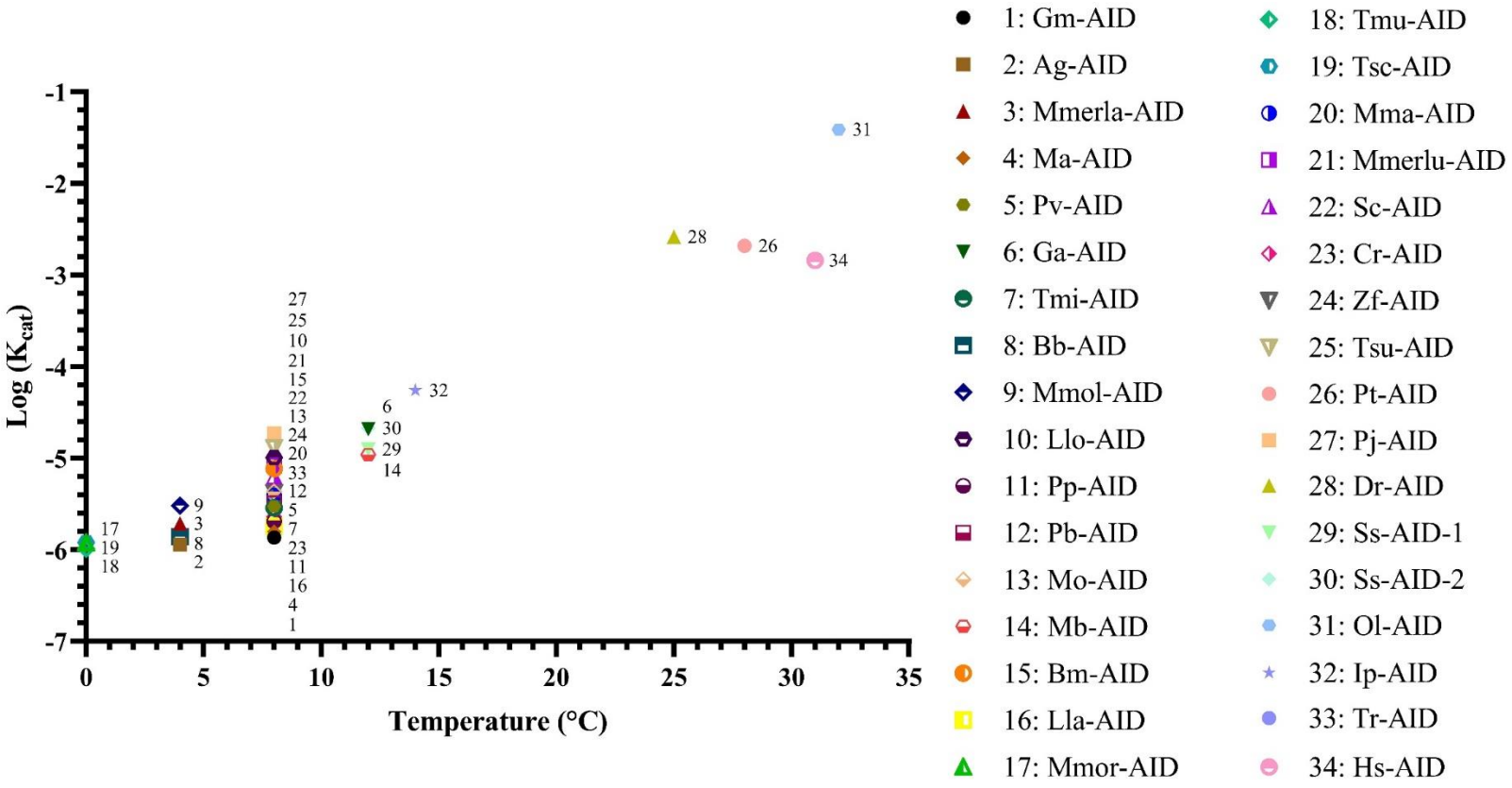


Figure 4-11: The relationship between optimal temperature and log K_{cat} of extant AID homologs studied here. For abbreviations, refer to Table 4-1.

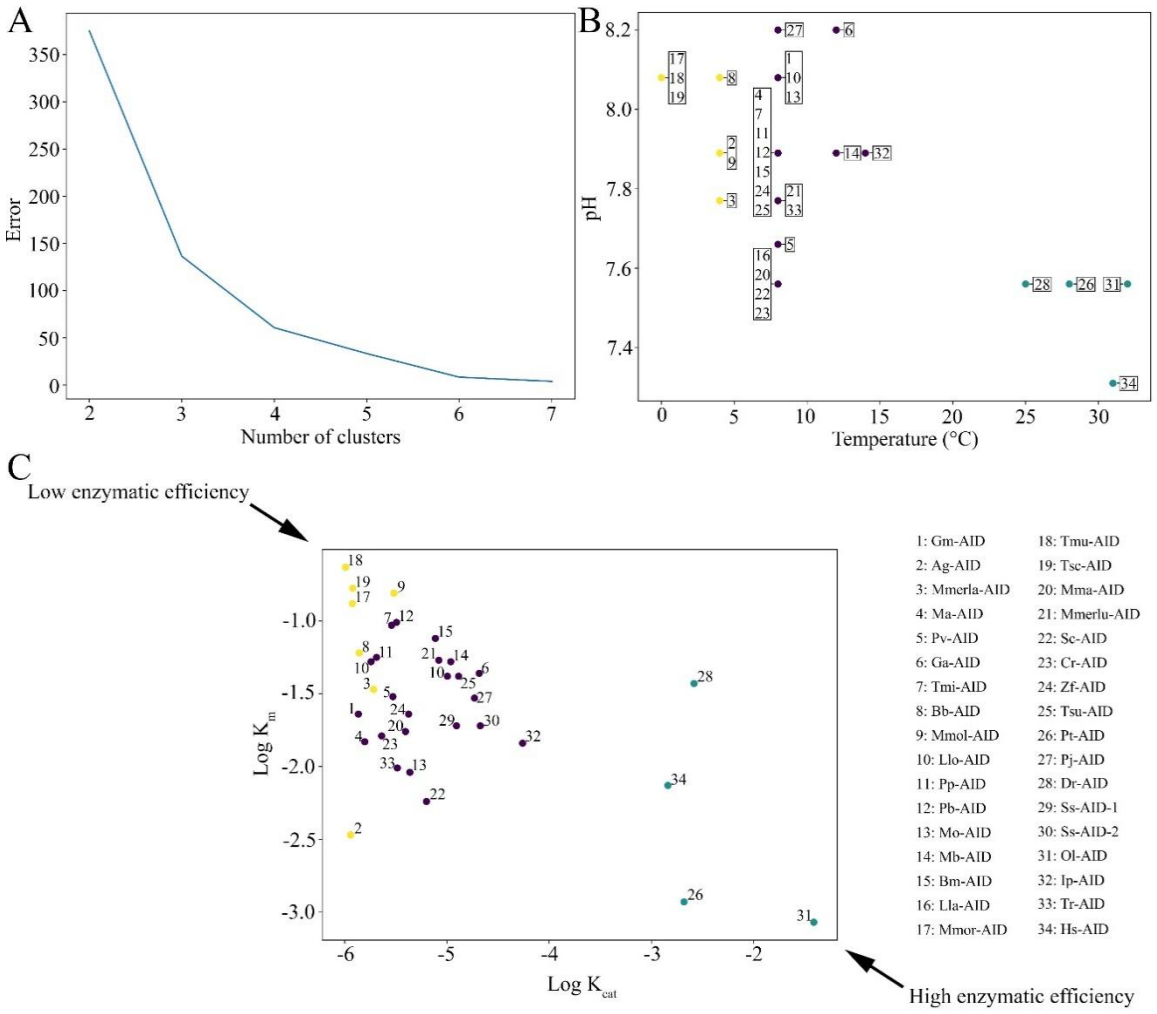


Figure 4-12: Clustering of extant AIDs based on their optimal temperature using machine learning algorithm of K-means clustering. A) The optimal number of clusters was estimated as three according to the K-means clustering model and elbow method. The K-means error was calculated for a given number of clusters (n = 2 to 8). On the error vs. number of clusters graph, the number of clusters where the "elbow" is bent was considered as the optimal number of clusters for the dataset. B) The dataset was divided into three distinct clusters based on the optimal temperature using K-means clustering model. The three clusters are colored cyan, violet, and yellow. C) The catalytic efficiency of the extant AID proteins was compared amongst the three clusters. The color scheme in B and C sections are the same. For abbreviations, refer to Table 4-1.

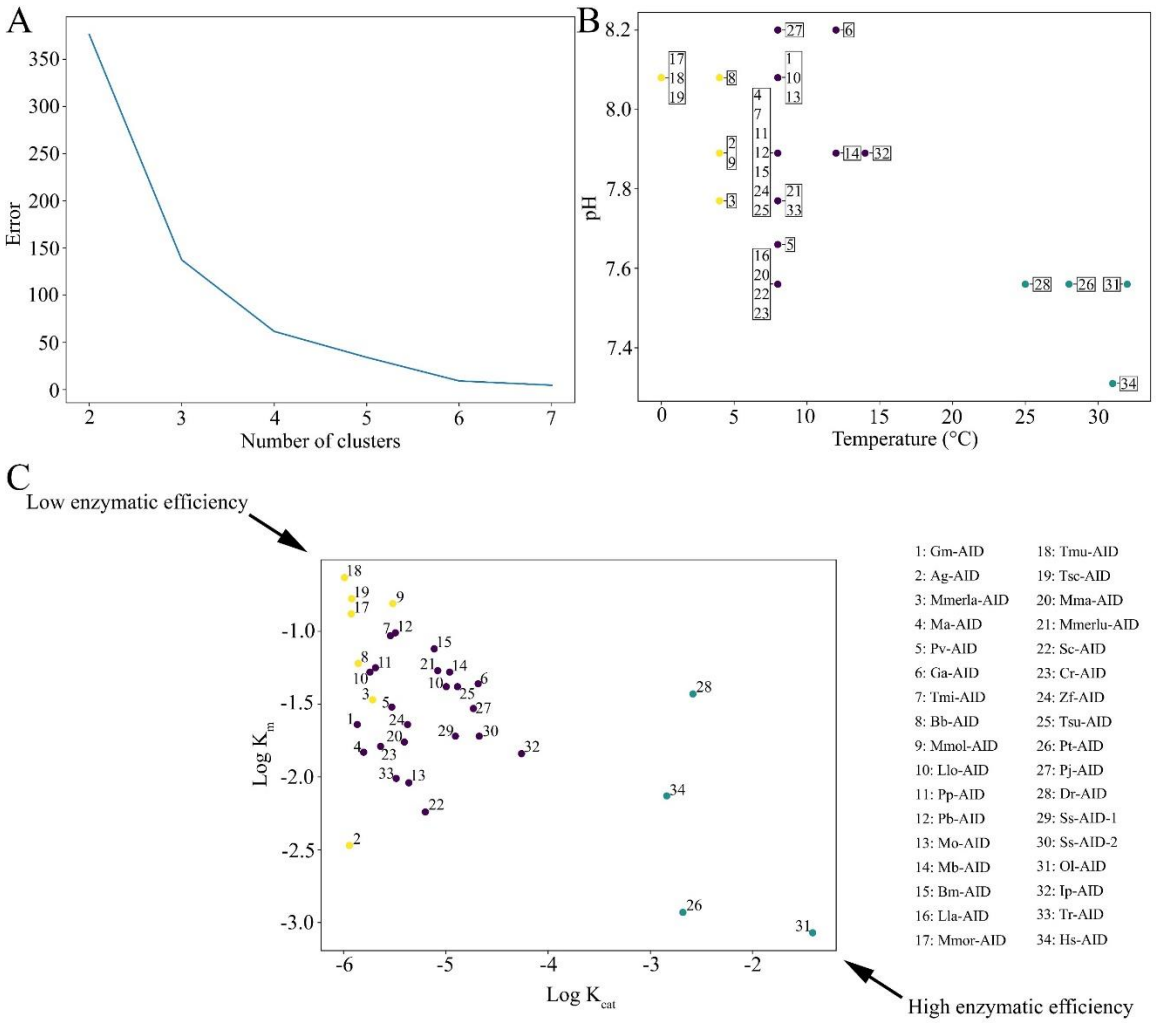


Figure 4-13: Clustering of extant AIDs based on their optimal temperature and optimal pH using machine learning algorithm of K-means clustering. A) The optimal number of clusters was estimated as three according to the K-means clustering model and elbow method. The K-means error was calculated for a given number of clusters (n = 2 to 8). On the error vs. number of clusters graph, the number of clusters where the "elbow" is bent was considered as the optimal number of clusters for the dataset. B) The dataset was divided into three distinct clusters based on the optimal temperature and pH using K-means clustering model. The three clusters are colored cyan, violet, and yellow. C) The catalytic efficiency of the extant AID proteins was compared amongst the three clusters obtained based on the optimal temperature and pH. The color scheme in B and C sections are the same. For abbreviations, refer to Table 4-1.

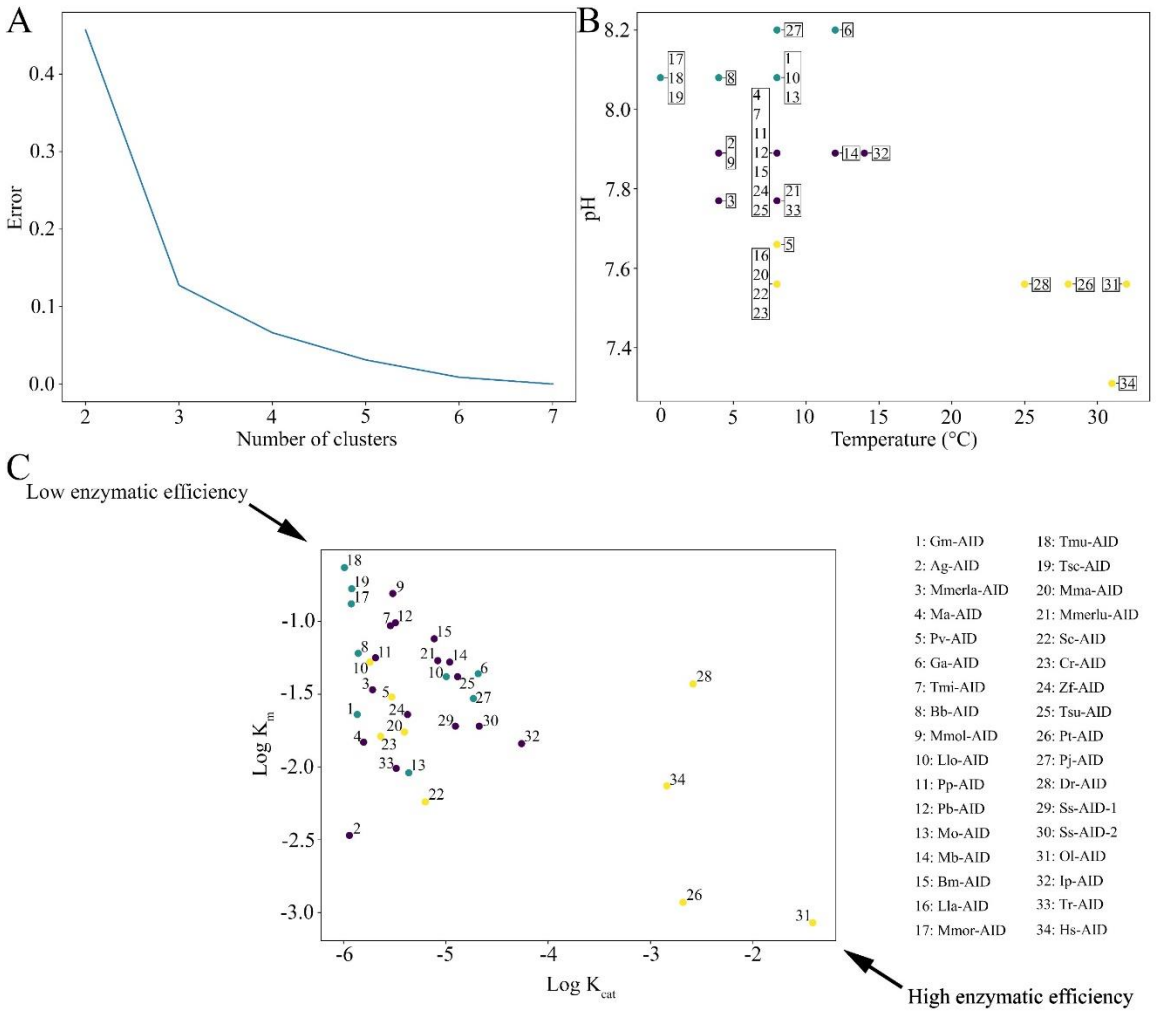


Figure 4-14: Clustering of extant AIDs based on their optimal pH using machine learning algorithm of K-means clustering. A) The optimal number of clusters was estimated as three according to the K-means clustering model and elbow method. The K-means error was calculated for a given number of clusters (n = 2 to 8). On the error vs. number of clusters graph, the number of clusters where the "elbow" is bent was considered as the optimal number of clusters for the dataset. B) The dataset was divided into three distinct clusters based on the optimal pH using K-means clustering model. The three clusters are colored cyan, violet, and yellow. C) The catalytic efficiency of the extant AID proteins was compared amongst the three clusters obtained based on the optimal pH. The color scheme in B and C sections are the same. For abbreviations, refer to Table 4-1.

4.4.2 Co-evolution of Gadidae *Ig* genes with their nearly inactivated AID

As mentioned in sections 1.5.3 and 3.4.5, previous studies have revealed a co-evolution between AID substrate specificity and the *Ig* variable (V) genes of vertebrate species (Conticello et al., 2005; Detanico et al., 2016; Golub & Charlemagne, 1998; Jolly et al., 1996; Oreste & Coscia, 2002; Wagner et al., 1995). Since we observed that the functional impairment of AID is a common phenomenon amongst Gadidae species, we sought to analyze their *Ig* gene sequences. The WRC and WGCW motif analyses of other Gadidae species revealed low/no AID hotspot enrichment in CDRs of Gadidae species (Figure 4-15 and Table 4-5) despite comparable abundance of WRC in their entire IgV_H fragments and higher GC content of their CDSs (Table 4-6).

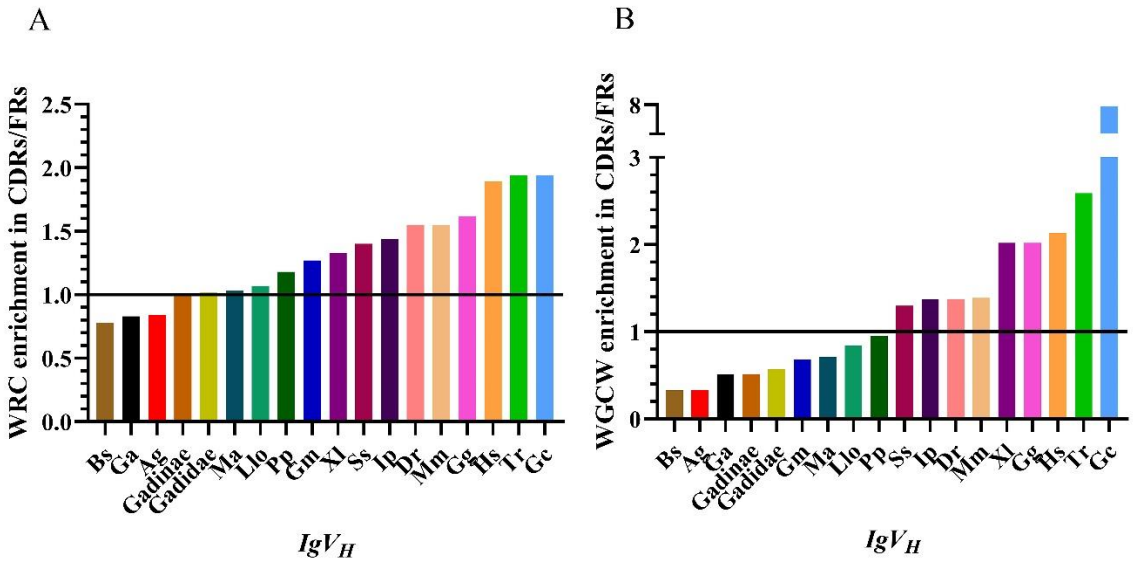


Figure 4-15: Co-evolution of AID activity with IgV_H gene sequences in Gadidae species. To assess the co-evolution of AID activity with IgV_H sequences in Gadidae species, enrichment of A) WRC motifs (AID hotspots on both strands) and B) WGCW motifs (overlapping AID hotspots on two strands) in CDRs of Gadidae species were compared to that of several other vertebrate species. Abbreviations: Bs: Polar cod; Ga: Silvery pout; Ag: Arctic cod; Ma: Haddock; Llo: Burbot; Pp: Forkbeard; Gm: Atlantic cod; Dr: zebrafish; Ss: Atlantic salmon; Ip: channel catfish; Tr: Japanese puffer fish; Gc: nurse shark; Xl: South African clawed toad; Gg: chicken; Mm: mouse; Hs: human.

Table 4-5: AID hotspot enrichment in IgV_H genes of various Gadidae and vertebrate species

	FR1			CDR1			FR2			CDR2			FR3					
	# AID hotspots	# nt. analyzed	index	# AID hotspots	# nt. analyzed	index	# AID hotspots	# nt. analyzed	index	# AID hotspots	# nt. analyzed	index	# AID hotspots	# nt. analyzed	index	Ave. FRs	Ave. CDRs	CDRs/FRs
Ag-IgV_H	879	5905	0.15	224	1328	0.17	168	1863	0.09	114	1968	0.06	1186	7224	0.16	0.13	0.11	0.84
Bs-IgV_H	405	2362	0.17	84	465	0.18	77	660	0.12	38	672	0.06	375	2201	0.17	0.15	0.12	0.78
Ma-IgV_H	167	882	0.19	35	156	0.22	23	231	0.10	21	240	0.09	128	775	0.17	0.15	0.16	1.03
Ga-IgV_H	226	1603	0.14	74	422	0.18	63	645	0.10	36	729	0.05	407	2407	0.17	0.14	0.11	0.83
Llo-IgV_H	345	1935	0.18	76	357	0.21	37	495	0.07	44	528	0.08	253	1566	0.16	0.14	0.15	1.07
Pp-IgV_H	377	2207	0.17	107	450	0.24	40	623	0.06	56	670	0.08	356	2042	0.17	0.14	0.16	1.18
Gadinae	2467	18589	0.13	852	4744	0.18	694	9009	0.08	383	5598	0.07	4051	25372	0.16	0.12	0.12	1.01
Gadidae	3189	22731	0.14	1035	5551	0.19	771	10127	0.08	483	6796	0.07	4660	28980	0.16	0.13	0.13	1.02
Gm-IgV_H	790	7837	0.10	435	2373	0.18	363	5610	0.06	174	1989	0.09	1955	12765	0.15	0.11	0.14	1.27
Ip-IgV_H	652	7199	0.09	482	2709	0.18	662	5498	0.12	379	2360	0.16	1753	12381	0.14	0.12	0.17	1.44
Tr-IgV_H	309	3675	0.08	219	1245	0.18	183	2361	0.08	268	1215	0.22	803	5517	0.15	0.10	0.20	1.94
Dr-IgV_H	410	5234	0.08	307	1786	0.17	396	3774	0.10	220	1510	0.15	1127	9143	0.12	0.10	0.16	1.55
Ss-IgV_H	2509	28445	0.09	1571	9215	0.17	2201	19629	0.11	1196	8333	0.14	6042	44363	0.14	0.11	0.16	1.40
Gc-IgV_H	727	7407	0.10	578	3102	0.19	664	6579	0.10	569	3027	0.19	1284	14250	0.09	0.10	0.19	1.94
Xl-IgV_H	88	902	0.10	50	292	0.17	67	611	0.11	33	252	0.13	192	1449	0.13	0.11	0.15	1.33
Gg-IgV_H	1218	15455	0.08	995	5010	0.20	1391	10627	0.13	1011	5031	0.20	3903	24359	0.16	0.12	0.20	1.62
Mm-IgV_H	3112	20493	0.15	1054	4209	0.25	689	11341	0.06	1730	13394	0.13	3907	25318	0.15	0.12	0.19	1.55
Hs-IgV_H	3322	27855	0.12	1452	5900	0.25	932	15503	0.06	2424	19590	0.12	4328	38075	0.11	0.10	0.18	1.89

Abbreviations: Ag: Arctic cod; Bs: Polar cod; Ma: Haddock; Ga: Silvery pout; Llo: Burbot; Pp: Forkbeard; Gm: Atlantic cod; Dr: zebrafish; Ss: Atlantic salmon; Ip: channel catfish; Tr: Japanese puffer fish; Gc: nurse shark; Xl: South African clawed toad; Gg: chicken; Mm: mouse; Hs: human.

Table 4-6: AID hotspot enrichment in the entire IgV_H genes and GC content of annotated complete protein coding genes (CDSs) of various Gadidae and vertebrate species

	IgV_H gene analysis				Genomic analysis	
	# AID hotspot	# nt. analyzed	AID hotspots/nt. analyzed	# transcripts	# CDSs	GC%
Ag-IgV_H	2571	18288	0.1406	87	8	60.35
Bs-IgV_H	979	6360	0.1539	20	73	61.66
Ma-IgV_H	374	2284	0.1637	7	44	54.80
Ga-IgV_H	806	5806	0.1388	25	6	60.87
Llo-IgV_H	755	4881	0.1547	15	32	59.19
Pp-IgV_H	936	5992	0.1562	19	9	60.23
Gadinae	8447	63312	0.1334	251	NA	59.44
Gadidae	10138	74185	0.1367	285	NA	59.52
Gm-IgV_H	3717	30574	0.1216	112	44330	59.53
Ip-IgV_H	3928	30147	0.1303	109	47956	51.46
Tr-IgV_H	1782	14013	0.1272	49	46294	54.11
Dr-IgV_H	2460	21447	0.1147	76	57060	49.85
Ss-IgV_H	13519	109985	0.1229	405	97576	55.12
Gc-IgV_H	3822	34365	0.1112	129	1507	47.97
Xl-IgV_H	430	3506	0.1226	44	49356	45.62
Gg-IgV_H	8518	60482	0.1408	239	56680	50.23
Mm-IgV_H	10492	74755	0.1404	420	88579	51.96
Hs-IgV_H	12458	106923	0.1165	727	120426	51.02

NA: Since the number of analyzed IgV_H transcripts was very low for Gadidae species other than Atlantic cod, we decided not to report this parameter for Gadidae and Gadinae groups.
Abbreviations: Ag: Arctic cod; Bs: Polar cod; Ma: Haddock; Ga: Silvery pout; Llo: Burbot; Pp: Forkbeard; Gm: Atlantic cod; Dr: zebrafish; Ss: Atlantic salmon; Ip: channel catfish; Tr: Japanese puffer fish; Gc: nurse shark; Xl: South African clawed toad; Gg: chicken; Mm: mouse; Hs: human.

4.4.3 Resurrecting Gadiformes ancestral AIDs

To estimate the evolutionary point at which the functional alteration of AID begun within Gadiformes lineage and to infer the evolutionary trajectory of the functional alteration, we performed ASR analyses. The prediction of ancestral sequences requires four steps that are followed by the fifth step of resurrecting the ancestral proteins in the lab (R. Merkl & R. Sterner, 2016). In the first step, homologous extant sequences are retrieved from various database. Then, a multiple sequence alignment (MSA) is created based on which a phylogeny tree would be constructed in the next step. In the last step of ASR analysis, ancestral sequences are predicted using the MSA and the phylogeny tree. A critical step in ASR is examining the biochemical and functional properties of the predicted ancestors in the lab (R. Merkl & R. Sterner, 2016; Rainer Merkl & Reinhard Sterner, 2016).

4.4.3.1 Selected extant species for ancestral sequence reconstruction analyses

We included the AID gene sequence form 73 bony fish species (Appendix 6 and Appendix 7) and used *Lampetra tridentata* CDA1 as the outgroup. The amino acid alignment was guided by the predicted 3D structure of Gm-AID (Appendix 8). Interestingly, we could not find a complete or partial *aicda* gene in the striped codlet (*Bregmaceros cantori*). The genomic sequencing also revealed the lack of many other important immune genes in this species (Malmstrom et al., 2016). In our dataset, the striped codlet represents the most basal Gadiformes species and is characterized by the complete absence of *mhc I U*, *mhc II*, *cd4*, *cd8* and *aicda* genes (Malmstrom et al., 2016).

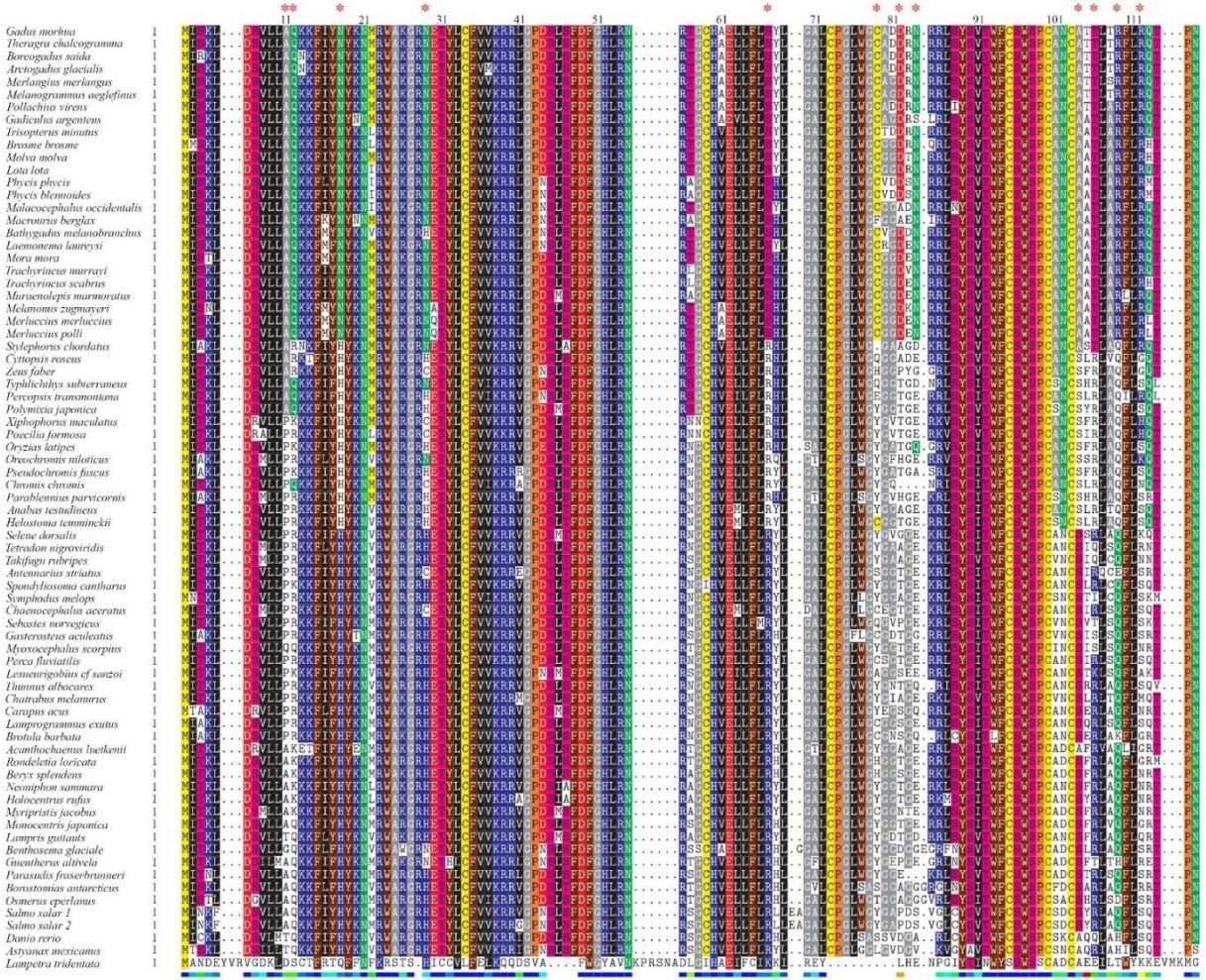
Previous studies have shown that factors such as the alignment algorithm, assumptions, and the rate of insertions and deletions impact the ASR results (R. Merkl &

R. Sterner, 2016; Vialle et al., 2018). With the goal of creating a more accurate MSA, we decided to guide the alignment algorithm with the 3D structure. Given the potential stronger conservation of structure *vs*. sequence in protein evolution, previous studies have concluded that the structure-guided alignments can outperform sequence-alignments (Ingles-Prieto et al., 2013; Kim & Lee, 2007). Here, we used our computationally predicted Gm-AID structure (Appendix 8) to guide MSA, which was manually inspected to verify the accuracy of the alignment such as the presence and the boundaries of the gaps (Figure 4-16).

We also noticed interesting amino acid differences mainly in the Gadiformes group compared to other bony fish species (Figure 4-17). The Aconthomorphata class has a conserved alanine (A) in position 11, with the exception of *L. guttatus* which has a tyrosine (T); however, the entire Percomorphaceae group has a proline (P) in this position, with the exception of *M. scorpius* which has a glutamine (Q). In position 12, majority of Euacanthomorphacea AIDs contain a positively charged amino acid (*i.e.*, R or K) while the rest of AIDs including Gadiformes mainly have a glutamine (Q) at this position. All Gadiformes species have an asparagine (N) amino acid at position 18 while the rest of extant species studied here contain a histidine (H). At position 29, we noticed that the entire Gadariae group, except for *B. melanobranchus*, contain asparagine (N) while the majority of other AIDs contain histidine (H) and some have cysteine (C) or asparagine (N). The position 67 in all Gadiformes AIDs is occupied with a serine (S) while the rest of extant species have the positively charged arginine (R). *M. berglax* and all non-Gadiformes AIDs contain an aromatic amino acid (phenylalanine [F] and tyrosine [T], respectively) at

position 79 while this residue has changed into cysteine (C) in the rest of Gadiformes species. In Gadiformes AIDs, position 82 has a mostly conserved aspartic acid (D), except for *M. berglax* and *M. occidentalis* which contain an alanine (A). This position is mostly occupied with an uncharged amino acid apart from *L. guttauts*, *D. rerio*, and *A. mexicanus* which also have an aspartic acid (D). At position 84, all Gadiformes species contain asparagine (N) apart from *G. argenteus* which has a serine (S). This position is mostly occupied with a negative residue (*i.e.*, E/D) in Acanthopterygii AIDs. While Gadariae AIDs have a conserved alanine (A) at 104 position, all other bony fish AIDs in this report have a serine (S) except for *A. luetkenii*, *D. rerio*, and *A. mexicanus* which also contain an alanine (A). At position 106, all Gadariae species have tyrosine (T) whilst most of other species have the positively charged arginine (R). The conserved amino acid at position 109 in the entire Gadiformes group is arginine (R), while this position is occupied mostly with glutamine (Q) and to lesser extent with glutamic acid (E), aspartic acid (D), lysine (K), and histidine (H) in all other extant species. At position 112, while the entire Gadariae group have an arginine (R), the rest of the species contain serine (S), glycine (G), arginine (R), histidine (H), lysine (K), asparagine (N), alanine (A), or glutamine (Q). At position 135, the whole Gadariae species show proline (P) while the rest of the extant AIDs mostly have an arginine (R). Position 136 is occupied by a histidine (H) in the entire Gadiformes group, and the rest of extant AID proteins mostly contain glutamic acid (E) with a few showing aspartic acid (D) or alanine (A). Position 151 in the entire Gadidae AIDs is occupied with a lysine (K) while the rest of extant AIDs contain tyrosine (T) with a few showing isoleucine (I), asparagine (N), or serine (S) at this position. While leucine (L) is the

conserved amino acid at position 170 in all Gadiformes AIDs, the rest of extant AIDs studied here contain glutamine (Q), lysine (K), tyrosine (T), alanine (A), leucine (L), or asparagine (N). While the most of non-Gadiformes AIDs contain glutamine (Q) in their 181 position, tyrosine (T) is conserved at this position in the entire Gadiformes group. Position 209 is occupied with phenylalanine (F), isoleucine (I), and leucine (L) in Gadidae, the rest of Gadiformes, and non-Gadiformes AIDs, respectively. All Zeiogadaria AIDs, except for *S. chordatus*, have an extra leucine (L) in position 212. Additionally, the entire Gadiformes group contain an extra tyrosine (T) or serine (S) at the C-terminus, making Gadiformes AIDs the longest AIDs studied thus far. Taken together, it seems that Gadiformes AIDs contain lineage-specific amino acid changes compared to the rest of our dataset. However, understanding the functional ramification of these lineage-specific amino acid replacements require further studies.


Gadus morhua
Theragra chalcogramma
Boreogadus saida
Arctogadus glacialis
Merlangius merlangus
Melanogrammus aeglefinus
Pollachius virens
Gadiculus argenteus
Trisopterus minutus
Brosme brosme
Molva molva
Lota lota
Phycis phycis
Phycis blennoides
Malacocephalus occidentalis
Macrourus berglax
Bathygadus melanobranchus
Laemonema laureysi
Mora mora
Trachyrincus murrayi
Trachyrincus scabrus
Muraenolepis marmoratus
Melamonus zugmayeri
Merluccius merluccius
Merluccius polli
Stylephorus chordatus
Cyttopsis roseus
Zeus faber
Typhlichthys subterraneus
Percopsis transmontana
Polymixia japonica
Xiphophorus maculatus
Poecilia formosa
Oryzias latipes
Oreochromis niloticus
Pseudochromis fuscus
Chromis chromis
Parablennius parvicornis
Anabas testudineus
Helostoma temminckii
Selene dorsalis
Tetraodon nigroviridis
Takifugu rubripes
Antennarius striatus
Spondyliosoma cantharus
Symphodus melops
Chaenocephalus aceratus
Sebastes norvegicus
Gasterosteus aculeatus
Myoxocephalus scorpius
Perca fluviatilis
Lesueurigobius cf sanzoi
Thunnus albacares
Chaetrabus melamarus
Carapus acus
Lamprogrammus exutus
Bratula barbata
Acanthochaenus luetkenii
Rondeletia loricata
Beryx splendens
Neoniphon sammara
Holocentrus rufus
Myripristis jacobus
Monocentris japonica
Lampris guttatus
Benthosema glaciale
Gnantherus altivela
Paraxadis fraserbrunneri
Borostomias antarcticus
Osmerus eperlanus
Salmo salar 1
Salmo salar 2
Danio rerio
Astyanax mexicanus
Lampetra tridentata
Gadidae sister group
Gadidae
Gadiformes
Gadariae
Zeiogadaria
Paracanthopterygii
Acanthomorpha
Ctenosquamata
Neoteleostei
Percomorphaceae
Euacanthomorphacea
Acanthopterygii

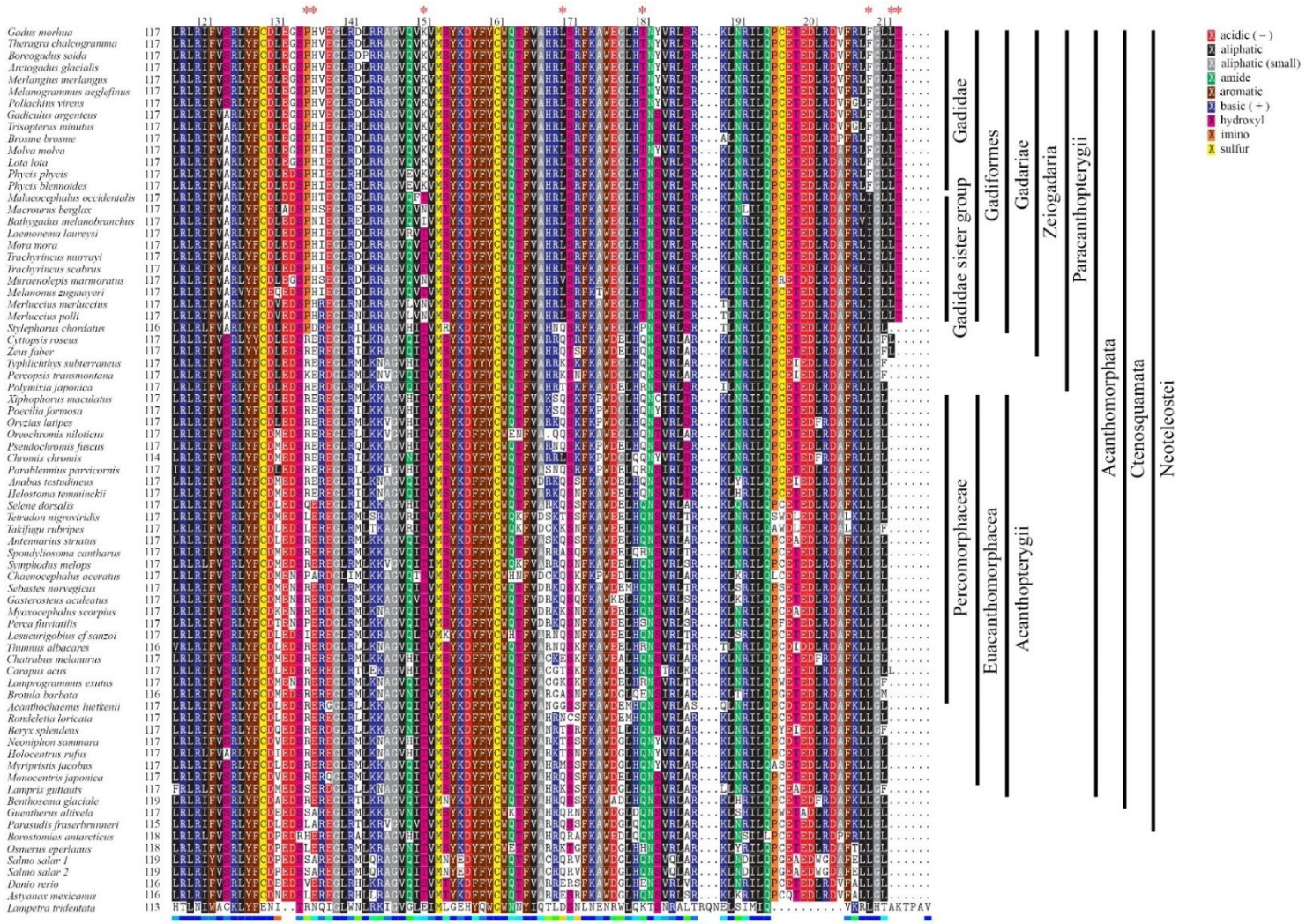


Figure 4-16: Amino acid alignment of extant genes used for ASR analyses. Positions with significant amino acid conservation within or outside the Gadiformes group is labeled with red star. For detailed explanation refer to the text.

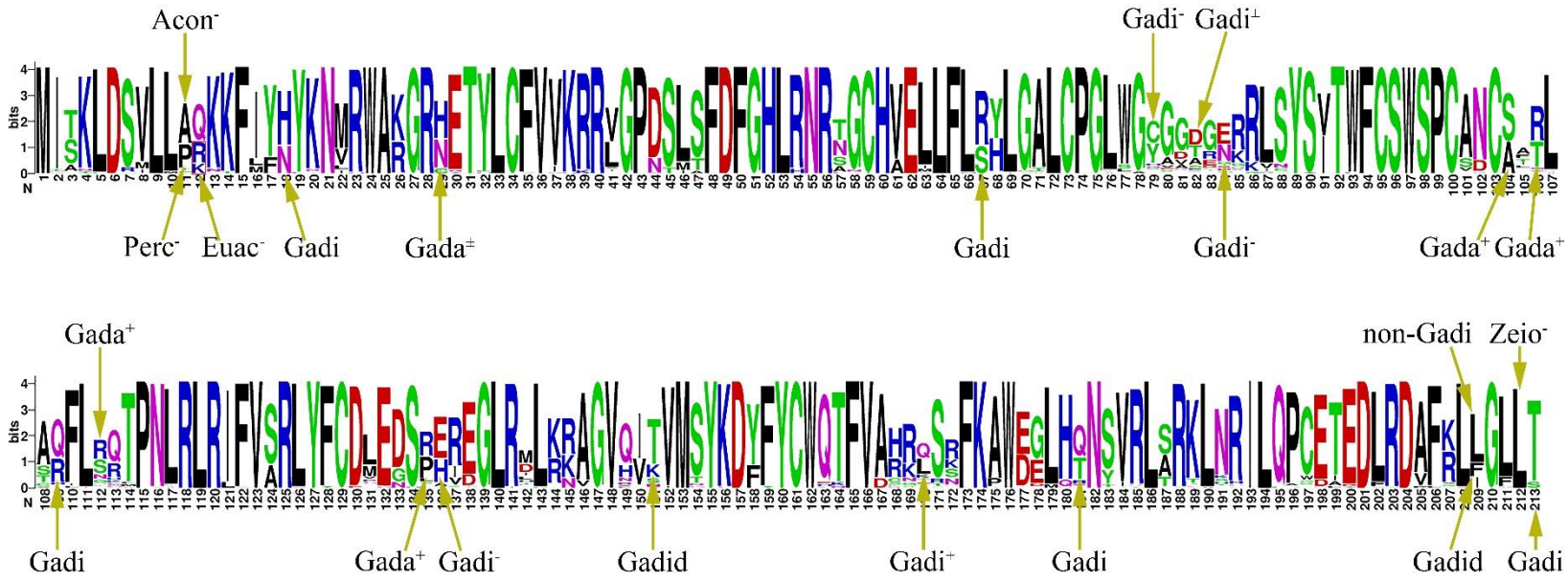


Figure 4-17: Amino acid conservation of extant AID homologs used in ASR analyses. Amino acid positions where a distinctive difference between various groups was observed are labeled. + sign emphasizes that a few members of other groups also contain the labeled amino acid at that position. – sign means that a few of the specified group are exceptions. ± sign indicates that while a few members of the specified group are exception, a few members of other groups show the specified amino acid at the labeled position. Abbreviation: Acon: Acanthomorphata; Euac: Euacanthomorphacea; Perc: Percomorphaceae; Zeio: Zeiogadaria; Gada: Gadariae; Gadi: Gadiformes; non-Gadi: non-Gadiformes; Gadid: Gadidae.

4.4.3.2 Gene tree *vs*. species tree

The best ML gene tree calculated based on the AID extant sequences is illustrated in Figure 4-18. The previously published species tree was constructed based on the genomic sequence of these species (Figure 4-19) (Malmstrom et al., 2016). Previous studies revealed that the phylogenetic uncertainty and inaccuracy could impact the ASR results (Duchêne & Lanfear, 2015; Groussin et al., 2015; R. Merkl & R. Sterner, 2016; Vialle et al., 2018). Specifically, the phylogenetic uncertainty could lead to the overestimation of the evolutionary transitions in the large datasets (Duchêne & Lanfear, 2015). In general, using a single tree to infer ancestral sequences assumes that the single tree demonstrates the true or close-to-true phylogenetic relationships amongst extant species (Joy et al., 2016; Pagel et al., 2004). Both ProtASR and RAxML accept a single input phylogenetic tree which would be used to deduce the phylogenetic relationships amongst the extant species (Arenas & Bastolla, 2019; Arenas et al., 2017; Stamatakis, 2014). We decided to use the previously-published species tree for ASR calculation using ProtASR and RAxML packages for two reasons. First, the previously published species tree has higher bootstrapping value and confidence compared to the best ML tree constructed using our *aicda* gene sequences (Malmstrom et al., 2016). Second, employing species-aware gene tree has been shown to improve the ASR results (Groussin et al., 2015).

Of the three ASR methods used here, the Bayesian inference seems to integrate the uncertainty concerning the tree topology and the evolutionary model parameters more adequately (Ronquist & Huelsenbeck, 2003; Ronquist et al., 2012). To predict the ancestral state of a given node, the MrBayes package can use a user-defined tree as the starting point

and combine the uncertainty regarding the tree topology of other nodes (*i.e.*, excluding the node for which the ancestral sequence is being calculated) and all other evolutionary parameters (Ronquist & Huelsenbeck, 2003; Ronquist et al., 2012). It is worth mentioning that to study the trait evolution, another Bayesian Markov chain Monte Carlo (MCMC) technique was developed where the uncertainty of the tree topology of both the given node and other ancestral nodes was taken into account (Pagel et al., 2004). This method sampled both better and worse trees to calculate the uncertainty about the existence of the ancestral node under study. Then, the estimated uncertainty was used to limit the confidence of the predicted ancestral state resulting in more realistic probability estimation (Pagel et al., 2004). However, this method was only used on a small dataset. Due to the size of our dataset, the availability of a high-confidence species tree for our dataset (which was used as the start tree in the MrBayes calculations), and the fact that the ancestral state in MrBayes package is calculated while integrating the uncertainty in all other parameters, including the topology of other parts of the tree, we decided to apply the ASR method implemented in MrBayes package. In other words, since the existence of the ancestral nodes studied in this thesis have been confirmed with genomic sequences and the fossil constraints in a previous study (Malmstrom et al., 2016), it is reasonable to assume that adding computationally intensive analyses to account for the uncertainty in the existence of these ancestral nodes was unnecessary.

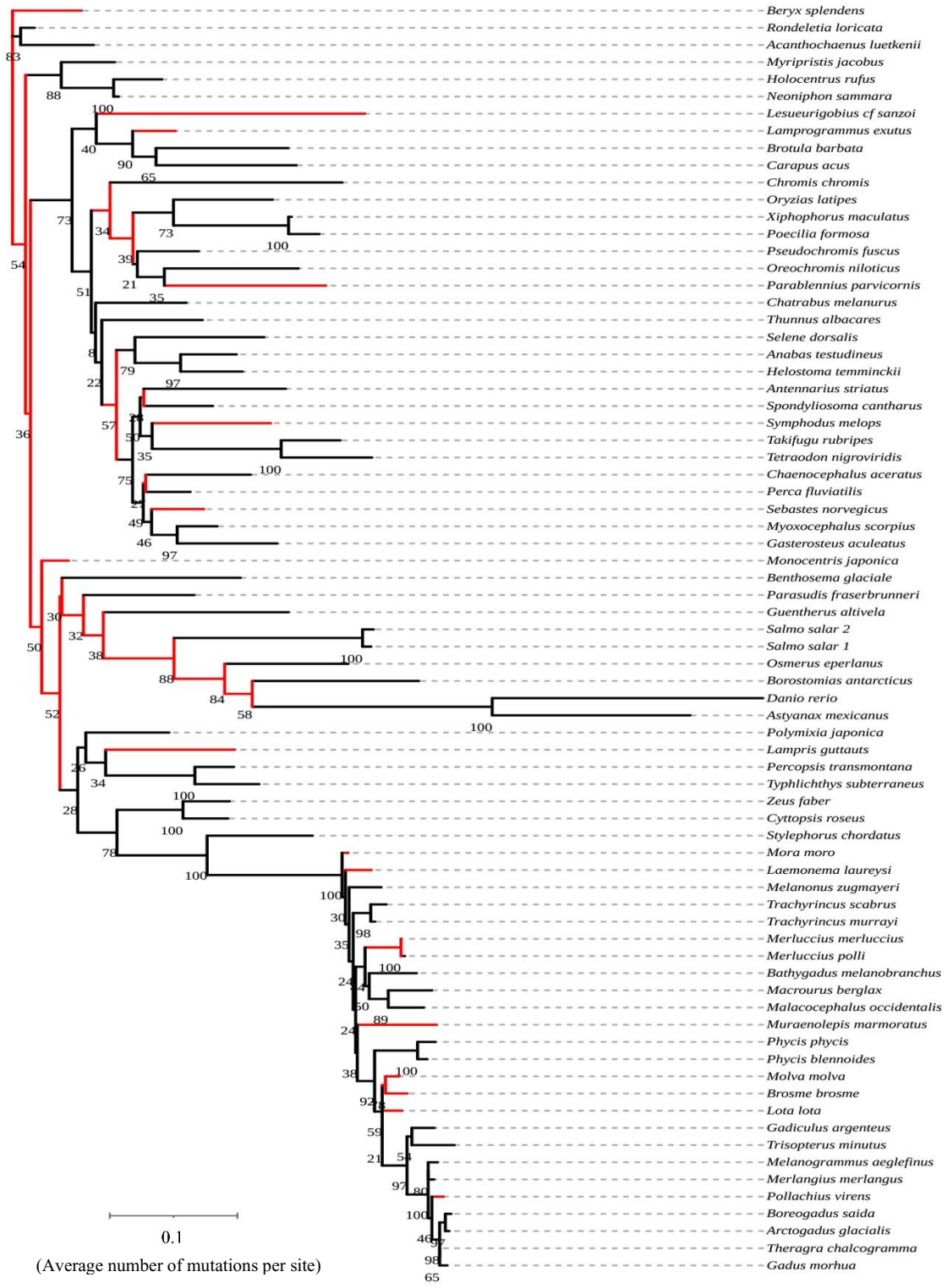


Figure 4-18: The best ML tree obtained in this thesis. The numbers represent the bootstrapping values. The major differences between the gene tree and the species tree are highlighted in red.

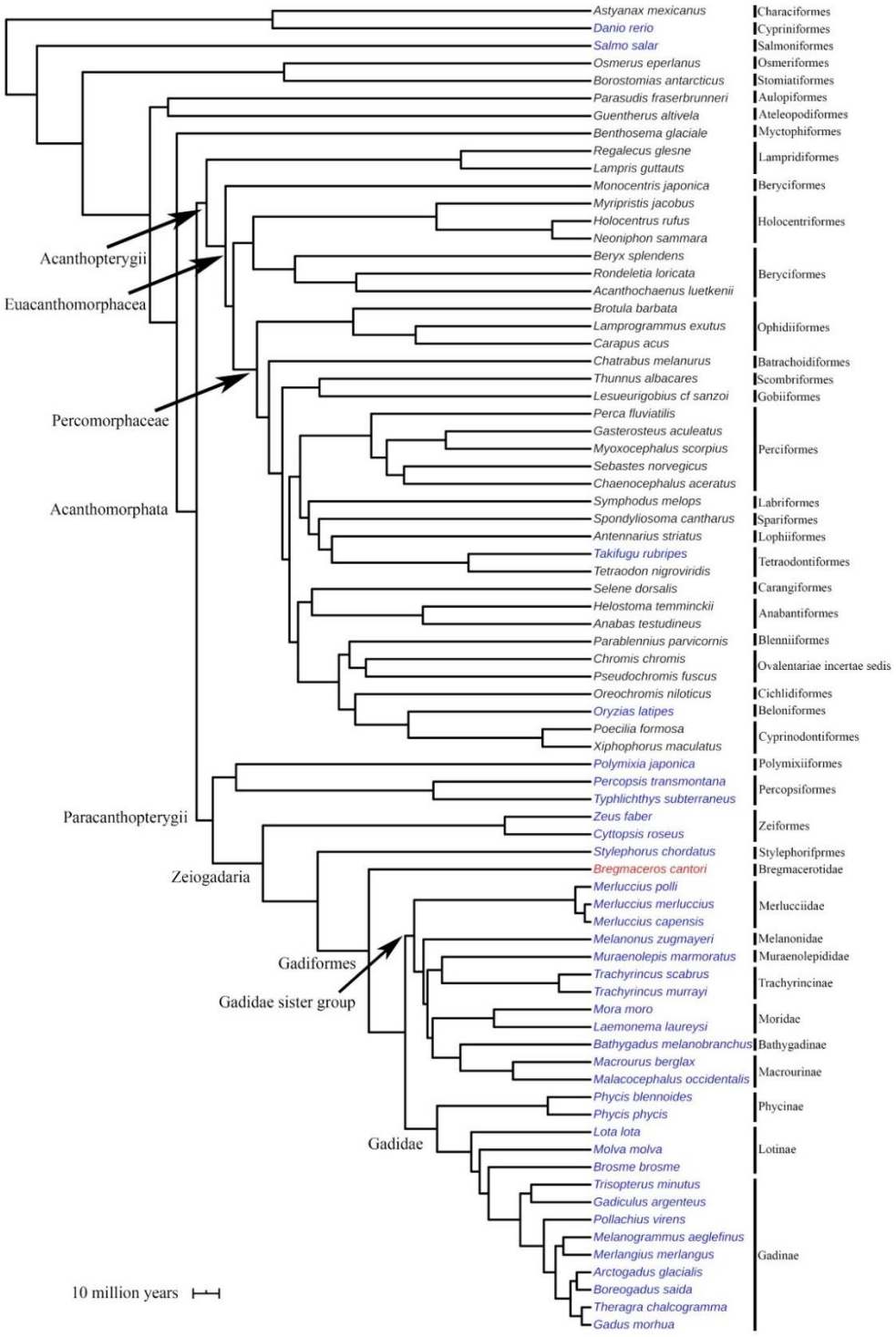


Figure 4-19: Previously published (Malmstrom et al., 2016) species tree used in this thesis. AID proteins from species colored blue were synthesized in the lab to study their biochemical properties. Channel catfish and human AIDs were also purified and tested. We could not find any aicda gene in the genomic sequence of B. cantori (colored in red).

4.4.3.3 Predicting ancestral AID sequences

Currently, the two methods of maximum likelihood (ML) and Bayesian inference are the most popular algorithms used to calculate ancestral genes (R. Merkl & R. Sterner, 2016). Amongst different ASR algorithms, the Bayesian methods incorporating rate variation model, seem to provide the most accurate results (Joy et al., 2016; Randall et al., 2016). We predicted the ancestral *aicda* gene sequence of Gadidae (Gd-ANC), its sister group (Gds-ANC), Gadiformes (Gf-ANC), and Zeiogadaria (Zg-ANC) using three different software packages: MrBayes, RAxML, and ProtASR.

MrBayes applies the Bayesian method to infer ancestral gene sequences from the extant protein alignment (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003; Ronquist et al., 2012). The RAxML package was used to predict ancestral genes using the ML algorithm (marginal ML) and protein alignment (Stamatakis, 2014). ProtASR is an ML-based package that takes advantage of a structurally constrained substitution model called "Mean-field" (Arenas & Bastolla, 2019; Arenas et al., 2017). Mean-field substitution model considers the unfolding and misfolding states of the protein under study which can outperform the empirical substitution models for data with larger sequence divergence (Arenas et al., 2015). ProtASR utilizes both marginal and joint maximum likelihood algorithms to predict ancestral sequences (Arenas & Bastolla, 2019; Arenas et al., 2017). In the marginal ML algorithm, the ancestral sequence is assigned while taking into account only the immediate descendants of a given node (Joy et al., 2016). In contrast, the joint ML method attempts to assign the ancestral state at each given node by maximizing the likelihood of the data throughout the entire tree (Joy et al., 2016). Therefore, it is more

likely to find global optima using the joint ML method (Joy et al., 2016). Additionally, ProtASR calculates the statistical probabilities at both global and local levels (Arenas & Bastolla, 2019; Arenas et al., 2017). Since the ancestral inference at the global level assumes that all sites evolve under a same evolutionary process, we decided to only consider the ASR results inferred using joint ML at the local level (*i.e.*, considering heterogeneous evolutionary processes across sites).

Figure 4-20 and Table 4-7, 4-8, 4-9, and 4-10 illustrate the predicted ancestral sequences obtained from each method. The Gadidae sister group was not formed as a monophyletic group in our gene tree constructed based on the nucleotide sequence of the extant *aicda* genes. Only a monophyletic group shares a common ancestor. Therefore, RAxML package was not able to infer the Gds-ANC when our *aicda* gene tree was used.

Amongst the applied ASR methods in this thesis, previous studies have shown that ASR results obtained from Bayesian inference, especially the hierarchical Bayes approach (*e.g.*, implemented in MrBayes package), outperform the results of other methods (Joy et al., 2016; Randall et al., 2016). Therefore, the predicted ancestral sequences were compared, and the consensus protein sequence for each ancestral node was predicted with a higher emphasis on MrBayes results. Figure 4-21 Shows the protein alignment of the predicted ancestral sequence. Variants of the ancestral AIDs were also generated if an amino acid position was predicted ambiguously (*i.e.*, positions with a statistical uncertainty of 0.2 or higher) (Eick et al., 2017).

The predicted Gd-ANC and Gds-ANC differ in 4 amino acid positions, which are occupied with amino acids that are substantially different regarding their biochemical

properties. The positions 17, 83, 151, and 209 are predicted to be occupied with an isoleucine (nonpolar aliphatic), an arginine (positively charged), a lysine (positively charged), and a phenylalanine (nonpolar aromatic) in Gd-ANC and a tyrosine (polar aromatic), a glutamic acid (negatively charged), a threonine/asparagine (polar aliphatic), and an isoleucine (nonpolar aliphatic) in Gds-ANC protein. Interestingly, amongst the predicted ancestral AIDs, Gds-ANC and Gf-ANC only differ in one amino acid position (I16M in Gds-ANC *vs*. Gf-ANC). In fact, position 16 in Gds-ANC was predicted with an ambiguity between I and M (Table 4-7). Zg-ANC was the most diverge ancestral AID compared with Gm-AID amino acid sequence and was predicted with more ambiguous amino acid sites compared to other predicted ancestral AIDs. Considering all the ASR methods, 4, 2, 5, and 22 sites showed uncertainty level of 0.2 or higher in Gd-ANC, Gds-ANC, Gf-ANC, and Zg-ANC, respectively (Table 4-7 through Table 4-10).

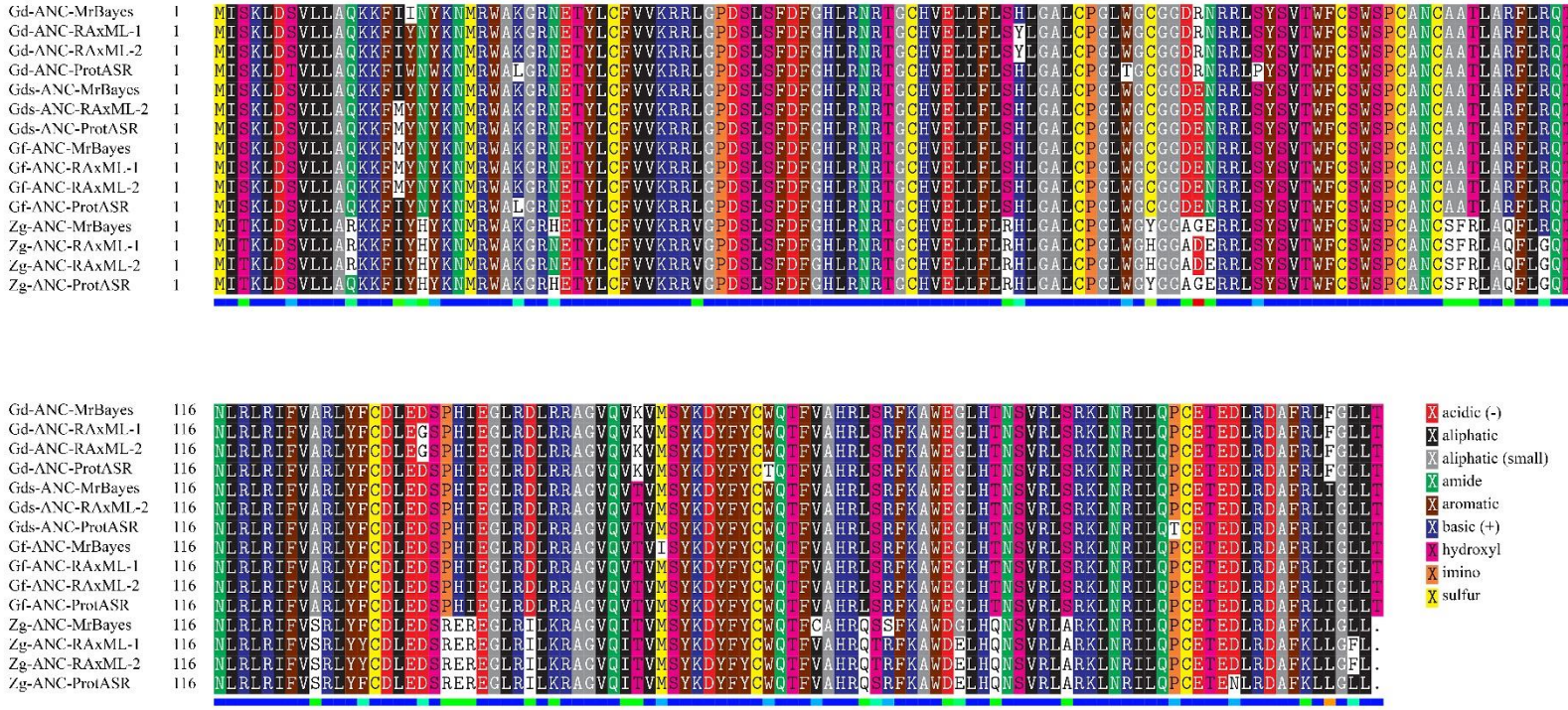


Figure 4-20: Amino acid alignment of the predicted ancestral AIDs using four different methods. Only amino acids with highest probability are shown. Predicted ancestors using MarBayes, RAxML, or ProtASR packages are labelled accordingly. In case of RAxML package, predicted ancestors using the AID gene tree or previously published species tree are labeled as 1 and 2, respectively. Amino acids are colored based on their chemical properties as indicated in the bottom right corner legends. Abbreviations: Gd-ANC: Gadidae ancestor; Gds-ANC: Gadidae sister group ancestor; Gf-ANC: Gadiformes ancestor; Zg-ANC: Zeiogadaria ancestor.

Table 4-7: Predicted ancestral sequences using MrBayes package and the species tree as the starting tree

Ancestral node	Predicted amino acid (aa) sequence	Length (aa)	Positions with < 0.8 certainty
Gd-ANC	MISKLDSVLLAQKKFIINYKNMRWAKGRNETYLCFVVKRRLGPD SLSFDFGHLRNRTGCHVELLFLSHLGALCPGLWGCGGDRNRRLS YSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLED SPHIEGLRDLRRAGVQVKVMSYKDYFYCWQTFVAHRLSRFKAW EGLHTNSVRLSRKLNRILQPCETEDLRDAFRLFGLLT.	213	133: D (71%) G (29%)
Gds-ANC	MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGP DSLSFDFGHLRNRTGCHVELLFLSHLGALCPGLWGCGGDENRRL SYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLE DSPHIEGLRDLRRAGVQVTVMSYKDYFYCWQTFVAHRLSRFKA WEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGLLT.	213	16: I (63%) M (37%) 151: T (79%) N (21%)
Gf-ANC	MISKLDSVLLAQKKFMYNYKNMRWAKGRNETYLCFVVKRRLGP DSLSFDFGHLRNRTGCHVELLFLSHLGALCPGLWGCGGDENRRL SYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLE DSPHIEGLRDLRRAGVQVTVISYKDYFYCWQTFVAHRLSRFKAW EGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGLLT.	213	
Zg-ANC	MITKLDSVLLARKKFIYHYKNMRWAKGRHETYLCFVVKRRVGP DSLSFDFGHLRNRTGCHVELLFLRHLGALCPGLWGYGGAGERRL SYSVTWFCSWSPCANCSFRLAQFLRQTPNLRLRIFVSRLYFCDLE DSREREGLRILKRAGVQITVMSYKDYFYCWQTFCAHRQSSFKAW DGLHQNSVRLARKLNRILQPCETEDLRDAFKLLGLL.	212	12: R (53%) Q (46%) 83: G (73%) D (27%) 172: S (62%) R (34%)

Table 4-8: Predicted ancestral sequences using RAxML package and the aicda gene tree

Ancestral node	Predicted amino acid (aa) sequence	Length (aa)	Positions with < 0.8 certainty
Gd-ANC	MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGP DSLSFDFGHLRNRTGCHVELLFLSYLGALCPGLWGCGGDRNRRL SYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLE GSPHIEGLRDLRRAGVQVKVMSYKDYFYCWQTFVAHRLSRFKA WEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLFGLLT.	213	22: M (71%) I (29%)
Gds-ANC*	NA		
Gf-ANC	MISKLDSVLLAQKKFMYNYKNMRWAKGRNETYLCFVVKRRLGP DSLSFDFGHLRNRTGCHVELLFLSHLGALCPGLWGCGGDENRRL SYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLE DSPHIEGLRDLRRAGVQVTVMSYKDYFYCWQTFVAHRLSRFKA WEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGLLT.	213	4: K (79%) T (21%)
Zg-ANC	MITKLDSVLLARKKFIYHYKNMRWAKGRNETYLCFVVKRRVGP DSLSFDFGHLRNRTGCHVELLFLRHLGALCPGLWGHGGADERRL SYSVTWFCSWSPCANCSFRLAQFLGQTPNLRLRIFVSRLYYCDLE DSREREGLRILKRAGVQITVMSYKDYFYCWQTFVAHRQTRFKA WDELHQNSVRLARKLNRILQPCETEDLRDAFKLLGFL.	212	29: N (40%) H (29%) Y (26%) 79: H (42%) R (16%) Y (14%) Q (13%) 84: E (60%) D (36%) 104: S (55%) A (35%) 105: F (40%) S (21%) L (20%) P (11%) 106: R (55%) T (36%) 112: G (56%) R (42%) 124: S (58%) A (37%) 135: R (47%) P (29%) 136: E (64%) D (35%) 144: K (62%) R (38%) 171: T (56%) S (40%) 177: D (54%) E (32%) 178: E (63%) G (36%) 181: Q (61%) P (36%) 187: A (49%) S (38%) 211: F (52%) L (35%)

*: The extant species which belong to the Gadidae sister group did not form a monophyletic group in our aicda gene tree. Therefore, RAxML was unable to assign an ancestral state for this group.

Table 4-9: Predicted ancestral sequences using RAxML package and the previously published species tree

Ancestral node	Predicted amino acid (aa) sequence	Length (aa)	Positions with < 0.8 certainty
Gd-ANC	MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGP DSLSFDFGHLRNRTGCHVELLFLSYLGALCPGLWGCGGDRNRRL SYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLE GSPHIEGLRDLRRAGVQVKVMSYKDYFYCWQTFVAHRLSRFKA WEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLFGLLT.	213	113: Q (71%) K (14%) L (12%)
Gds-ANC	MISKLDSVLLAQKKFMYNYKNMRWAKGRNETYLCFVVKRRLGP DSLSFDFGHLRNRTGCHVELLFLSHLGALCPGLWGCGGDENRRL SYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLE DSPHIEGLRDLRRAGVQVTVMSYKDYFYCWQTFVAHRLSRFKA WEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGLLT.	213	
Gf-ANC	MISKLDSVLLAQKKFMYNYKNMRWAKGRNETYLCFVVKRRLGP DSLSFDFGHLRNRTGCHVELLFLSHLGALCPGLWGCGGDENRRL SYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLE DSPHIEGLRDLRRAGVQVTVMSYKDYFYCWQTFVAHRLSRFKA WEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGLLT.	213	16: M (75%) I (25%) 83: E (62%) K (19%) G (14%) 151: T (70%) N (21%) 209: I (79%) F (21%)
Zg-ANC	MITKLDSVLLARKKFIYHYKNMRWAKGRNETYLCFVVKRRVGP DSLSFDFGHLRNRTGCHVELLFLRHLGALCPGLWGHGGADERRL SYSVTWFCSWSPCANCSFRLAQFLGQTPNLRLRIFVSRLYYCDLE DSREREGLRILKRAGVQITVMSYKDYFYCWQTFVAHRQTRFKA WDELHQNSVRLARKLNRILQPCETEDLRDAFKLLGFL.	212	29: N (40%) H (30%) Y (27%) 79: H (38%) R (14%) Q (12%) Y (12%) 84: E (61%) D (36%) 103: S (60%) A (37%) 104: F (44%) S (22%) L (18%) 105: R (57%) T (35%) 112: G (59%) R (41%) 128: Y (60%) F (39%) 135: R (55%) P (31%) 136: E (55%) D (35%) 144: K (63%) R (36%) 171: T (58%) S (39%) 177: D (55%) E (33%) 178: E (64%) G (35%) 181: Q (62%) P (35%) 187: A (51%) S (29%) T (11%) 211: F (53%) L (33%)

Table 4-10: Predicted ancestral sequences using RrotASR package, our computationally predicted Gm-AID 3D structure, and the previously published species tree

Ancestral node	Predicted amino acid (aa) sequence (Joint ML)	Length (aa)	Positions with < 0.8 certainty*
Gd-ANC	MISKLDTVLLAQKKFIWNWKNMRWALGRNETYLCFVVKRRLGP DSLSFDFGHLRNRTGCHVELLFLSHLGALCPGLTGCGGDRNRRLP YSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLED SPHIEGLRDLRRAGVQVKVMSYKDYFYCTQTFVAHRLSRFKAWE GLHTNSVRLSRKLNRILQPCETEDLRDAFRLFGLLT.	213	83: R (29%) G (24%) E (23%) S (21%)
Gds-ANC	MISKLDSVLLAQKKFMYNYKNMRWAKGRNETYLCFVVKRRLGP DSLSFDFGHLRNRTGCHVELLFLSHLGALCPGLWGCGGDENRRL SYSVTWFCSWSPCANCAATLARFLRQTTNLRLRIFVARLYFCDLE DSPHIEGLRDLRRAGVQVTVMSYKDYFYCWQTFVAHRLSRFKA WEGLHTNSVRLSRKLNRILQTCETEDLRDAFRLIGLLT.	213	
Gf-ANC	MISKLDSVLLAQKKFIYNYKNMRWALGRNETYLCFVVKRRLGP DSLSFDFGHLRNRTGCHVELLFLSHLGALCPGLWGCGGDENRRL SYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLE DSPHIEGLRDLRRAGVQVTVMSYKDYFYCWQTFVAHRLSRFKA WEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGLLT.	213	83: E (56%) G (32%) 209: I (73%) F (20%)
Zg-ANC	MITKLDSVLLAQKKFIYHYKNMRWAKGRHETYLCFVVKRRVGP DSLSFDFGHLRNRTGCHVELLFLRHLGALCPGLWGYGGAGERRL SYSVTWFCSWSPCANCSFRLAQFLGQTPNLRLRIFVSRLYFCDLE DSREREGLRILKRAGVQITVMSYKDYFYCWQTFVAHRQSRFKA WDELHQNSVRLARKLNRILQPCETENLRDAFKLLGLL.	212	12: Q (72%) R (28%) 112: G (51%) S (43%) 178: E (73%) G (27%)

*: Site-specifc (local level) probablities

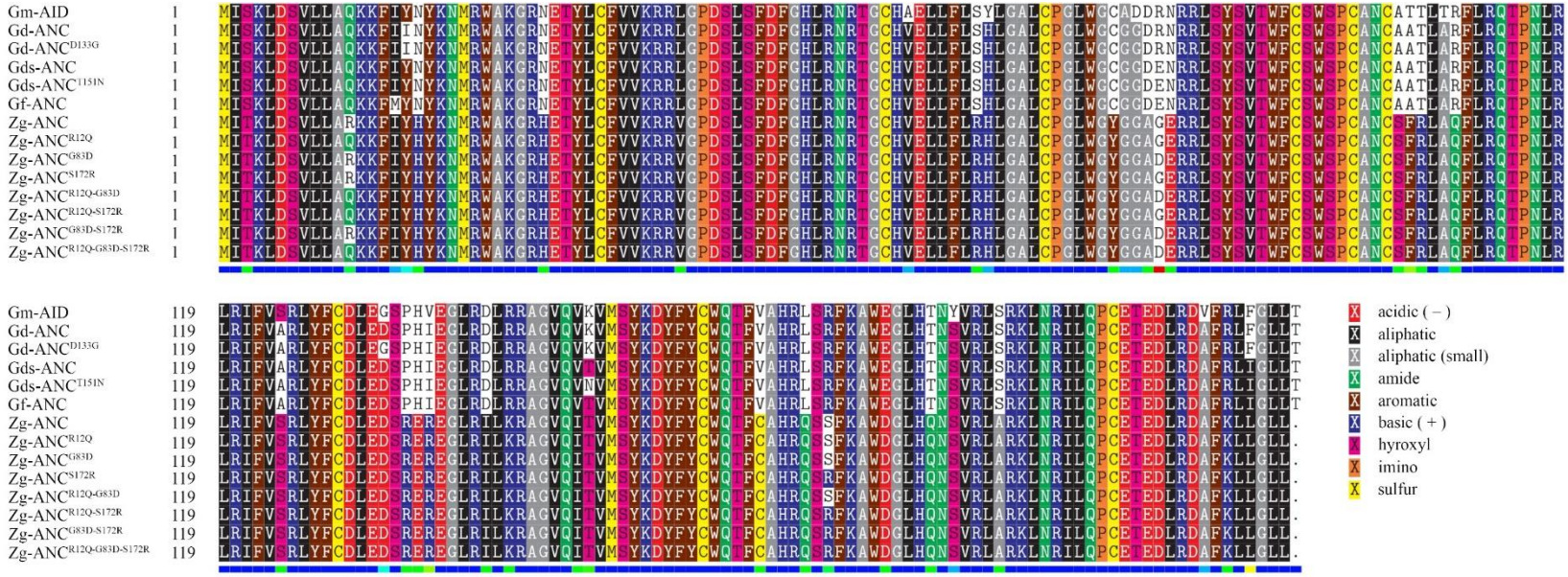


Figure 4-21: Amino acid alignment of the expressed ancestral AIDs. Any amino acid positions with less than 0.8 probability were synthesized as mutants. Amino acids are colored based on their chemical properties as indicated in the bottom right corner legends. The arbitrary cut-off of 0.2 was used to generate variants of the predicted ancestral AIDs. Abbreviations: Gd-ANC: Gadidae ancestor; Gds-ANC: Gadidae sister group ancestor; Gf-ANC: Gadiformes ancestor; and Zg-ANC: Zeiogadaria ancestor.

4.4.3.4 Biochemical properties of the predicted ancestral AIDs

We then synthesized, expressed, and purified 13 predicted ancestral AIDs and their variants (Figure 4-21) as GST-tagged fusion proteins to examine their biochemical properties. Using the resurrected ancestral AIDs, we sought to explore the effect on optimal temperature, optimal pH, K_m, and K_{cat} during Gadiformes' evolution (Figure 4-22, Figure 4-23, and Table 4-11). We found that the optimal temperature of AID was reduced from 12 °C to 8 °C in the Gadiformes common ancestor. At the same evolutionary time, the optimal pH of the AID was increased from 7.56 to 7.89. We also observed another increase in AID's optimal pH in the ancestor of Gadidae species (from 7.89 to 8.08).

Our results showed a reduction of about 15-fold in the catalytic rate (K_{cat}) of the Gadiformes ancestor compared to Zeiogadaria ancestor (1.90E-06 *vs.* 2.77E-05; Table 4-11). A more considerable reduction in the catalytic rate of AID was observed in the predicted ancestor of Gadidae species (~ 35- and ~500- fold reduction compared to Gadiformes and Zeiogadaria ancestor, respectively). We observed a 10-fold improvement when the K_m of Gf-ANC was compared to that of Zf-ANC (12.41 *vs.* 124.5). However, we observed a decline (less than 4-fold) in the K_m of Gd-ANC compared to that of Gf-ANC (46.7 *vs.* 12.41). The changes in the K_{cat} and K_m of ancestral AIDs resulted in ~ 30% and 99.5% reduction of catalytic efficiency (K_{cat}/K_m ratio) in the Gadiformes and Gadidae ancestral AIDs compared with Zg-ANC, respectively. Taken together, these results suggest that the functional impairment of AID likely occurred in the common ancestor of Gadidae group.

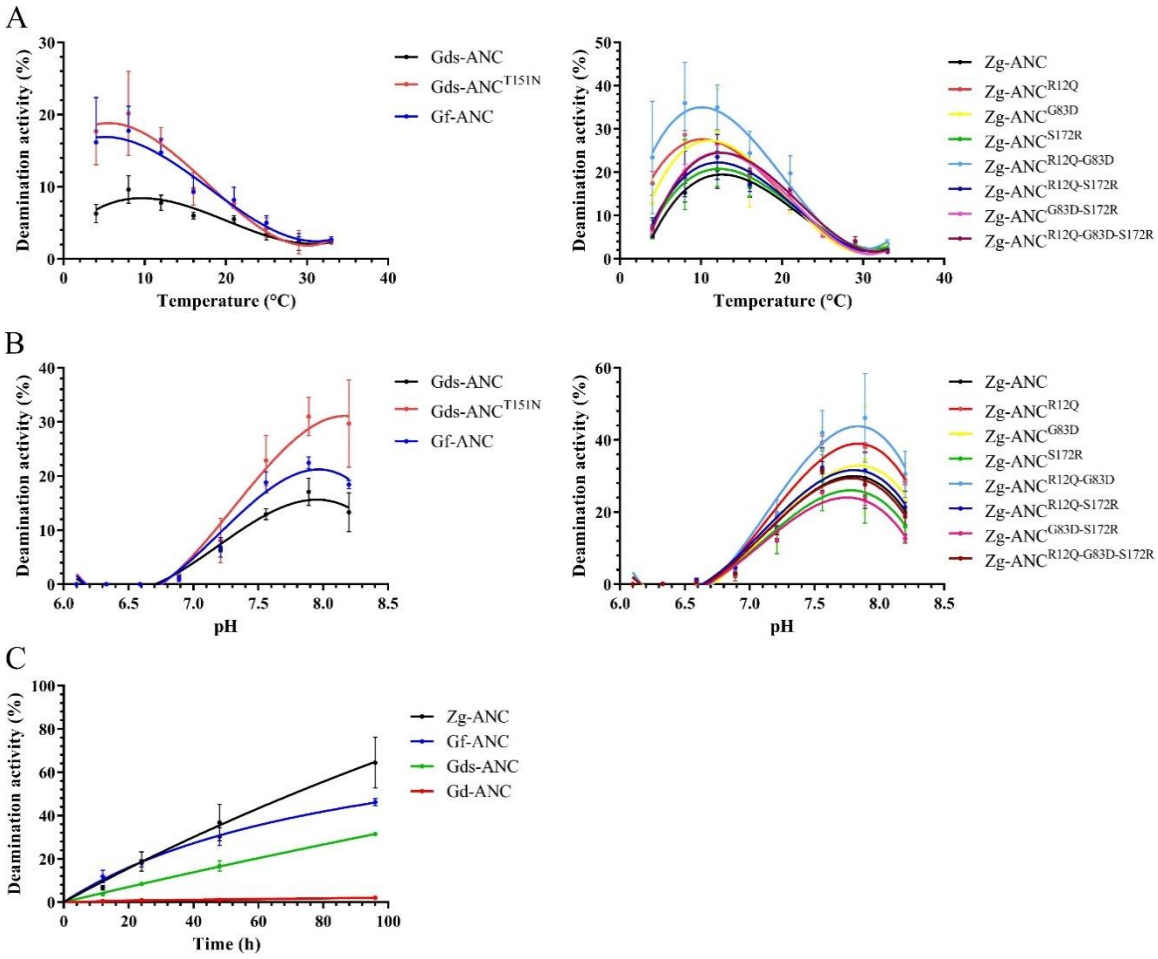


Figure 4-22: Biochemical properties of resurrected ancestral AIDs and their variants. Optimal temperature (A), optimal pH (B), and time-course kinetic (C) of predicted ancestral AIDs were measured using our standard alkaline cleavage assay. Two independent protein preparations of each ancestral AID were tested in duplicate. Data is presented as Mean ± SEM (n ≥ 4). Abbreviations: Gd-ANC: Gadidae ancestor; Gds-ANC: Gadidae sister group ancestor; Gf-ANC: Gadiformes ancestor; and Zg-ANC: Zeiogadaria ancestor.

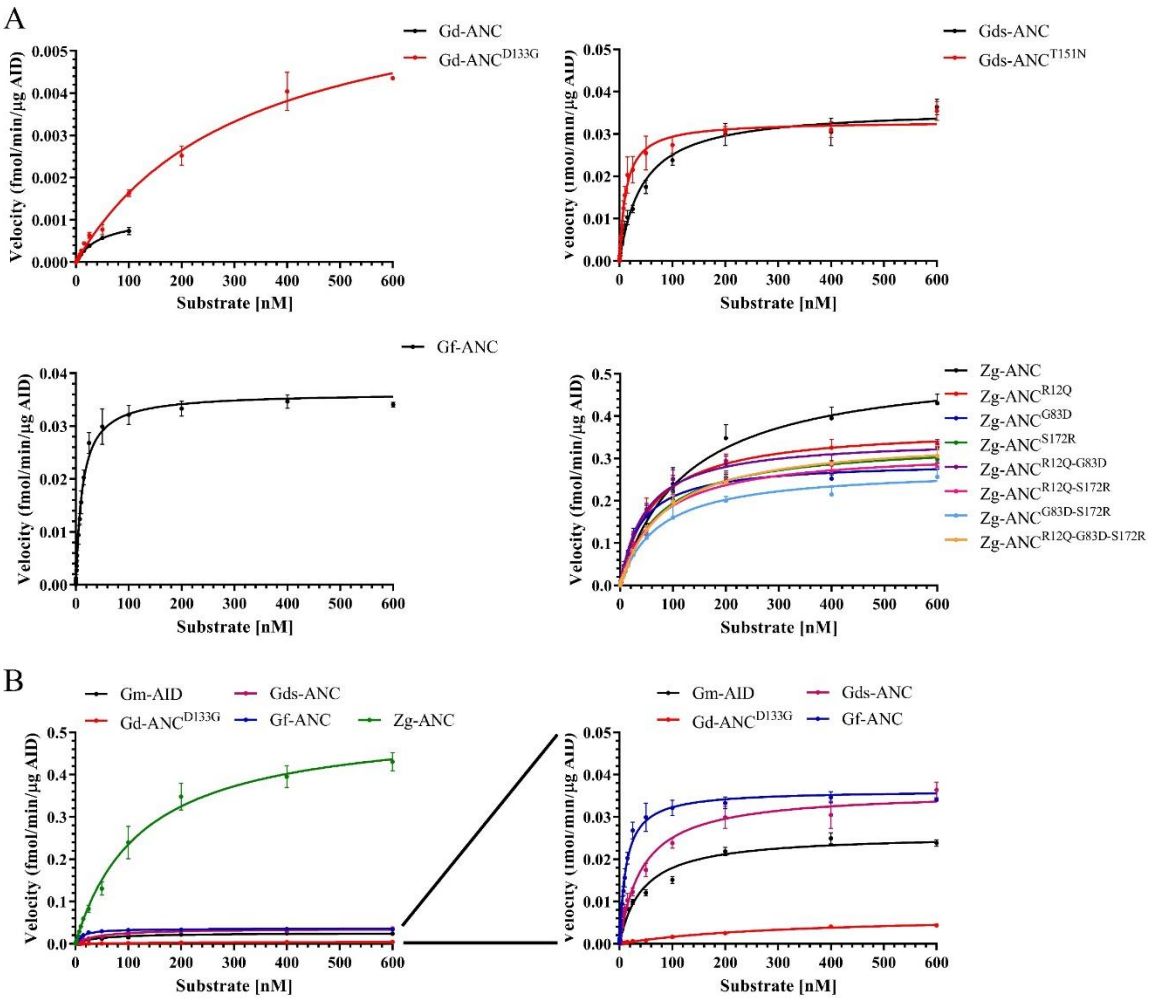


Figure 4-23: Comparison of the catalytic rate of predicted ancestral AIDs and their variants. A) The catalytic rate of resurrected ancestral AIDs and their variants was measured through Michaelis-Menten kinetics. At least two independent protein preparations of each AID protein were tested at their optimal temperature and pH with 0.03125-600 fmol range of TGCbub7 substrate. Each reaction was carried out in duplicate. Data is represented as mean ± SEM (n ≥ 4). Due to difference in the catalytic activity of ancestral AIDs, each ancestral AID was plotted separately. Please note that the y-axes have different scales. B) For better comparison, the results for ancestral AIDs were plotted with Gm-AID. For each ancestral node, only the most probable AID protein was included. In the case of common ancestor of Gadidae, the variant (Gd-ANC^{D133G}) was used due to the extremely low activity of the Gd-ANC. Abbreviations: Gd-ANC: Gadidae ancestor; Gds-ANC: Gadidae sister group ancestor; Gf-ANC: Gadiformes ancestor; and Zg-ANC: Zeiogadaria ancestor.

Table 4-11: The enzymatic parameters measured for predicted ancestral AIDs

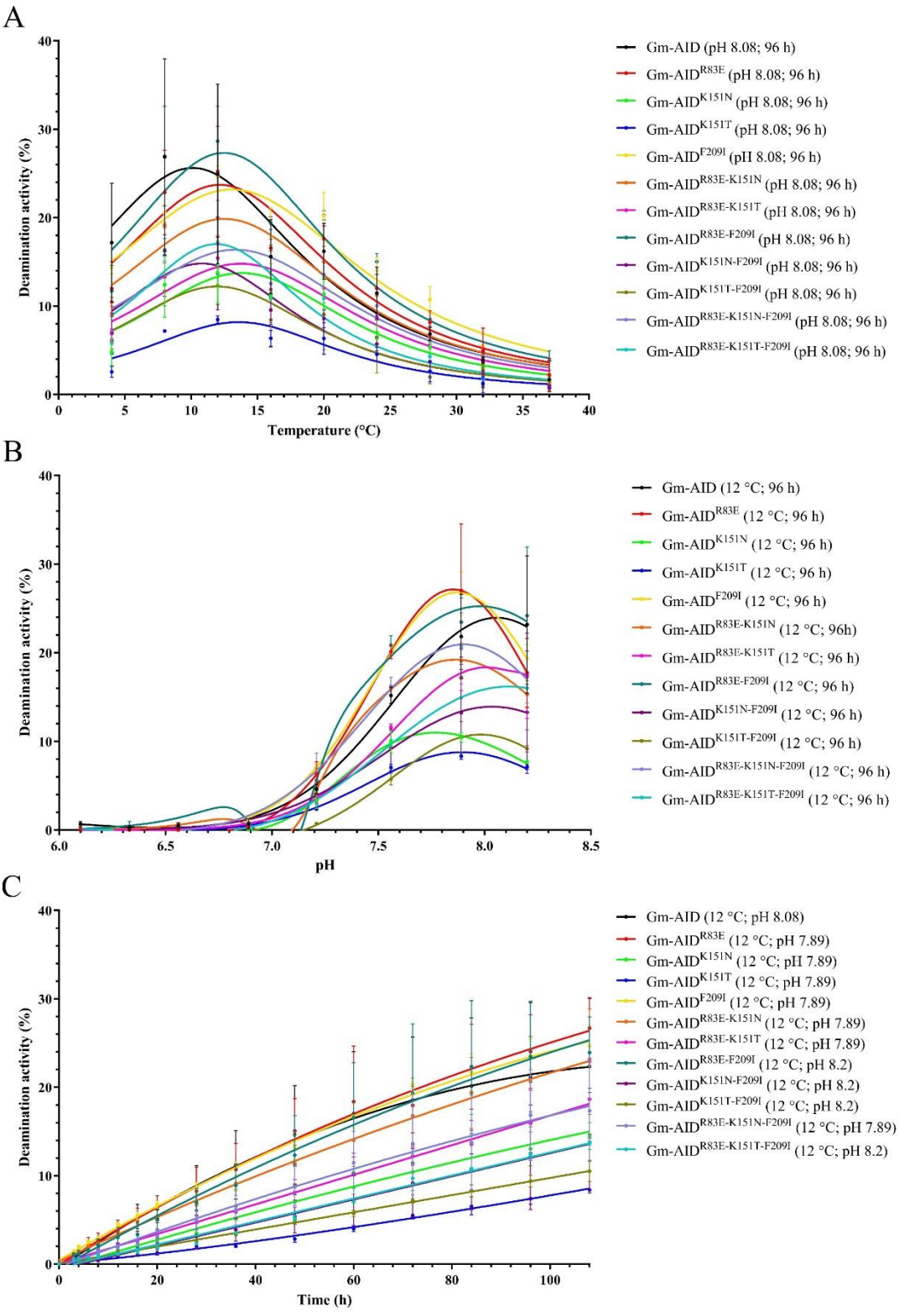
	Temp. (°C)	pH	K_{cat} (min^{-1})	K_m (nM)	V_{max} (fmol/	Std. Error		R^2	K_{cat}/K_m ($min^{-1}nM^{-1}$)	K_{cat}/K_m change (%)*
						K_{cat} (m	K_m			
Gd-ANC	8	8.08	5.72E-08	46.7	0.001	2.56E-09	3.969	0.98	1.22E-09	0.55
Gd-ANC^{D133G}	8	8.08	3.58E-07	316.9	0.007	1.60E-08	29.55	0.99	1.13E-09	0.51
Gds-ANC	8	7.89	1.89E-06	43.82	0.036	5.56E-08	4.475	0.95	4.31E-08	19.39
Gds-ANC^{T151N}	8	7.89	1.73E-06	12.33	0.032	5.33E-08	1.513	0.92	1.40E-07	63.05
Gf-ANC	8	7.89	1.90E-06	12.41	0.036	5.05E-08	1.084	0.96	1.53E-07	68.95
Zg-ANC	12	7.56	2.77E-05	124.5	0.527	1.05E-06	13.26	0.97	2.22E-07	100.00
Zg-ANC^{R12Q}	12	7.56	1.97E-05	60.9	0.375	5.52E-07	5.636	0.97	3.23E-07	145.37
Zg-ANC^{G83D}	12	7.56	1.54E-05	39.77	0.293	4.29E-07	3.895	0.96	3.87E-07	174.11
Zg-ANC^{S172R}	12	7.56	1.79E-05	75.58	0.340	4.97E-07	6.649	0.97	2.37E-07	106.55
Zg-ANC$^{R12Q-G83D}$	12	7.56	1.82E-05	47.62	0.347	4.48E-07	4.01	0.97	3.83E-07	172.22
Zg-ANC$^{R12Q-S172R}$	12	7.56	1.68E-05	69.12	0.319	4.45E-07	5.911	0.97	2.43E-07	109.09
Zg-ANC$^{G83D-S172R}$	12	7.56	1.46E-05	73.22	0.277	5.04E-07	7.633	0.96	1.99E-07	89.53
Zg-ANC$^{R12Q-G83D-S172R}$	12	7.56	1.85E-05	84.3	0.350	3.82E-07	4.688	0.99	2.19E-07	98.40

*: The change in the catalytic efficiency was compared to the K_{cat}/K_m of Zg-ANC.
Abbreviations: Gd-ANC: Gadidae ancestor; Gds-ANC: Gadidae sister group ancestor; Gf-ANC: Gadiformes ancestor; and Zg-ANC: Zeiogadaria ancestor.

4.4.4 The potential functional effects of AID's ancestral amino acid mutations

Next, we explored the effect of the amino acid changes observed in the predicted ancestral AIDs. The ~ 35-fold reduction in the K_{cat} of the Gd-ANC compared with Gds-ANC was the result of four amino acid differences (*i.e.,* I17Y, R83E, K151T, and F209I in Gd-ANC *vs.* Gds-ANC; Figure 4-21). In Gm-AID, these positions are the same as Gds-ANC except for position 17. Therefore, we changed the other three positions in Gm-AID into the corresponding amino acids in the Gds-ANC and studied their functional impact on the biochemical properties of AID. We explored optimal temperature, optimal pH, K_{cat}, K_m, and enzymatic efficiency of the Gm-AID mutants (Figure 4-24 and Table 4-12).

All the mutants revealed a higher optimal temperature (12 °C) compared with wildtype Gm-AID (8 °C, Figure 4-24 A). However, the effect on the optimal pH was minor and not consistent (Figure 4-24 B). Among the mutants, only Gm-AID^{F209I} and Gm-AID$^{R83E-K151N}$ exhibited higher K_{cat} than Gm-AID while Gm-AID^{R83E} and Gm-AID^{K151T} had similar K_{cat} to Gm-AID. All other mutants showed reduced K_{cat} (Table 4-12). The estimated K_m data showed that Gm-AID^{F209I}, Gm-AID$^{R83E-F209I}$, and Gm-AID$^{R83E-K151T-F209I}$ positioned dC more efficiently in the catalytic pocket compared to the wildtype Gm-AID. Also, we observed that among three mutations studied here, K151T/N has the highest deteriorating effect on the substrate binding affinity (*i.e.,* positioning dC in the catalytic pocket) of Gm-AID (Table 4-12). We concluded that the change of F209 to I might be responsible for the difference in the catalytic rate of Gd-ANC *vs.* Gds-ANC. This change only requires a T to A mutation in the first codon position (TTT and TTC encode F, and ATT and ATC encode I).


A
Deamination activity (%)
Temperature (°C)
Gm-AID (pH 8.08; 96 h)
Gm-AID^R83E (pH 8.08; 96 h)
Gm-AID^K151N (pH 8.08; 96 h)
Gm-AID^K151T (pH 8.08; 96 h)
Gm-AID^F209I (pII 8.08; 96 h)
Gm-AID^R83E-K151N (pH 8.08; 96 h)
Gm-AID^R83E-K151T (pH 8.08; 96 h)
Gm-AID^R83E-F209I (pH 8.08; 96 h)
Gm-AID^K151N-F209I (pH 8.08; 96 h)
Gm-AID^K151T-F209I (pH 8.08; 96 h)
Gm-AID^R83E-K151N-F209I (pH 8.08; 96 h)
Gm-AID^R83E-K151T-F209I (pH 8.08; 96 h)
B
Deamination activity (%)
pH
Gm-AID (12 °C; 96 h)
Gm-AID^R83E (12 °C; 96 h)
Gm-AID^K151N (12 °C; 96 h)
Gm-AID^K151T (12 °C; 96 h)
Gm-AID^F209I (12 °C; 96 h)
Gm-AID^R83E-K151N (12 °C; 96h)
Gm-AID^R83E-K151T (12 °C; 96 h)
Gm-AID^R83E-F209I (12 °C; 96 h)
Gm-AID^K151N-F209I (12 °C; 96 h)
Gm-AID^K151T-F209I (12 °C; 96 h)
Gm-AID^R83E-K151N-F209I (12 °C; 96 h)
Gm-AID^R83E-K151T-F209I (12 °C; 96 h)
C
Deamination activity (%)
Time (h)
Gm-AID (12 °C; pH 8.08)
Gm-AID^R83E (12 °C; pH 7.89)
Gm-AID^K151N (12 °C; pII 7.89)
Gm-AID^K151T (12 °C; pH 7.89)
Gm-AID^F209I (12 °C; pH 7.89)
Gm-AID^R83E-K151N (12 °C; pH 7.89)
Gm-AID^R83E-K151T (12 °C; pH 7.89)
Gm-AID^R83E-F209I (12 °C; pH 8.2)
Gm-AID^K151N-F209I (12 °C; pH 8.2)
Gm-AID^K151T-F209I (12 °C; pH 8.2)
Gm-AID^R83E-K151N-F209I (12 °C; pH 7.89)
Gm-AID^R83E-K151T-F209I (12 °C; pH 8.2)

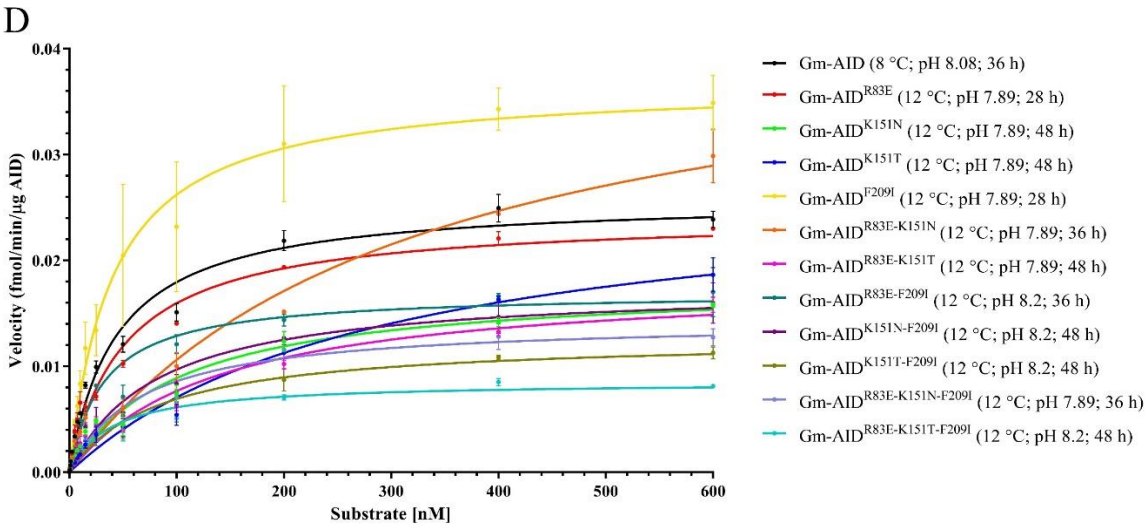


Figure 4-24: Biochemical properties of Atlantic cod AID mutants. To explore the functional impact of some ancestral mutations predicted during the evolution of AID within the Gadiformes lineage, we substituted amino acids in three positions in Gm-AID with that of corresponding predicted amino acid(s) in Gds-ANC. The optimal temperature (A), optimal pH (B), time course kinetics (C), and the catalytic rate (Michaelis-Menten kinetics; D) were compared to that of wildtype Gm-AID. At least two independent protein preparations of each AID were tested in duplicate ($n \geq 4$). Data is represented as Mean ± SEM.

Table 4-12: The enzymatic parameters measured for Gm-AID ancestral mutants

	Temp. (°C)	pH	K_{cat} (min^{-1})	K_m (nM)	V_{max} (fmol/min/µg)	Std. Error		R^2	K_{cat}/K_m ($min^{-1}nM^{-1}$)	K_{cat}/K_m change (%)*
						K_{cat} (min^{-1})	K_m (nM)			
Gm-AID	8	8.08	1.36E-06	44.05	0.026	3.05E-08	3.421	0.97	3.09E-08	100.00
Gm-AID^{R83E}	12	7.89	1.28E-06	51.95	0.024	4.00E-08	5.519	0.98	2.46E-08	79.44
Gm-AID^{K151N}	12	7.89	9.45E-07	102	0.018	4.35E-08	13.9	0.97	9.26E-09	29.96
Gm-AID^{K151T}	12	7.89	1.46E-06	293.9	0.028	1.03E-07	44.15	0.98	4.98E-09	16.11
Gm-AID^{F209I}	12	7.89	1.93E-06	40.22	0.037	8.89E-08	6.487	0.95	4.81E-08	155.52
Gm-AID$^{R83E-K151N}$	12	7.89	2.33E-06	317.3	0.044	2.14E-07	60.75	0.96	7.34E-09	23.74
Gm-AID$^{R83E-K151T}$	12	8.08	9.63E-07	138.8	0.018	5.95E-08	23.39	0.96	6.94E-09	22.43
Gm-AID$^{R83E-F209I}$	12	8.08	8.95E-07	32.21	0.017	3.46E-08	4.481	0.95	2.78E-08	89.86
Gm-AID$^{K151N-F209I}$	12	8.08	9.27E-07	81.85	0.018	4.33E-08	11.91	0.96	1.13E-08	36.63
Gm-AID$^{K151T-F209I}$	12	8.08	6.69E-07	84	0.013	2.88E-08	11.22	0.97	7.96E-09	25.74
Gm-AID$^{R83E-K151N-F209I}$	12	7.89	7.55E-07	67.18	0.014	3.51E-08	10.12	0.96	1.12E-08	36.33
Gm-AID$^{R83E-K151T-F209I}$	12	8.08	4.49E-07	39.47	0.008	1.59E-08	4.926	0.97	1.14E-08	36.78

*: The change in the catalytic efficiency was compared to the K_{cat}/K_m of Gm-AID.

4.5 Discussion

Previous studies in jawed vertebrates have shown that during a humoral antibody response, activation of B cells leads to expression of the *aicda* gene (Maul & Gearhart, 2010; Owen, 2019). The product of this gene, AID, introduces mutations in *Ig* genes, leading to generation of antibodies with higher affinity for cognate antigen (Betz et al., 1993; Bromage et al., 2006; Cain et al., 2002; Diaz et al., 1999; Dooley & Flajnik, 2005; Dooley et al., 2006; Hsu, 2016; Jenne et al., 2003; Kaattari et al., 2002; Lee et al., 2002; Malecek et al., 2005; Marianes & Zimmerman, 2011; Mehr et al., 2004; Wilson et al., 1992; Yang et al., 2006). Interestingly, the humoral immune system of Atlantic cod differs from other studied vertebrates. This species lacks antigen-specific high affinity antibodies (Arnesen et al., 2002; Lund et al., 2008; Lund et al., 2006; Magnadottir et al., 2001; Schroder et al., 2009; Solem & Stenvik, 2006). In previous chapters, we showed that despite the conservation of *aicda* gene synteny in Atlantic cod, the enzyme (Gm-AID) itself lacks robust catalytic efficiency compared with other examined AID homologs.

In this chapter, we sought to explore the evolutionary trajectory of AID's enzymatic properties leading to its functional impairment in Atlantic cod to ask three questions: first, is the deactivation of Gm-AID unique within the Gadiformes family? Second, if it is not unique, which other family members share this trait, and at what evolutionary point did the functional impairment of AID protein occur? Third, what are the amino acid changes responsible for AID functional impairment?

To answer these questions, we expressed, purified, and studied the biochemical properties of 36 extant homologs within and outside of the Gadiformes lineage.

Additionally, we applied ASR methodology to predict the AID sequence in the common ancestors of Gadidae, its sister group, Gadifomes, and Zeiogadaria species. We found that during adaptation of AID enzyme to the ambient temperature of the Gadiformes species, its catalytic efficiency was gradually reduced in the evolutionary branches leading to the Atlantic cod. Given the previously observed remodeling of immune system in the Gadiformes species where the loss of the main humoral immune genes coincided with the expansion of genes involved in the cell-mediated and innate immune systems, our findings suggests that the functional impairment of AID gene within Gadifomres is most likely a continuation of their immune system drastic remodeling.

Here we studied the biochemical properties of 36 extant AID homologs using our established alkaline cleavage assay (Abdouni et al., 2013; Abdouni et al., 2018; Dancyger et al., 2012; Emma M. Quinlan, 2017; King et al., 2015; Larijani & Martin, 2007). Interestingly, we could not detect any cytidine deaminase activity for Bs-AID (polar cod) and Mz-AID (Arrowtail) in our assays. Also, amongst the studied extant species here, Ag-AID (arctic cod) is the only AID exhibiting significantly lower catalytic efficiency than Gm-AID (~ 8-fold less). There are only three amino acid differences between Bs-AID and Gm-AID: S3R, K13N, and L143P in Gm-AID *vs*. Bs-AID. Amongst these differences, it seems that the drastic change from leucine (L) to proline (P) at position 143 might be crucial in the absence of cytidine deaminase activity in Bs-AID. Gm-AID^{L143} resides in the α3, and its replacement with a proline most likely resulted in a shorter α3 in Bs-AID. Considering the close phylogenetic relationship between Atlantic cod, polar cod, and arctic cod and their extremely low catalytic efficiencies (in Ag-AID and Gm-AID) or the lack of

cytidine deaminase activity (in Bs-AID), it seems that evolution in Gadidae family might be directed towards loss of AID activity. In line with these findings, we observed that the CDRs of Gadidae species exhibited no or lowest enrichments of AID hotspots compared with other vertebrates. Understanding the structural basis of the lack of cytidine deaminase activity of Bs-AID and Mz-AID requires more detailed computational and mutational analyses which was beyond the scope of this thesis.

We also discovered a cold adaptation of AID enzyme amongst species studied here which seemed to be governed by their habitat temperature as suggested before (Appendix 6) (Barreto et al., 2005; Conticello et al., 2005; Dancyger et al., 2012; Ichikawa et al., 2006; Wakae et al., 2006). For the first time, here, we showed that some AID homologs exhibit cytidine deaminase activity in the temperatures below 0 °C. Tsc-AID, Tmu-AID, and Mmor-AID demonstrated optimal temperature of 0 °C while maintaining more than 50% of their maximum catalytic activity at -10 °C. *T. scarbus*, *T. murrayi*, and *M. mora* live in the deep-water (as low as 2000 m) and this might explain their lower optimal temperature (www.fishbase.se). Additionally, the result of our correlation and clustering analyses uncovered a strong positive relationship between optimal temperature and the catalytic efficiency of extant AIDs. In other words, unlike previously studied cold-adapted enzymes which retained their catalytic efficiency, it appears that AID enzyme has lost its catalytic efficiency as the result of adaptation to the colder temperatures. Another possible scenario is that the cold adaptation and low catalytic activity are not related but only occurred at a close evolutionary time.

Although the exact structural adjustments occurred during the evolution of the cold-adapted enzymes are not fully understood, reducing thermal stability was proposed as the mechanism to increase catalytic efficiency at low temperatures (Pucci & Rooman, 2017; Smalas et al., 2000). The reduced stability may be accomplished by structural changes such as intra-molecular hydrogen bonds and ion-pairs, proline-, methionine-, glycine-, or arginine content, surface hydrophilicity, helix stability, and core packing (Marshall, 1997; Smalas et al., 2000). Siddiqui and Cavicchioli reviewed what is known about cold adaptation and found seven strategies employed by cold-adapted enzymes: on the surface, cold-adapted enzymes tend to have more hydrophobic residues; more surface exposed negatively charged amino acids have been observed in cold-adapted enzymes; serine (S) can be replaced by an alanine (A) to reduce the intramolecular H-bonds in cold-adapted enzymes; reduced Arginine/Lysine ratio was also observed in some cold-adapted enzymes while this ratio was increased in some others; aromatic interactions and salt bridges may also be less in cold-adapted enzymes; generally, in the cold-adapted enzymes, the Ncap (N terminus) and the Ccap (C terminus) of the a-helix are more positively and negatively charged, respectively; in the loops of cold-adapted enzymes, the number of the prolines is less while the abundance of glycine residues is increased (Siddiqui & Cavicchioli, 2006). In our dataset, we noticed an interesting 20-degree optimal temperature difference between AIDs of two closely related species of *T. subterraneus* and *P. transmontana*. Examining the amino acid differences amongst these two AIDs and our preliminary computational analyses (data not shown) did not reveal many significant structural adjustments as seen in previously studied cold-adapted enzymes. The only structural adjustment noticed here is

that Tsu-AID α3 has a more positive Ncap and less positive Ccap compared to that of Pt-AID (Pt-AID$^{A101\text{-}L105\text{-}I110\text{-}R112}$ *vs*. Tsu-AID$^{S101\text{-}H105\text{-}F110\text{-}S112}$). Further mutational and biochemical analyses are required to investigate the impact of this structural adjustment in their optimal temperature difference.

To explore the potential mechanism(s) employed by Pt-AID to improve its thermoresistance compared to Tsu-AID, we predicted their stability curve using SCooP server (Pucci et al., 2017; Pucci & Rooman, 2014, 2016). We found that most likely the higher thermoresistance of Pt-AID is due to its more negative enthalpy change at the temperature of maximum stability (ΔH_s) which caused its ΔG_r to decrease as well. It is important to note that although pH has a significant influence on the evaluation of the thermodynamic parameters, the current version of SCooP predicts the stability curve only at pH 7 (Pucci et al., 2017; Pucci & Rooman, 2014, 2016). Pt-AID showed an optimal pH of 7.6 while the optimal pH of Tsu-AID was measured at 7.9 in our assays. Therefore, measuring their stability curves at their corresponding optimal pH in the lab or developing a software capable of predicting the stability curve at a given pH is required to confirm these results.

In general, cold-adapted and temperate enzymes usually exhibit a similar catalytic efficiency at their corresponding optimal temperature (Marshall, 1997). This is usually accomplished by increasing the K_{cat} and decreasing the K_m during the process of cold adaptation. The increased catalytic activity of cold-adapted enzymes was attributed to optimization of electrostatic interactions at and around the active site which results in more flexibility around the active site (Siddiqui & Cavicchioli, 2006; Smalas et al., 2000). The

higher local flexibility around the catalytic pocket usually results in higher K_{cat} and K_m (Siddiqui & Cavicchioli, 2006). However, studies on A4-lactate dehydrogenase (A4-LDH) and cytosolic malate dehydrogenase (cMDH) showed that substrate affinity decreases during evolution of cold-adapted enzymes to increase catalytic rate efficiency (Fields et al., 2015). Importantly, in our dataset, we did not detect catalytic efficiency retention during cold adaptation process of AID enzyme. In other words, it seems that unlike metabolic enzymes that maintain their catalytic efficiency in the psychrophilic organisms, AID catalytic efficiency was reduced in Gadiformes species studied here. This may suggest that, in these species, the cost of antibody maturation may outcompete its benefit, eliminating the need to maintain an active AID enzyme.

By resurrecting the ancestral AIDs as old as 120 million years ago (Zg-AID), we successfully pinpointed the major AID's functional changes occurred during the evolution of Gadoformes species. The measured biochemical properties of the predicted ancestral AIDs confirmed the cold adaptation of AID enzyme while losing the catalytic efficiency. Specifically, we observed a four-degree reduction in the optimal temperature of Zg-ANC to Gf-ANC (12 °C to 8 °C) while losing 30% of its catalytic efficiency. Although the optimal temperature of Gd-ANC was similar to that of Gf-ANC, its catalytic efficiency was significantly impaired (99.2%) due to 97% reduction in K_{cat} and 376% increase in K_m values. These findings further confirmed the earlier suggested scenario where the cost of maintaining the antibody maturation process in Gadiformes, and especially Gadidae family, outcompeted its benefits. Also, these findings suggest that the first reduction in the catalytic activity of AID, occurred in the Gadiformes ancestor, could be due to adaptation

to the cold temperature while the second reduction in the catalytic efficiency, observed in the Gadidae ancestor, was independent of cold adaptation.

The K_{cat}/K_m ratio obtained for the Gds-ANC and its variant indicated two possible scenarios. If the Gds-ANC is the true ancestral AID at this node, a ~ 30% reduction in the catalytic efficiency has occurred during the evolution of AID in the common ancestor of Gadidae sister group. If Gds-ANC^{T151N} is the true ancestral AID at this node, then one can conclude that AID catalytic efficiency did not change during the evolution of the Gadidae sister group compared to that of Gf-ANC.

Comparison of amino acid sequence of Gf-ANC with Gd-ANC revealed five amino acid differences. These variations which are responsible for the 97% reduction in K_{cat} and 376% increase in K_m values are: M16I, Y17I, E83R, T151K, ad I209F in Gf-ANC *vs*. Gd-ANC. Since Gds-ANC which contains an isoleucine (I) in position 16 (similar to Gf-ANC) exhibited the same K_{cat} as Gf-ANC (1.89E-06 ± 5.56E-08 *vs*. 1.90E-06 ± 5.05E-08, respectively) but the same K_m as Gd-ANC (43.82 ± 4.475 *vs*. 44.05 ± 3.421, respectively), we concluded that position 16 could be responsible for the 376% increase in the K_m of the Gd-ANC compared with Gf-ANC. Therefore, all or a portion of the other remaining four amino acid differences have contributed to the 97% reduction of K_{cat} in Gd-ANC compared with that of Gf-ANC. Replacement of these amino acids in Gm-AID revealed a 1.5-fold increase in K_{cat}. This slight improvement in the K_{cat} of Gm-AID is far less than the 33-fold improvement in the K_{cat} of Gf-ANC compared with Gd-ANC, suggesting the presence of epistatic mutations in Gm-AID which prevented the positive effect of causative mutation(s)

to be observed. Further mutational analyses are required to figure out the exact position(s) responsible for the drastic reduction in Gd-ANC K_{cat} value.

In summary, here we reported that a similar reduction in catalytic activity of AID, detected in Atlantic cod in the previous chapter, could also be observed in other species of Gadiformes order, especially within the Gadidae family. For the first time, here, we investigated the functional evolutionary trajectory of AID enzyme within the Gadiformes order. We found that while AID was evolved to adapt to the lower temperatures mirroring the ambient temperature of Gadiformes species, it lost its catalytic efficiency. However, the more drastic reduction of catalytic efficiency observed in Gadidae ancestor seems to be a purposeful event to inactivate AID in this family of fish. Reduced catalytic efficiency (specially in Gadidae species), lack of cytidine deaminase activity (Bs-AID and Mz-AID), and potential exclusion of *aicda* gene from the genome (*B. cantori*) are some of the variations found in this report. These variations could indicate the presence of a previously unknown vast plasticity in the humoral and adaptive immune system of bony fish. Our comprehensive evolutionary comparative approach applied here could be a powerful tool to unmask the potential plasticity in other biological settings.

Chapter 5:

Discussion

5.1 Overview

Diversification of the B cell antigen receptors (*i.e.,* immunoglobulin genes [*Ig*]) is a vital step in the arms race between the host's humoral immune response and pathogens. The *Ig* genes undergo primary and secondary diversifications to generate the naïve and activated B cell antigen receptor repertoires, respectively (Maul & Gearhart, 2010; Owen, 2019). *Ig* gene secondary diversification is initiated when the enzyme activation-induced cytidine deaminase (AID, encoded by *aicda* gene) mutates deoxycytidine (dC) into deoxyuridine (dU) at *Ig* genes of activated B cells (Maul & Gearhart, 2010; Methot & Di Noia, 2017; Owen, 2019). As a result of secondary diversification, the affinity of the antibodies for the cognate antigen could increase as much as 1000-fold, enhancing the efficient recognition and neutralization of the pathogen by activated B cells (Magor, 2015; Meffre et al., 2001).

The presence of the AID gene and the antibody maturation process have been reported in many jawed vertebrate species studied thus far (Abos et al., 2018; Bromage et al., 2006; Cain et al., 2002; Dooley & Flajnik, 2005; Dooley et al., 2006; Hsu, 2016; Jenne et al., 2003; Kaattari et al., 2002; Malecek et al., 2005; Marianes & Zimmerman, 2011; Mehr et al., 2004; Reynaud et al., 1991; Wiens et al., 2003; Wilson et al., 1992; Yang et al., 2006). Some examples of these species are nurse shark (*Ginglymostoma cirratum*) (Diaz et al., 1999; Dooley & Flajnik, 2005; Dooley et al., 2006; Hsu, 1998, 2016; Lee et al., 2002; Malecek et al., 2005; Voss & Sigel, 1972; Zhu & Hsu, 2010), rainbow trout (*Oncorhynchus mykiss*) (Bromage et al., 2006; Cain et al., 2002; Kaattari et al., 2002; Ye et al., 2010), Atlantic salmon (*Salmo salar*) (Solem & Stenvik, 2006), channel catfish

(*Ictalurus punctatus*) (Yang et al., 2006), zebrafish (*Danio rerio*) (Marianes & Zimmerman, 2011), African clawed frog (*Xenopus laevis*) (Hsu, 1998; Wilson et al., 1992), rabbits (Mehr et al., 2004), chicken (Mehr et al., 2004), sheep (Reynaud et al., 1991), mouse (Betz et al., 1993; Chi et al., 2020; Owen, 2019; Rajewsky et al., 1987; Wiens et al., 2003; Yeap & Meng, 2019), and human (Chi et al., 2020; Imkeller & Wardemann, 2018; Owen, 2019; Yeap & Meng, 2019). Based on these reports, the current consensus in immunology is that antibody affinity maturation is an ancient process, present in all vertebrate species, and dating back to the ancestor of jawed vertebrates.

Interestingly, the immune responses of Gadiformes species seems to differ from other studied vertebrate species. One of the most studied Gadiformes species is the Atlantic cod (*Gadus mohua*). This species is important for the marine ecosystems and the economy of many nations with coast lines in the North Atlantic Ocean (*e.g*., the eastern Canadian provinces and several Scandinavian countries) due to being harvested in commercial food fisheries and forming a vital link in the aquatic food chain. Disease outbreaks in Atlantic cod stocks resulting in high mortality rates have been reported in Newfoundland, Nova Scotia, New Brunswick, and along the east coast of the USA (Frenette et al., 2017; Grove et al., 2003; Gudmundsdottir et al., 2006; Hong, 2013; Samuelsen et al., 2006). Functional analyses of the Atlantic cod humoral responses revealed lack of antibody maturation in this species (Arnesen et al., 2002; Corripio-Miyar et al., 2007; Magnadottir et al., 1999; Magnadottir et al., 2001; Solem & Stenvik, 2006). For example, immunization of Atlantic cod against *Vibrio anguillarum* did not generate any humoral response despite mounting protective immunity, suggesting the involvement of cell-mediated and/or other types of

immunity (Caipang et al., 2009; Gudmundsdóttir et al., 2009; Lund et al., 2007; Mikkelsen et al., 2011; Solbakken, Jentoft, Reitan, Mikkelsen, Jakobsen, et al., 2019). Also, in response to *Brucella pinnipedialis, Francisella noatunensis,* and *Vibrio salmonicida* infections, only a weak T-cell independent anti-LPS antibody response was detected in Atlantic cod (Ellingsen et al., 2011; Espelid et al., 1991; Lund et al., 2008; Lund et al., 2006; Nymo et al., 2016). Other than Atlantic cod, the vaccination of the gadoid haddock (*Melanogrammus aeglefinus*), another Gadiformes species, despite successful reduction in mortality, did not mount an antigen-specific antibody response (Corripio-Miyar et al., 2007).

The comparison of genomic sequences of 72 Gadiformes species to that of other bony fish revealed a unique absence of numerous genes that are central to humoral immune system (Malmstrom et al., 2016). These immune genes, involved in T-cell/B-cell interactions, include major histocompatibility complex (*mhc*) class II, cluster of differentiation 4 (*cd4*; pseudogene), and invariant chain (*Ii*) genes (Malmstrom et al., 2016; Star et al., 2011). In contrast, the *mhc I* and some Toll-like receptor (*tlr*) loci are significantly expanded in the Gadiformes fish compared to other teleost fish (Malmstrom et al., 2016; Parham, 2015, 2016; Solbakken, Rise, et al., 2016; Solbakken, Torresen, et al., 2016; Star et al., 2011; Torresen et al., 2017). Based on these observations and studies on the immune response of Atlantic cod, it was suggested that Gadiformes species may utilize alternative immune strategies to compensate for the lack of these genes including MHC I cross-presentation and T cell-independent activation of B cells (Malmstrom et al., 2013; Solbakken, Jentoft, Reitan, Mikkelsen, Gregers, et al., 2019).

Introducing AID-mediated somatic mutations in the *Ig* genes of the activated B cells is the initial step to generate antibodies with higher affinity for the cognate antigen (Bransteitter et al., 2003; Kolar et al., 2007; Larijani, Frieder, Basit, et al., 2005; Meffre et al., 2001; Muramatsu et al., 1999; Muto et al., 2000). Previous studies have shown WRC (W=A/T; R=A/G) motifs as AID's mutational hotspots (Bransteitter et al., 2003; Dancyger et al., 2012; Emma M. Quinlan, 2017; Larijani, Frieder, Basit, et al., 2005; Larijani et al., 2007; Meffre et al., 2001). Given the central role of AID activity in initiating antibody maturation process, a clear enrichment of WRC motifs was observed in the CDR portion of *Ig* genes of studied vertebrates (Conticello et al., 2005; Detanico et al., 2016; Golub & Charlemagne, 1998; Jolly et al., 1996; Oreste & Coscia, 2002; Wagner et al., 1995; Wei et al., 2015). The importance of AID substrate specificity co-evolution with the *Ig* gene sequence was validated when the replacement of these WRC motifs with AID coldspots reduced mutation frequency in *IgV* region (Wei et al., 2015). Additionally, analyzing the crystal structure of the antibody-antigen complex showed that the majority of the antibody residues interacting with the antigen are subject to AID-mediated mutations (Detanico et al., 2016). It is important to note that although creating the diversity in the adaptive immune antigen receptors is crucial to protect the host, any deviation and mis-regulation of this genome-damaging system is costly.

The diversification of adaptive immune antigen receptors is a unique example of deliberate controlled self-DNA mutation and rearrangement in vertebrates. One source of structural variations (SV) of chromosomes and mutations in B cells is the mis-targeted activity of AID (Choudhary et al., 2018; Trancoso et al., 2020). For example, in patients

with chronic myeloid leukemia (CML), AID-mediated hypermutation of tumor repressor and DNA repair genes have been associated with progress into fatal B lymphoid blast crisis and an Imatinib-resistance phenotype (Klemm et al., 2009). In diffuse large B cell lymphomas (DLBCL), somatic hypermutation (SHM) has been reported in proto-oncogenes (Seifert et al., 2019). The *IgH-cMYC* translocation is observed in Burkitt lymphoma where the frequency of this translocation was correlated with AID activity level (Takizawa et al., 2008). AID-induced hypermutations have also been observed in chronic lymphoid leukemia (CLL) (Burns et al., 2017). There has also been evidence of AID-mediated carcinogenesis in germinal center B cells as the result of Epstein-Barr virus (EBV)-induced AID expression (Mohri et al., 2017). Moreover, AID-mediated mutations have been observed in ovarian cancer (Lindley et al., 2016). Interestingly, under strong inflammatory stimuli, the premature expression of *aicda* during B cell development could drive the clonal evolution of childhood B cell acute lymphoblastic leukemia (B-ALL) (Swaminathan et al., 2015). It was proposed that aberrant AID-mediated mutations in CpG islands would create T:G mismatches which would cause SV (Swaminathan et al., 2015). Taken together, AID acts as a double-edged sword in immunity and cancer.

AID plays a central role in protecting vertebrate species by initiating the antibody affinity maturation process. Unlike other vertebrates, it seems that the antibody responses of Atlantic cod lack antibody maturation, exhibiting high levels of low affinity antibodies and lack of high affinity ones upon immunization (Arnesen et al., 2002; Magnadottir et al., 1999; Magnadottir et al., 2001; Solem & Stenvik, 2006). This scenario struck us as being

similar to patient with hyper IgM syndrome type II (HIGM II) where lack of AID activity results in the absence of antibody maturation (Minegishi et al., 2000; Revy et al., 2000).

5.2 Findings and significance

In this thesis, we sought to identify the molecular basis behind the lack of antibody maturation in Atlantic cod. We applied a comparative molecular-biochemical-computational methodology to study the genetics, expression, function, and evolutionary trajectory of AID in Atlantic cod. Our objectives were to answer four main questions: 1) Is the gene synteny and transcript expression of *aicda* gene conserved in Atlantic cod compared with other studied vertebrates? 2) Is Atlantic cod AID a functional cytidine deaminase? 3) Is the Atlantic cod case an exception among Gadiformes species? 4) At what evolutionary point did the adaptation(s) resulting in the lack of antibody maturation in Atlantic cod occur? We attempted to answer the first two questions in chapter 2 and 3, respectively, and the last two questions in chapter 4.

5.2.1 Summary of findings

In chapter 2, we found that the gene synteny and transcript expression of Atlantic cod *aicda* gene is similar to those of other Teleostei species studied here. Specifically, the analyses of *aicda* locus revealed conserved synteny throughout the Teleostei infraclass. In addition to the primary transcript containing the full-size AID coding sequence (CDS), we also detected a truncated transcript (*T-Gm-acida*), in which the first exon was missing. Although various AID isoforms have been identified in tetrapods (Albesiano et al., 2003; Marr et al., 2007; McCarthy et al., 2003; Muramatsu et al., 1999; Ohmori et al., 2004; Verma et al., 2010; Wu et al., 2008), exclusion of the first exon has only been observed in the lizard Iberian ribbed newt (*Pleurodeles waltl*) (Bascove & Frippiat, 2010), Moreover, no alternative *aicda* transcript isoform has been discovered in other bony fish species

studied thus far (Saunders & Magor, 2004; Zhao et al., 2005). Similar to other studied vertebrates, *Gm-aicda* transcripts were mainly expressed in immune-related tissues (Bascove & Frippiat, 2010; Marr et al., 2007; Muramatsu et al., 1999; Muto et al., 2000; Ohmori et al., 2004; Saunders & Magor, 2004; Verma et al., 2010). Also, like other vertebrate species (Pone et al., 2012), we detected an increased overall expression of *Gm-aicda* transcript in response to viral and bacterial mimic immune stimulations. Given these findings, we concluded that the genetics and expression of the *Gm-aicda* were mainly conserved compared to other teleost species.

In chapter 3, we synthesized, purified, and compared the enzymatic properties of Atlantic cod AID (Gm-AID) to human (Hs-AID) and several other fish AIDs. We found that despite having all the functional domains of the AID/APOBEC family, Gm-AID catalytic rate was orders of magnitude lower than any other studied AID homologs thus far. In line with the functional impairment of Gm-AID, we observed the lowest WRC and WGCW enrichment in the complementarity-determining regions (CDRs) of Atlantic cod *Ig* genes compared to other studied vertebrates. As expected, the optimal temperature of Gm-AID was estimated between 4 to 8 °C, indicating evolutionary adaptation of AID enzyme to the Atlantic cod's ambient temperature. Computational simulations detected a significant increase in an alternative ssDNA binding mode in Gm-AID where the substrate did not fit in the classical DNA binding grooves previously identified in Hs-AID (King et al., 2015; Qiao et al., 2017). When the potential amino acid positions involved in the alternative binding mode in Gm-AID were replaced with their counterparts in Hs-AID or zebrafish AID (Dr-AID), the catalytic rate of the Gm-AID mutants was improved up to 10-

fold but still remained considerably lower than the other AID homologs examined. Based on these findings, two models became possible: either that the lack of antibody maturation in Atlantic cod is directly due to the functional impairment of its AID enzyme; or, alternatively, that because the Atlantic cod does not have the necessary mechanisms upstream of AID activity (*e.g.*, T-cell/B-cell interaction receptors) to initiate antibody affinity maturation. Previous studies in which B cell activation upon immune stimulation in Atlantic cod was observed (Solbakken, Jentoft, Reitan, Mikkelsen, Gregers, et al., 2019; Solbakken, Jentoft, Reitan, Mikkelsen, Jakobsen, et al., 2019) and our findings here that *Gm-aicda* expression was increased upon immune stimulation indirectly support the former scenario. However, to clearly distinguish between these two scenarios, functional analyses of antibody responses of other Gadiformes species are required. Since the major reduction in AID catalytic efficiency has occurred in the Gadidae ancestor, lack of affinity-matured antibodies in non-Gadidae species would prove the latter scenario. However, presence of affinity-matured antibodies in *Gadiculus argenteus*, a Gadidae species with an active AID, would prove the former scenario.

In chapter 4, we expanded our biochemical analyses to 36 species within and outside of Gadiformes lineage to investigate the functional properties of other Gadiformes AIDs (Figure 5-1), in order to shed light on the "which came first? The chicken or the egg" nature of the loss of MHC II pathway and functional impairment of AID. Within this lineage, we found AID homologs with no detectable cysteine deaminase activity (*i.e.*, Bs-AID and Mz-AID) and with catalytic efficiency lower than Gm-AID (*i.e.*, Ag-AID). Using ancestral sequence reconstruction (ASR) methods, we pinpointed the cold adaptation (12

°C to 8 °C) and functional impairment (99.2% reduction in catalytic efficiency) of AID enzyme in the common ancestor of Gadiformes and Gadidae, respectively. The asynchronous cold adaptation and functional impairment in the ancestral AIDs suggest that the functional impairment of the AID enzyme is a purposeful event not a byproduct of cold adaptation. Since the loss of *mhc II*, *cd4*, and *Ii* genes has occurred in the common ancestor of Gadiformes (Malmstrom et al., 2016), while the functional impairment of AID enzyme was identified in the common ancestor of Gadidae group, it seems that most likely the inactivation of AID was a consequence of the loss of central genes involved in the necessary mechanisms upstream of AID activity (*i.e.*, AM).

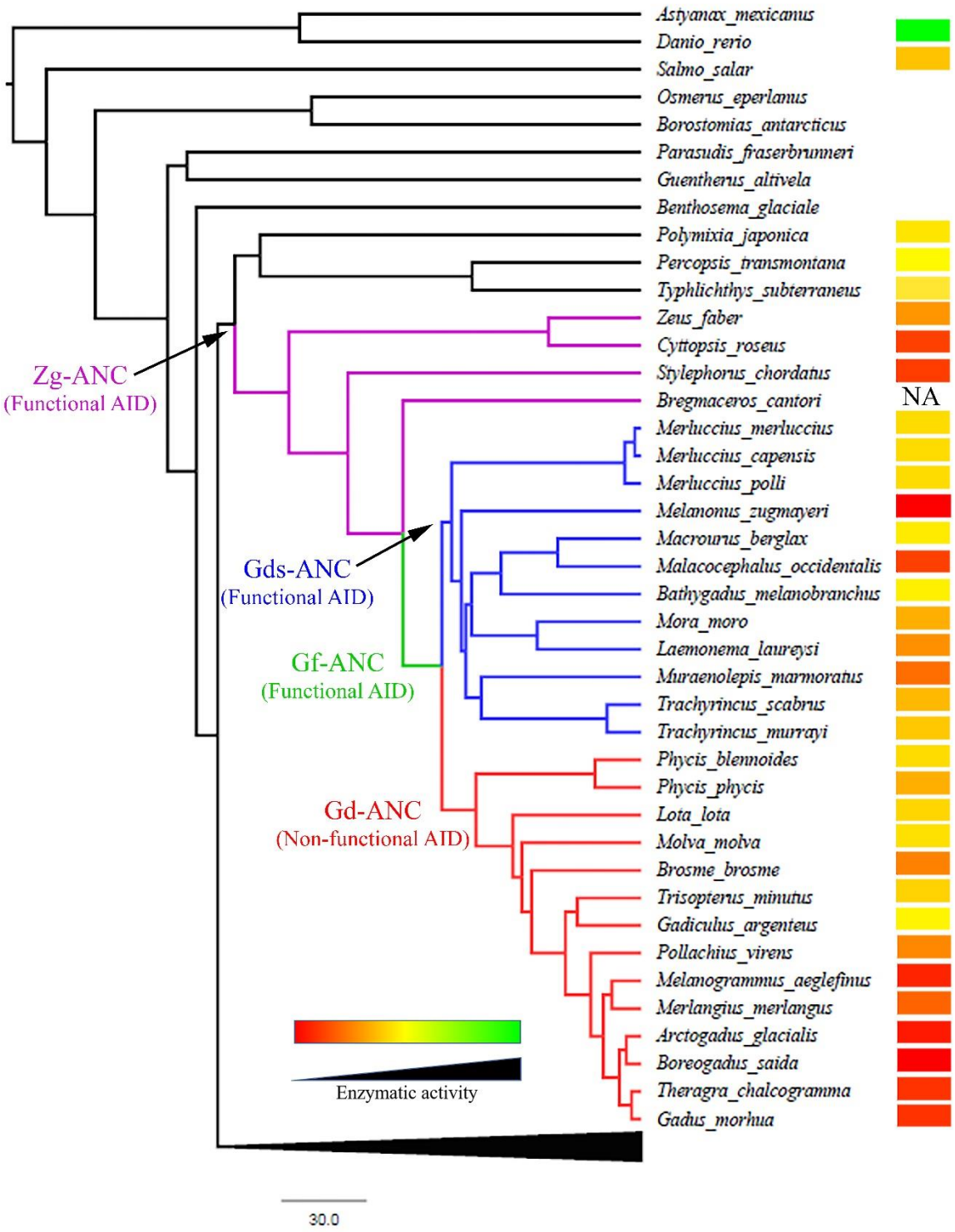


Figure 5-1: Comparison of catalytic rate of Gadiformes AIDs. Red to green color change indicates the low to high catalytic effeciency of AIDs. The K_{cat}/K_m data from Table 4-4 is used to draw this figure. NA indicates species without aicda gene.

5.2.2 Significance and future directions

In this thesis, for the first time, we reported two vertebrate species with functionally impaired AID enzymes (*B. saida* and *M. zugmayeri*). In human and mouse models, deficiency in AID function leads to the hyper IgM syndrome type II (HIGM II) characterized by a lack of affinity matured antibodies (Minegishi et al., 2000; Revy et al., 2000). Patients with HIGM II are susceptible to recurrent bacterial and opportunistic infections in respiratory and gastrointestinal tracts, autoimmunity, lymphoproliferation, and malignancies (Qamar & Fuleihan, 2014; Yazdani et al., 2019). However, Gadiformes species with functionally impaired AID, such as Atlantic cod, are healthy within their natural habitats (Parham, 2016).

It was suggested that the evolution of self-DNA-mutating enzymes such as AID has shaped the evolution of vertebrates' adaptive immune system where the invention of cellular machinery capable of introducing somatic mutations in the lymphocyte antigen receptors facilitated the evolution of the adaptive immune system (Trancoso et al., 2020). Interestingly, in the case of the Gadiformes lineage, it seems that the change in their common ancestor's habitat has altered their reliance on different branches of the immune system (Parham, 2016; Solbakken et al., 2017). Here, we proposed that the reduced dependency on humoral immunity, in return, has shaped the evolution of Gadiformes AID. In this scenario, the absence of strong reliance on the antibody response has eliminated the selective pressure to maintain AID functional, while the genome-wide collateral damage caused by AID off-target activity has formed a selective pressure to reduce/eliminate its activity.

Interestingly, we could not find evidence of the *aicda* gene in the striped codlet (*Bregmaceros cantori*) which also lacks *mhc I U*, *mhc II*, *cd4*, and *cd8* genes (Malmstrom et al., 2016). These genes are central to cell-mediated and humoral immune systems. On the other hand, we found that the teleost fish *Gouania willdenowi* which lacks *Ig* genes (Mirete-Bachiller et al., 2019), has maintained its *aicda* gene. Although biochemical analyses are required to confirm the activity or inactivity of its AID enzyme, the presence of the *aicda* gene in the absence of *Ig* genes shows a deeper level of plasticity within the vertebrate immune system, especially amongst bony fish. These new findings along with the previous studies prove that the vertebrate immune system dynamic is more flexible than currently believed. It seems that in the right environmental conditions, alternative immune strategies where one branch of immune system is shrinking can be successful in protecting the host (Figure 5-2).

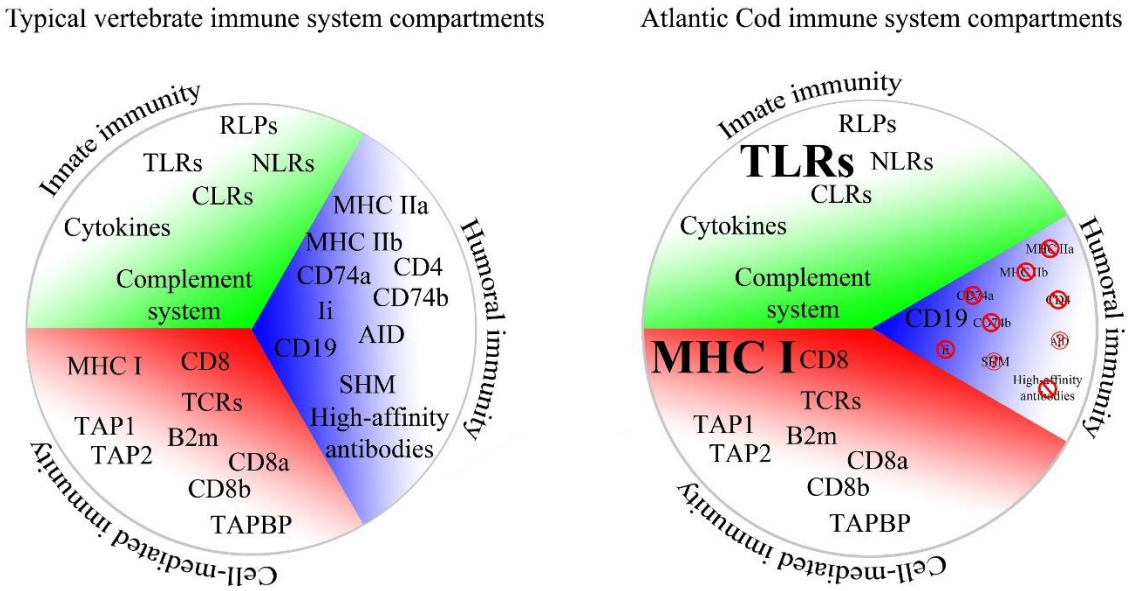


Figure 5-2: Model of a uniquely but successful compartmentalized immune system in Atlantic cod. Atlantic cod is genetically unique amongst all studied bony fish and vertebrates in that it is missing several genes that are essential for a robust secondary antibody response. On the other hand, it exhibits a unique expansion of other genes involved in cell-mediated and innate immunity. This alternative immune system is also present in other Gadiformes species.

Based on our current understanding of vertebrate immune system, the three branches of innate, cell-mediated, and humoral immunity are necessary to protect a species (Smith et al., 2019). However, the Gadiforms lineage is an exception where a drastic re-modeling of their immune system has re-invented the interwoven interaction between different branches of immune system. The first evidence came to light when, unlike other vertebrate species, the functional analyses of the Atlantic cod immune system revealed a weak humoral immune response upon immune stimulation and infection (Arnesen et al., 2002; Magnadottir et al., 1999; Magnadottir et al., 2001; Solem & Stenvik, 2006). The second line of evidence was provided when the genome of 72 teleost species, within and outside of Gadiformes lineage, were compared. This comparison uncovered a series of

gene losses and expansions unique to the Gadiformes lineage (Malmstrom et al., 2016; Solbakken, Rise, et al., 2016; Solbakken, Torresen, et al., 2016; Solbakken et al., 2017). For example, *mhc II*, *cd4*, and *cd74* (*i.e.*, invariant chain) genes were lost in the common ancestor of Gadiformes (Solbakken, Rise, et al., 2016). These genes play crucial roles in antibody responses (Owen, 2019). Concurrent with these gene losses, *mhc I* and some of the *tlr* genes were expanded in the common ancestor of Gadiformes lineage (Malmstrom et al., 2016; Solbakken, Rise, et al., 2016; Solbakken, Torresen, et al., 2016). These genes are involved in the cell-mediated and innate immunity, respectively (Owen, 2019). Interestingly, examining the binding domain and the sorting signaling sequence of the expanded *mhc I* genes in Atlantic cod showed novel signaling motifs similar to the ones involved in MHC II and cross-presentation pathways (Malmstrom et al., 2013). Based on these findings, it was suggested that the expansion and neofunctionalization of *mhc I* genes in Atlantic cod has led to generation of MHC II-like MHC I molecules (Malmstrom et al., 2013). However, the inactivation of Gadidae AIDs may suggest that the re-modeling of their immune system is moving towards shrinking the role of antibody response rather than converting other genes (*e.g.*, *mhc I*) to play the role of the lost genes involve in antibody response (*e.g.*, *mhc II*). It would be interesting to investigate these two scenarios in the future.

In addition to describing the novel insights on evolution of immunity, the findings presented in this thesis, are also significant from an enzyme structure:function perspective. The work presented in chapter 4 is the first endeavor to carry out full biochemical characterization on a large family of extant and ancestral versions of an enzyme involved

in human disease, and the first marriage of such evolutionary comparative enzymology with machine learning to shed light on structural and functional aspects of enzyme evolution. Since the discovery of AID and APOBECs, there has been hundreds of research papers investigating their evolution, regulation, structure, and function. Given the importance of these enzymes in human immunity and cancer, understandably, much of the research has been focused on human and to lesser extent to mouse counterparts. Studying the evolutionary trajectory of the current-day species is a powerful tool for understanding biology in the molecular, cellular, and organismal levels. In the comparative biology, exploring the differences between various species leads to discovering how natural selection has forced the evolution of the extant species (Martinez, 2018). Similarly, at the molecular level, comparative approaches can be used to study the enzymatic/biochemical properties of a protein. Comparative enzymology aims to discover the diversifying molecular mechanisms that altered enzymes' structure and function in response to the evolutionary pressures (Storey, 2016). Understanding these diversifying molecular mechanisms is an effective tool in understanding the proteins' structure-function relationship. To complement our comparative enzymology of chapter 4, future work should also focus on determination of the 3D structure of the AID homologs examined in chapter 4 by X-ray crystallography and NMR.

We have taken a comparative approach to study structure and function of AID. We hypothesized that since different AID homologs possess different enzymatic properties, examining their predicted structure side-by-side, would assist us in pinpointing the functional motifs of AID (Abdouni et al., 2013; Abdouni et al., 2018; Emma M. Quinlan,

2017; King & Larijani, 2017; King et al., 2015). Moreover, our comparative approach would assist in a better understanding the biological variations in the immune system of other species. Since vertebrate's immune system and their DNA/RNA editing enzymes, such as AID, have strongly influenced each other's evolution (Trancoso et al., 2020), examining the biochemical properties and evolutionary trajectory of other vertebrates' AID would shed light on other possible alternative immune strategies within this class. This is an area of research that has been neglected and the higher frequency of disease emergence in animals, due to environmental changes (*e.g.*, global warming) (Maslo et al., 2017), has created a strong need to put more effort into this type of research.

References

Abascal, F., Zardoya, R., & Telford, M. J. (2010, Jul). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res, 38*(Web Server issue), W7-13. https://doi.org/10.1093/nar/gkq291

Abdouni, H., King, J. J., Suliman, M., Quinlan, M., Fifield, H., & Larijani, M. (2013, May 01). Zebrafish AID is capable of deaminating methylated deoxycytidines. *Nucleic Acids Res, 41*(10), 5457-5468. https://doi.org/10.1093/nar/gkt212

Abdouni, H. S., King, J. J., Ghorbani, A., Fifield, H., Berghuis, L., & Larijani, M. (2018, Jan). DNA/RNA hybrid substrates modulate the catalytic activity of purified AID. *Mol Immunol, 93*, 94-106. https://doi.org/10.1016/j.molimm.2017.11.012

Abos, B., Estensoro, I., Perdiguero, P., Faber, M., Hu, Y., Diaz Rosales, P., Granja, A. G., Secombes, C. J., Holland, J. W., & Tafalla, C. (2018). Dysregulation of B Cell Activity During Proliferative Kidney Disease in Rainbow Trout. *Front Immunol, 9*, 1203. https://doi.org/10.3389/fimmu.2018.01203

Agius, C., & Roberts, R. J. (2003, Sep). Melano-macrophage centres and their role in fish pathology. *J Fish Dis, 26*(9), 499-509. http://www.ncbi.nlm.nih.gov/pubmed/14575368

Akanuma, S. (2017, Aug 6). Characterization of Reconstructed Ancestral Proteins Suggests a Change in Temperature of the Ancient Biosphere. *Life (Basel), 7*(3). https://doi.org/10.3390/life7030033

Albesiano, E., Messmer, B. T., Damle, R. N., Allen, S. L., Rai, K. R., & Chiorazzi, N. (2003, Nov 01). Activation-induced cytidine deaminase in chronic lymphocytic leukemia B cells: expression as multiple forms in a dynamic, variably sized fraction of the clone. *Blood, 102*(9), 3333-3339. https://doi.org/10.1182/blood-2003-05-1585

Alexaki, A., Kames, J., Holcomb, D. D., Athey, J., Santana-Quintero, L. V., Lam, P. V. N., Hamasaki-Katagiri, N., Osipova, E., Simonyan, V., Bar, H., Komar, A. A., & Kimchi-Sarfaty, C. (2019, Jun 14). Codon and Codon-Pair Usage Tables (CoCoPUTs): Facilitating Genetic Variation Analyses and Recombinant Gene Design. *J Mol Biol, 431*(13), 2434-2441. https://doi.org/10.1016/j.jmb.2019.04.021

Aleyd, E., Heineke, M. H., & van Egmond, M. (2015, Nov). The era of the immunoglobulin A Fc receptor FcαRI; its function and potential as target in disease. *Immunol Rev, 268*(1), 123-138. https://doi.org/10.1111/imr.12337

Aleyd, E., van Hout, M. W., Ganzevles, S. H., Hoeben, K. A., Everts, V., Bakema, J. E., & van Egmond, M. (2014, Mar 1). IgA enhances NETosis and release of neutrophil

extracellular traps by polymorphonuclear cells via Fcα receptor I. *J Immunol, 192*(5), 2374-2383. https://doi.org/10.4049/jimmunol.1300261

Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., & Ronquist, F. (2004, Feb 12). Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics, 20*(3), 407-415. https://doi.org/10.1093/bioinformatics/btg427

Ao, X., Sa, R., Wang, J., Dao, R., Wang, H., & Yu, H. (2016, Dec). Activation-induced cytidine deaminase selectively catalyzed active DNA demethylation in pluripotency gene and improved cell reprogramming in bovine SCNT embryo. *Cytotechnology, 68*(6), 2637-2648. https://doi.org/10.1007/s10616-016-9988-8

Arciuli, M., Fiocco, D., Fontana, S., Arena, M. P., Frassanito, M. A., & Gallone, A. (2017, Sep). Administration of a polyphenol-enriched feed to farmed sea bass (Dicentrarchus labrax L.): Kidney melanomacrophages response. *Fish Shellfish Immunol, 68*, 404-410. https://doi.org/10.1016/j.fsi.2017.07.043

Arenas, M., & Bastolla, U. (2019, 12/05). ProtASR2: Ancestral Reconstruction of Protein Sequences accounting for Folding Stability. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.13341

Arenas, M., Sánchez-Cobos, A., & Bastolla, U. (2015, Aug). Maximum-Likelihood Phylogenetic Inference with Selection on Protein Folding Stability. *Mol Biol Evol, 32*(8), 2195-2207. https://doi.org/10.1093/molbev/msv085

Arenas, M., Weber, C. C., Liberles, D. A., & Bastolla, U. (2017, Nov 1). ProtASR: An Evolutionary Framework for Ancestral Protein Reconstruction with Selection on Folding Stability. *Syst Biol, 66*(6), 1054-1064. https://doi.org/10.1093/sysbio/syw121

Arnesen, S. M., Schroder, M. B., Dalmo, R. A., & Bogwald, J. (2002, Aug). Antigen uptake and immunoglobulin production in Atlantic cod (Gadus morhua L.) after intraperitoneal injection of Vibrio anguillarum. *Fish Shellfish Immunol, 13*(2), 159-170. http://www.ncbi.nlm.nih.gov/pubmed/12400865

Averill, B. a. E., P. (2011). General Chemistry: Principles, Patterns, and Applications.

Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., Huelsenbeck, J. P., Ronquist, F., Swofford, D. L., Cummings, M. P., Rambaut, A., & Suchard, M. A. (2012, Jan). BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst Biol, 61*(1), 170-173. https://doi.org/10.1093/sysbio/syr100

Aziz, M., & Iheanacho, F. (2019). Physiology, Antibody. In *StatPearls*. https://www.ncbi.nlm.nih.gov/pubmed/31536276

Babbage, G., Garand, R., Robillard, N., Zojer, N., Stevenson, F. K., & Sahota, S. S. (2004, Apr 01). Mantle cell lymphoma with t(11;14) and unmutated or mutated VH genes expresses AID and undergoes isotype switch events. *Blood, 103*(7), 2795-2798. https://doi.org/10.1182/blood-2003-05-1632

Babkova, P., Dunajova, Z., Chaloupkova, R., Damborsky, J., Bednar, D., & Marek, M. (2020, 2020/01/01/). Structures of hyperstable ancestral haloalkane dehalogenases show restricted conformational dynamics. *Computational and Structural Biotechnology Journal, 18*, 1497-1508. https://doi.org/https://doi.org/10.1016/j.csbj.2020.06.021

Barik, S. (2016). What Really Rigs Up RIG-I? *J Innate Immun, 8*(5), 429-436. https://doi.org/10.1159/000447947

Barreto, V. M., & Magor, B. G. (2011, Sep). Activation-induced cytidine deaminase structure and functions: a species comparative view. *Dev Comp Immunol, 35*(9), 991-1007. https://doi.org/10.1016/j.dci.2011.02.005

Barreto, V. M., Pan-Hammarstrom, Q., Zhao, Y., Hammarstrom, L., Misulovin, Z., & Nussenzweig, M. C. (2005, Sep 19). AID from bony fish catalyzes class switch recombination. *J Exp Med, 202*(6), 733-738. https://doi.org/10.1084/jem.20051378

Bascove, M., & Frippiat, J. P. (2010, Apr). Molecular characterization of Pleurodeles waltl activation-induced cytidine deaminase. *Mol Immunol, 47*(7-8), 1640-1649. https://doi.org/10.1016/j.molimm.2010.01.005

Bastianello, G., & Arakawa, H. (2017, Jan 9). A double-strand break can trigger immunoglobulin gene conversion. *Nucleic Acids Res, 45*(1), 231-243. https://doi.org/10.1093/nar/gkw887

Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J. P., Manis, J. P., & Alt, F. W. (2005, Nov 24). The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature, 438*(7067), 508-511. https://doi.org/10.1038/nature04255

Basu, U., Wang, Y., & Alt, F. W. (2008, 2008/10/24/). Evolution of Phosphorylation-Dependent Regulation of Activation-Induced Cytidine Deaminase. *Mol Cell, 32*(2), 285-291. https://doi.org/https://doi.org/10.1016/j.molcel.2008.08.019

Baumgarth, N. (2011, Jan). The double life of a B-1 cell: self-reactivity selects for protective effector functions. *Nat Rev Immunol, 11*(1), 34-46. https://doi.org/10.1038/nri2901

Berg, J. M., Stryer, L., & Tymoczko, J. L. (2002). *Biochemistry. 5th edition*. W. H. Freeman 2002. http://lib.ugent.be/catalog/ebk01:3450000000002008

Betz, A. G., Milstein, C., González-Fernández, A., Pannell, R., Larson, T., & Neuberger, M. S. (1994, Apr 22). Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. *Cell, 77*(2), 239-248. https://doi.org/10.1016/0092-8674(94)90316-6

Betz, A. G., Rada, C., Pannell, R., Milstein, C., & Neuberger, M. S. (1993, Mar 15). Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. *Proc Natl Acad Sci U S A, 90*(6), 2385-2388. http://www.ncbi.nlm.nih.gov/pubmed/8460148

Bhutani, N., Decker, M. N., Brady, J. J., Bussat, R. T., Burns, D. M., Corbel, S. Y., & Blau, H. M. (2013, Mar). A critical role for AID in the initiation of reprogramming to induced pluripotent stem cells. *FASEB J, 27*(3), 1107-1113. https://doi.org/10.1096/fj.12-222125

Boehm, T., Hess, I., & Swann, J. B. (2012, Jun). Evolution of lymphoid tissues. *Trends Immunol, 33*(6), 315-321. https://doi.org/10.1016/j.it.2012.02.005

Bohn, M. F., Shandilya, S. M., Albin, J. S., Kouno, T., Anderson, B. D., McDougle, R. M., Carpenter, M. A., Rathore, A., Evans, L., Davis, A. N., Zhang, J., Lu, Y., Somasundaran, M., Matsuo, H., Harris, R. S., & Schiffer, C. A. (2013, Jun 04). Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. *Structure, 21*(6), 1042-1050. https://doi.org/10.1016/j.str.2013.04.010

Bransteitter, R., Pham, P., Scharff, M. D., & Goodman, M. F. (2003, Apr 01). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci U S A, 100*(7), 4102-4107. https://doi.org/10.1073/pnas.0730835100

Bransteitter, R., Sneeden, J. L., Allen, S., Pham, P., & Goodman, M. F. (2006, Jun 23). First AID (activation-induced cytidine deaminase) is needed to produce high affinity isotype-switched antibodies. *J Biol Chem, 281*(25), 16833-16836. https://doi.org/10.1074/jbc.R600006200

Branton, S. A., Ghorbani, A., Bolt, B. N., Fifield, H., Berghuis, L. M., & Larijani, M. (2020, Jul). Activation-induced cytidine deaminase can target multiple topologies of double-stranded DNA in a transcription-independent manner. *FASEB J, 34*(7), 9245-9268. https://doi.org/10.1096/fj.201903036RR

Brar, S. S., Watson, M., & Diaz, M. (2004, Jun 18). Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. *J Biol Chem, 279*(25), 26395-26401. https://doi.org/10.1074/jbc.M403503200

Briney, B. S., & Crowe, J. E., Jr. (2013). Secondary mechanisms of diversification in the human antibody repertoire. *Front Immunol, 4*, 42. https://doi.org/10.3389/fimmu.2013.00042

Bromage, E. S., Ye, J., & Kaattari, S. L. (2006, Jan). Antibody structural variation in rainbow trout fluids. *Comp Biochem Physiol B Biochem Mol Biol, 143*(1), 61-69. https://doi.org/10.1016/j.cbpb.2005.10.003

Brown, G. D., Willment, J. A., & Whitehead, L. (2018, Jun). C-type lectins in immunity and homeostasis. *Nat Rev Immunol, 18*(6), 374-389. https://doi.org/10.1038/s41577-018-0004-8

Buerstedde, J. M., & Arakawa, H. (2006). Immunoglobulin gene conversion or hypermutation: that's the question. *Subcell Biochem, 40*, 11-24. https://doi.org/10.1007/978-1-4020-4896-8_2

Buonocore, F., & Gerdol, M. (2016, Jan). Alternative adaptive immunity strategies: coelacanth, cod and shark immunity. *Mol Immunol, 69*, 157-169. https://doi.org/10.1016/j.molimm.2015.09.003

Burns, A., Alsolami, R., Becq, J., Timbs, A., Bruce, D., Robbe, P., Vavoulis, D., Cabes, M., Dreau, H., Taylor, J., Knight, S. J. L., Mansson, R., Bentley, D., Beekman, R., Martin-Subero, J. I., Campo, E., Houlston, R. S., Ridout, K. E., & Schuh, A. (2017, Jun 06). Whole-genome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHVmut and IgHVunmut subgroups. *Leukemia*. https://doi.org/10.1038/leu.2017.177

Byeon, I. J., Ahn, J., Mitra, M., Byeon, C. H., Hercik, K., Hritz, J., Charlton, L. M., Levin, J. G., & Gronenborn, A. M. (2013). NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nat Commun, 4*, 1890. https://doi.org/10.1038/ncomms2883

Cain, K. D., Jones, D. R., & Raison, R. L. (2002, Mar). Antibody-antigen kinetics following immunization of rainbow trout (Oncorhynchus mykiss) with a T-cell dependent antigen. *Dev Comp Immunol, 26*(2), 181-190. http://www.ncbi.nlm.nih.gov/pubmed/11696383

Caipang, C. M., Brinchmann, M. F., & Kiron, V. (2009, Jul). Profiling gene expression in the spleen of Atlantic cod, Gadus morhua upon vaccination with Vibrio anguillarum antigen. *Comp Biochem Physiol B Biochem Mol Biol, 153*(3), 261-267. https://doi.org/10.1016/j.cbpb.2009.03.005

Cao, Y. (2018). *Advances in Membrane Proteins: Part I: Mass Processing and Transportation*. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-0532-0

Carletti, M. S., Monzon, A. M., Garcia-Rios, E., Benitez, G., Hirsh, L., Fornasari, M. S., & Parisi, G. (2020, Jan 1). Revenant: a database of resurrected proteins. *Database (Oxford), 2020*. https://doi.org/10.1093/database/baaa031

Carpenter, M. A., Rajagurubandara, E., Wijesinghe, P., & Bhagwat, A. S. (2010, May 04). Determinants of sequence-specificity within human AID and APOBEC3G. *DNA Repair (Amst), 9*(5), 579-587. https://doi.org/10.1016/j.dnarep.2010.02.010

Catchen, J. M., Conery, J. S., & Postlethwait, J. H. (2009, Aug). Automated identification of conserved synteny after whole-genome duplication. *Genome Res, 19*(8), 1497-1505. https://doi.org/10.1101/gr.090480.108

Chakravarty, S., Ung, A. R., Moore, B., Shore, J., & Alshamrani, M. (2018, Mar 27). A Comprehensive Analysis of Anion-Quadrupole Interactions in Protein Structures. *Biochemistry, 57*(12), 1852-1867. https://doi.org/10.1021/acs.biochem.7b01006

Chandra, V., Bortnick, A., & Murre, C. (2015, Sep). AID targeting: old mysteries and new challenges. *Trends Immunol, 36*(9), 527-535. https://doi.org/10.1016/j.it.2015.07.003

Chang, B. S. (2003, Aug). Ancestral gene reconstruction and synthesis of ancient rhodopsins in the laboratory. *Integr Comp Biol, 43*(4), 500-507. https://doi.org/10.1093/icb/43.4.500

Chang, B. S., Jönsson, K., Kazmi, M. A., Donoghue, M. J., & Sakmar, T. P. (2002, Sep). Recreating a functional ancestral archosaur visual pigment. *Mol Biol Evol, 19*(9), 1483-1489. https://doi.org/10.1093/oxfordjournals.molbev.a004211

Chaudhuri, J., & Alt, F. W. (2004, Jul). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol, 4*(7), 541-552. https://doi.org/10.1038/nri1395

Chaudhuri, J., Khuong, C., & Alt, F. W. (2004, Aug 26). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature, 430*(7003), 992-998. https://doi.org/10.1038/nature02821

Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., & Alt, F. W. (2003, Apr 17). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature, 422*(6933), 726-730. https://doi.org/10.1038/nature01574

Chi, X., Li, Y., & Qiu, X. (2020, Feb 7). V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology*. https://doi.org/10.1111/imm.13176

Choi, B., Rempala, G., & Kim, J. K. (2017, 12/05). Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Sci Rep, 7*. https://doi.org/10.1038/s41598-017-17072-z

Choudhary, M., Tamrakar, A., Singh, A. K., Jain, M., Jaiswal, A., & Kodgire, P. (2017, Sep 21). AID Biology: A pathological and clinical perspective. *Int Rev Immunol*, 1-20. https://doi.org/10.1080/08830185.2017.1369980

Choudhary, M., Tamrakar, A., Singh, A. K., Jain, M., Jaiswal, A., & Kodgire, P. (2018, Jan 2). AID Biology: A pathological and clinical perspective. *Int Rev Immunol, 37*(1), 37-56. https://doi.org/10.1080/08830185.2017.1369980

Colonna, M. (2018, Jun 19). Innate Lymphoid Cells: Diversity, Plasticity, and Unique Functions in Immunity. *Immunity, 48*(6), 1104-1117. https://doi.org/10.1016/j.immuni.2018.05.013

Conticello, S. G. (2008). The AID/APOBEC family of nucleic acid mutators. *Genome Biol, 9*(6), 229. https://doi.org/10.1186/gb-2008-9-6-229

Conticello, S. G., Ganesh, K., Xue, K., Lu, M., Rada, C., & Neuberger, M. S. (2008, Aug 22). Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. *Mol Cell, 31*(4), 474-484. https://doi.org/10.1016/j.molcel.2008.07.009

Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K., & Neuberger, M. S. (2005, Feb). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol, 22*(2), 367-377. https://doi.org/10.1093/molbev/msi026

Corripio-Miyar, Y., Mazorra de Quero, C., Treasurer, J. W., Ford, L., Smith, P. D., & Secombes, C. J. (2007, Jul 15). Vaccination experiments in the gadoid haddock, Melanogrammus aeglefinus L., against the bacterial pathogen Vibrio anguillarum. *Vet Immunol Immunopathol, 118*(1-2), 147-153. https://doi.org/10.1016/j.vetimm.2007.04.011

Coutinho, A., Gronowicz, E., Bullock, W. W., & Möller, G. (1974, Jan 1). Mechanism of thymus-independent immunocyte triggering. Mitogenic activation of B cells results in specific immune responses. *J Exp Med, 139*(1), 74-92. https://doi.org/10.1084/jem.139.1.74

Covello, J. M., Bird, S., Morrison, R. N., Bridle, A. R., Battaglene, S. C., Secombes, C. J., & Nowak, B. F. (2013, Mar). Isolation of RAG-1 and IgM transcripts from the striped trumpeter (Latris lineata), and their expression as markers for development of the adaptive

immune response. *Fish Shellfish Immunol, 34*(3), 778-788. https://doi.org/10.1016/j.fsi.2012.12.015

Creacy, S. D., Routh, E. D., Iwamoto, F., Nagamine, Y., Akman, S. A., & Vaughn, J. P. (2008, Dec 12). G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J Biol Chem, 283*(50), 34626-34634. https://doi.org/10.1074/jbc.M806277200

Dancyger, A. M., King, J. J., Quinlan, M. J., Fifield, H., Tucker, S., Saunders, H. L., Berru, M., Magor, B. G., Martin, A., & Larijani, M. (2012, Apr). Differences in the enzymatic efficiency of human and bony fish AID are mediated by a single residue in the C terminus modulating single-stranded DNA binding. *FASEB J, 26*(4), 1517-1525. https://doi.org/10.1096/fj.11-198135

Davidson, G. A., Lin, S. H., Secombes, C. J., & Ellis, A. E. (1997, Sep 19). Detection of specific and 'constitutive' antibody secreting cells in the gills, head kidney and peripheral blood leucocytes of dab (Limanda limanda). *Vet Immunol Immunopathol, 58*(3-4), 363-374. http://www.ncbi.nlm.nih.gov/pubmed/9436279

de Yebenes, V. G., & Ramiro, A. R. (2006, Sep). Activation-induced deaminase: light and dark sides. *Trends Mol Med, 12*(9), 432-439. https://doi.org/10.1016/j.molmed.2006.07.001

DeFranco, A. L. (2016). The germinal center antibody response in health and disease. *F1000Res, 5*. https://doi.org/10.12688/f1000research.7717.1

Detanico, T., Phillips, M., & Wysocki, L. J. (2016). Functional Versatility of AGY Serine Codons in Immunoglobulin Variable Region Genes. *Front Immunol, 7*, 525. https://doi.org/10.3389/fimmu.2016.00525

Di Noia, J. M., & Neuberger, M. S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem, 76*, 1-22. https://doi.org/10.1146/annurev.biochem.76.061705.090740

Diaz-Satizabal, L., & Magor, B. G. (2015, Jan). Isolation and cytochemical characterization of melanomacrophages and melanomacrophage clusters from goldfish (Carassius auratus, L.). *Dev Comp Immunol, 48*(1), 221-228. https://doi.org/10.1016/j.dci.2014.10.003

Diaz, M., Velez, J., Singh, M., Cerny, J., & Flajnik, M. F. (1999, May). Mutational pattern of the nurse shark antigen receptor gene (NAR) is similar to that of mammalian Ig genes and to spontaneous mutations in evolution: the translesion synthesis model of somatic

hypermutation. *Int Immunol, 11*(5), 825-833. http://www.ncbi.nlm.nih.gov/pubmed/10330287

Dickinson, G. S., Akkoyunlu, M., Bram, R. J., & Alugupalli, K. R. (2015, Dec). BAFF receptor and TACI in B-1b cell maintenance and antibacterial responses. *Ann N Y Acad Sci, 1362*, 57-67. https://doi.org/10.1111/nyas.12772

Doan, T. (2013). Lippincott's Illustrated Reviews: Immunology. http://meded.lwwhealthlibrary.com/book.aspx?bookid=777

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., & Baker, N. A. (2004, Jul 1). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res, 32*(Web Server issue), W665-667. https://doi.org/10.1093/nar/gkh381

Dominguez, P. M., Teater, M., Chambwe, N., Kormaksson, M., Redmond, D., Ishii, J., Vuong, B., Chaudhuri, J., Melnick, A., Vasanthakumar, A., Godley, L. A., Papavasiliou, F. N., Elemento, O., & Shaknovich, R. (2015, Sep 29). DNA Methylation Dynamics of Germinal Center B Cells Are Mediated by AID. *Cell Rep, 12*(12), 2086-2098. https://doi.org/10.1016/j.celrep.2015.08.036

Dooley, H., & Flajnik, M. F. (2005, Mar). Shark immunity bites back: affinity maturation and memory response in the nurse shark, Ginglymostoma cirratum. *Eur J Immunol, 35*(3), 936-945. https://doi.org/10.1002/eji.200425760

Dooley, H., Stanfield, R. L., Brady, R. A., & Flajnik, M. F. (2006, Feb 07). First molecular and biochemical analysis of in vivo affinity maturation in an ectothermic vertebrate. *Proc Natl Acad Sci U S A, 103*(6), 1846-1851. https://doi.org/10.1073/pnas.0508341103

Duchêne, S., & Lanfear, R. (2015, Sep). Phylogenetic uncertainty can bias the number of evolutionary transitions estimated from ancestral state reconstruction methods. *J Exp Zool B Mol Dev Evol, 324*(6), 517-524. https://doi.org/10.1002/jez.b.22638

Eibel, H., Kraus, H., Sic, H., Kienzler, A. K., & Rizzi, M. (2014, May). B cell biology: an overview. *Curr Allergy Asthma Rep, 14*(5), 434. https://doi.org/10.1007/s11882-014-0434-8

Eick, G. N., Bridgham, J. T., Anderson, D. P., Harms, M. J., & Thornton, J. W. (2017, Feb 1). Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. *Mol Biol Evol, 34*(2), 247-261. https://doi.org/10.1093/molbev/msw223

Ellingsen, T., Inami, M., Gjessing, M. C., Van Nieuwenhove, K., Larsen, R., Seppola, M., Lund, V., & Schroder, M. B. (2011, Aug). Francisella noatunensis in Atlantic cod (Gadus morhua L.); waterborne transmission and immune responses. *Fish Shellfish Immunol, 31*(2), 326-333. https://doi.org/10.1016/j.fsi.2011.05.021

Ellis, A. E. (1980, 1980/09/01). Antigen-trapping in the spleen and kidney of the plaice Pleuronectes platessa L [https://doi.org/10.1111/j.1365-2761.1980.tb00425.x]. *J Fish Dis, 3*(5), 413-426. https://doi.org/https://doi.org/10.1111/j.1365-2761.1980.tb00425.x

Emma M. Quinlan, J. J. K., Chris T. Amemiya, Ellen Hsu and Mani Larijani. (2017). Biochemical regulatory features of AID remain conserved from lamprey to humans. *Mol Cell Biol, 37*(15). https://doi.org/10.1128/MCB.00077-17

Eslamloo, K., Ghorbani, A., Xue, X., Inkpen, S. M., Larijani, M., & Rise, M. L. (2019). Characterization and Transcript Expression Analyses of Atlantic Cod Viperin. *Front Immunol, 10*, 311. https://doi.org/10.3389/fimmu.2019.00311

Eslamloo, K., Inkpen, S. M., Rise, M. L., & Andreassen, R. (2018, Jan). Discovery of microRNAs associated with the antiviral immune response of Atlantic cod macrophages. *Mol Immunol, 93*, 152-161. https://doi.org/10.1016/j.molimm.2017.11.015

Eslamloo, K., Xue, X., Booman, M., Smith, N. C., & Rise, M. L. (2016, Oct). Transcriptome profiling of the antiviral immune response in Atlantic cod macrophages. *Dev Comp Immunol, 63*, 187-205. https://doi.org/10.1016/j.dci.2016.05.021

Espelid, S., Rodseth, O. M., & Jorgensent, Ø. (1991). Vaccination experiments and studies of the humoral immune responses in cod, Gadus morhua L., to four strains of monoclonal-defined Vibrio anguillarum.

Faili, A., Stary, A., Delbos, F., Weller, S., Aoufouchi, S., Sarasin, A., Weill, J. C., & Reynaud, C. A. (2009, May 15). A backup role of DNA polymerase kappa in Ig gene hypermutation only takes place in the complete absence of DNA polymerase eta. *J Immunol, 182*(10), 6353-6359. https://doi.org/10.4049/jimmunol.0900177

Fields, P. A., Dong, Y., Meng, X., & Somero, G. N. (2015, Jun). Adaptations of protein structure and function to temperature: there is more than one way to 'skin a cat'. *J Exp Biol, 218*(Pt 12), 1801-1811. https://doi.org/10.1242/jeb.114298

Frenette, A. P., Eydal, M., Hansen, H., Burt, M. D., & Duffy, M. S. (2017, Jan). Integrative Approach for the Reliable Detection and Specific Identification of the Microsporidium Loma morhua in Atlantic Cod (Gadus morhua). *J Eukaryot Microbiol, 64*(1), 67-77. https://doi.org/10.1111/jeu.12339

Frieder, D., Larijani, M., Tang, E., Parsa, J. Y., Basit, W., & Martin, A. (2006). Antibody diversification: mutational mechanisms and oncogenesis. *Immunol Res, 35*(1-2), 75-88. https://doi.org/10.1385/IR:35:1:75

Fu, X., Zhang, F., Watabe, S., & Asakawa, S. (2017, Jan 18). Immunoglobulin light chain (IGL) genes in torafugu: Genomic organization and identification of a third teleost IGL isotype. *Sci Rep, 7*, 40416. https://doi.org/10.1038/srep40416

Fu, X., Zhang, H., Tan, E., Watabe, S., & Asakawa, S. (2015, Mar). Characterization of the torafugu (Takifugu rubripes) immunoglobulin heavy chain gene locus. *Immunogenetics, 67*(3), 179-193. https://doi.org/10.1007/s00251-014-0824-z

Fugmann, S. D., & Schatz, D. G. (2003, May). RNA AIDS DNA. *Nat Immunol, 4*(5), 429-430. https://doi.org/10.1038/ni0503-429

Fukita, Y., Jacobs, H., & Rajewsky, K. (1998, Jul). Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity, 9*(1), 105-114. https://doi.org/10.1016/s1074-7613(00)80592-0

Furukawa, R., Toma, W., Yamazaki, K., & Akanuma, S. (2020, Sep 23). Ancestral sequence reconstruction produces thermally stable enzymes with mesophilic enzyme-like catalytic properties. *Sci Rep, 10*(1), 15493. https://doi.org/10.1038/s41598-020-72418-4

Futoma-Kołoch, B. (2016, 06/24). Immune Response against Bacterial Lipopolysaccharide. *Journal of Molecular Immunology, 1*. https://doi.org/10.4172/jmi.1000e106

Gajula, K. S., Huwe, P. J., Mo, C. Y., Crawford, D. J., Stivers, J. T., Radhakrishnan, R., & Kohli, R. M. (2014, Sep). High-throughput mutagenesis reveals functional determinants for DNA targeting by activation-induced deaminase. *Nucleic Acids Res, 42*(15), 9964-9975. https://doi.org/10.1093/nar/gku689

Gars, E., Butzmann, A., Ohgami, R., Balakrishna, J. P., & O'Malley, D. P. (2019, Nov 13). The life and death of the germinal center. *Ann Diagn Pathol, 44*, 151421. https://doi.org/10.1016/j.anndiagpath.2019.151421

Gaucher, E. A., Thomson, J. M., Burgan, M. F., & Benner, S. A. (2003, Sep 18). Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature, 425*(6955), 285-288. https://doi.org/10.1038/nature01977

Ghosn, E., Yoshimoto, M., Nakauchi, H., Weissman, I. L., & Herzenberg, L. A. (2019, Aug 1). Hematopoietic stem cell-independent hematopoiesis and the origins of innate-like B lymphocytes. *Development, 146*(15). https://doi.org/10.1242/dev.170571

Giudicelli, V., Chaume, D., & Lefranc, M. P. (2005, Jan 1). IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res, 33*(Database issue), D256-261. https://doi.org/10.1093/nar/gki010

Glasauer, S. M., & Neuhauss, S. C. (2014, Dec). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics, 289*(6), 1045-1060. https://doi.org/10.1007/s00438-014-0889-2

Golub, R., & Charlemagne, J. (1998, Feb 01). Structure, diversity, and repertoire of VH families in the Mexican axolotl. *J Immunol, 160*(3), 1233-1239. http://www.ncbi.nlm.nih.gov/pubmed/9570539

Gomez-Fernandez, B. J., Risso, V. A., Rueda, A., Sanchez-Ruiz, J. M., & Alcalde, M. (2020, Jul 2). Ancestral Resurrection and Directed Evolution of Fungal Mesozoic Laccases. *Appl Environ Microbiol, 86*(14). https://doi.org/10.1128/aem.00778-20

Good-Jacobson, K. L. (2018, Jul). Strength in diversity: Phenotypic, functional, and molecular heterogeneity within the memory B cell repertoire. *Immunol Rev, 284*(1), 67-78. https://doi.org/10.1111/imr.12663

Goyal, S., Castrillon-Betancur, J. C., Klaile, E., & Slevogt, H. (2018). The Interaction of Human Pathogenic Fungi With C-Type Lectin Receptors. *Front Immunol, 9*, 1261. https://doi.org/10.3389/fimmu.2018.01261

Goyenechea, B., Klix, N., Yélamos, J., Williams, G. T., Riddell, A., Neuberger, M. S., & Milstein, C. (1997, Jul 1). Cells strongly expressing Ig(kappa) transgenes show clonal recruitment of hypermutation: a role for both MAR and the enhancers. *EMBO J, 16*(13), 3987-3994. https://doi.org/10.1093/emboj/16.13.3987

Granato, A., Chen, Y., & Wesemann, D. R. (2015, Apr). Primary immunoglobulin repertoire development: time and space matter. *Curr Opin Immunol, 33*, 126-131. https://doi.org/10.1016/j.coi.2015.02.011

Grasseau, A., Boudigou, M., Le Pottier, L., Chriti, N., Cornec, D., Pers, J. O., Renaudineau, Y., & Hillion, S. (2019, Jun 10). Innate B Cells: the Archetype of Protective Immune Cells. *Clin Rev Allergy Immunol*. https://doi.org/10.1007/s12016-019-08748-7

Greeve, J., Philipsen, A., Krause, K., Klapper, W., Heidorn, K., Castle, B. E., Janda, J., Marcu, K. B., & Parwaresch, R. (2003, May 01). Expression of activation-induced cytidine deaminase in human B-cell non-Hodgkin lymphomas. *Blood, 101*(9), 3574-3580. https://doi.org/10.1182/blood-2002-08-2424

Groussin, M., Hobbs, J. K., Szollosi, G. J., Gribaldo, S., Arcus, V. L., & Gouy, M. (2015, Jan). Toward more accurate ancestral protein genotype-phenotype reconstructions with the use of species tree-aware gene trees. *Mol Biol Evol, 32*(1), 13-22. https://doi.org/10.1093/molbev/msu305

Grove, S., Johansen, R., Dannevig, B. H., Reitan, L. J., & Ranheim, T. (2003, Feb 27). Experimental infection of Atlantic halibut Hippoglossus hippoglossus with nodavirus: tissue distribution and immune response. *Dis Aquat Organ, 53*(3), 211-221. https://doi.org/10.3354/dao053211

Gudmundsdottir, B. K., Bjornsdottir, B., Gudmundsdottir, S., & Bambir, S. H. (2006, Aug). A comparative study of susceptibility and induced pathology of cod, Gadus morhua (L.), and halibut, Hippoglossus hippoglossus (L.), following experimental infection with Moritella viscosa. *J Fish Dis, 29*(8), 481-487. https://doi.org/10.1111/j.1365-2761.2006.00741.x

Gudmundsdóttir, S., Magnadóttir, B., Björnsdóttir, B., Arnadóttir, H., & Gudmundsdóttir, B. K. (2009, Apr). Specific and natural antibody response of cod juveniles vaccinated against Vibrio anguillarum. *Fish Shellfish Immunol, 26*(4), 619-624. https://doi.org/10.1016/j.fsi.2008.09.017

Gumulya, Y., & Gillam, E. M. (2017, Jan 1). Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro' approach to protein engineering. *Biochem J, 474*(1), 1-19. https://doi.org/10.1042/bcj20160507

Guo, Y., Bao, Y., Meng, Q., Hu, X., Meng, Q., Ren, L., Li, N., & Zhao, Y. (2012). Immunoglobulin genomics in the guinea pig (Cavia porcellus). *PLoS One, 7*(6), e39298. https://doi.org/10.1371/journal.pone.0039298

Haakenson, J. K., Huang, R., & Smider, V. V. (2018). Diversity in the Cow Ultralong CDR H3 Antibody Repertoire. *Front Immunol, 9*, 1262. https://doi.org/10.3389/fimmu.2018.01262

Haas, K. M. (2015, Dec). B-1 lymphocytes in mice and nonhuman primates. *Ann N Y Acad Sci, 1362*, 98-109. https://doi.org/10.1111/nyas.12760

Habib, O., Habib, G., Do, J. T., Moon, S. H., & Chung, H. M. (2014, Feb 01). Activation-induced deaminase-coupled DNA demethylation is not crucial for the generation of induced pluripotent stem cells. *Stem Cells Dev, 23*(3), 209-218. https://doi.org/10.1089/scd.2013.0337

Hackney, J. A., Misaghi, S., Senger, K., Garris, C., Sun, Y., Lorenzo, M. N., & Zarrin, A. A. (2009). DNA targets of AID evolutionary link between antibody somatic hypermutation and class switch recombination. *Adv Immunol, 101*, 163-189. https://doi.org/10.1016/S0065-2776(08)01005-5

Hall, T. E., Smith, P., & Johnston, I. A. (2004, Mar). Stages of embryonic development in the Atlantic cod Gadus morhua. *J Morphol, 259*(3), 255-270. https://doi.org/10.1002/jmor.10222

Harjes, S., Solomon, W. C., Li, M., Chen, K. M., Harjes, E., Harris, R. S., & Matsuo, H. (2013, Jun). Impact of H216 on the DNA binding and catalytic activities of the HIV restriction factor APOBEC3G. *J Virol, 87*(12), 7008-7014. https://doi.org/10.1128/JVI.03173-12

Harms, M. J., & Thornton, J. W. (2010, Jun). Analyzing protein structure and function using ancestral gene reconstruction. *Curr Opin Struct Biol, 20*(3), 360-366. https://doi.org/10.1016/j.sbi.2010.03.005

Harris, R. S., Petersen-Mahrt, S. K., & Neuberger, M. S. (2002, Nov). RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell, 10*(5), 1247-1253. http://www.ncbi.nlm.nih.gov/pubmed/12453430

Harwood, N. E., & Batista, F. D. (2010). Early events in B cell activation. *Annu Rev Immunol, 28*, 185-210. https://doi.org/10.1146/annurev-immunol-030409-101216

Häsler, J., Rada, C., & Neuberger, M. S. (2012, Aug). The cytoplasmic AID complex. *Semin Immunol, 24*(4), 273-280. https://doi.org/10.1016/j.smim.2012.05.004

Hayashi, F., Nagata, T., Nagashima, T., Muto, Y., Inoue, M., Kigawa, T., Yokoyama, S., RIKEN (2009). *Solution structure of the monomeric form of mouse APOBEC2*. http://www.rcsb.org/pdb/explore/explore.do?structureId=2RPZ.

Heineke, M. H., & van Egmond, M. (2017, Feb). Immunoglobulin A: magic bullet or Trojan horse? *Eur J Clin Invest, 47*(2), 184-192. https://doi.org/10.1111/eci.12716

Hendricks, J., Bos, N. A., & Kroese, F. G. M. (2018). Heterogeneity of Memory Marginal Zone B Cells. *Crit Rev Immunol, 38*(2), 145-158. https://doi.org/10.1615/CritRevImmunol.2018024985

Hillion, S., Arleevskaya, M. I., Blanco, P., Bordron, A., Brooks, W. H., Cesbron, J. Y., Kaveri, S., Vivier, E., & Renaudineau, Y. (2019, Jun 1). The Innate Part of the Adaptive Immune System. *Clin Rev Allergy Immunol*. https://doi.org/10.1007/s12016-019-08740-1

Hirbod-Mobarakeh, A., Aghamohammadi, A., & Rezaei, N. (2014, Jan). Immunoglobulin class switch recombination deficiency type 1 or CD40 ligand deficiency: from bedside to bench and back again. *Expert Rev Clin Immunol, 10*(1), 91-105. https://doi.org/10.1586/1744666X.2014.864554

Hogenbirk, M. A., Heideman, M. R., Velds, A., van den Berk, P. C., Kerkhoven, R. M., van Steensel, B., & Jacobs, H. (2013). Differential programming of B cells in AID deficient mice. *PLoS One, 8*(7), e69815. https://doi.org/10.1371/journal.pone.0069815

Holden, L. G., Prochnow, C., Chang, Y. P., Bransteitter, R., Chelico, L., Sen, U., Stevens, R. C., Goodman, M. F., & Chen, X. S. (2008, Nov 6). Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature, 456*(7218), 121-124. https://doi.org/10.1038/nature07357

Holinski, A., Heyn, K., Merkl, R., & Sterner, R. (2017, Feb). Combining ancestral sequence reconstruction with protein design to identify an interface hotspot in a key metabolic enzyme complex. *Proteins, 85*(2), 312-321. https://doi.org/10.1002/prot.25225

Hong, J. R. (2013, Feb 12). Betanodavirus: Mitochondrial disruption and necrotic cell death. *World J Virol, 2*(1), 1-5. https://doi.org/10.5501/wjv.v2.i1.1

Hori, T. S., Gamperl, A. K., Booman, M., Nash, G. W., & Rise, M. L. (2012, Aug 28). A moderate increase in ambient temperature modulates the Atlantic cod (Gadus morhua) spleen transcriptome response to intraperitoneal viral mimic injection. *BMC Genomics, 13*, 431. https://doi.org/10.1186/1471-2164-13-431

Hori, T. S., Gamperl, A. K., Nash, G., Booman, M., Barat, A., & Rise, M. L. (2013, Oct). The impact of a moderate chronic temperature increase on spleen immune-relevant gene transcription depends on whether Atlantic cod (Gadus morhua) are stimulated with bacterial versus viral antigens. *Genome, 56*(10), 567-576. https://doi.org/10.1139/gen-2013-0090

Hou, B., Saudan, P., Ott, G., Wheeler, M. L., Ji, M., Kuzmich, L., Lee, L. M., Coffman, R. L., Bachmann, M. F., & DeFranco, A. L. (2011, Mar 25). Selective utilization of Toll-like receptor and MyD88 signaling in B cells for enhancement of the antiviral germinal center response. *Immunity, 34*(3), 375-384. https://doi.org/10.1016/j.immuni.2011.01.011

Hou, Q., Bourgeas, R., Pucci, F., & Rooman, M. (2018, Oct 2). Computational analysis of the amino acid interactions that promote or decrease protein solubility. *Sci Rep, 8*(1), 14661. https://doi.org/10.1038/s41598-018-32988-w

Hou, S., Silvas, T. V., Leidner, F., Nalivaika, E. A., Matsuo, H., Kurt Yilmaz, N., & Schiffer, C. A. (2019, 2019/01/08). Structural Analysis of the Active Site and DNA Binding of Human Cytidine Deaminase APOBEC3B. *Journal of Chemical Theory and Computation, 15*(1), 637-647. https://doi.org/10.1021/acs.jctc.8b00545

Hsu, E. (1998, Apr). Mutation, selection, and memory in B lymphocytes of exothermic vertebrates. *Immunol Rev, 162*, 25-36. http://www.ncbi.nlm.nih.gov/pubmed/9602349

Hsu, E. (2016, May 01). Assembly and Expression of Shark Ig Genes. *J Immunol, 196*(9), 3517-3523. https://doi.org/10.4049/jimmunol.1600164

Hu, W., Begum, N. A., Mondal, S., Stanlie, A., & Honjo, T. (2015, May 5). Identification of DNA cleavage- and recombination-specific hnRNP cofactors for activation-induced cytidine deaminase. *Proc Natl Acad Sci U S A, 112*(18), 5791-5796. https://doi.org/10.1073/pnas.1506167112

Hu, Y., Ericsson, I., Torseth, K., Methot, S. P., Sundheim, O., Liabakk, N. B., Slupphaug, G., Di Noia, J. M., Krokan, H. E., & Kavli, B. (2013, Jan 23). A combined nuclear and nucleolar localization motif in activation-induced cytidine deaminase (AID) controls immunoglobulin class switching. *J Mol Biol, 425*(2), 424-443. https://doi.org/10.1016/j.jmb.2012.11.026

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics, 17*(8), 754-755. https://doi.org/10.1093/bioinformatics/17.8.754

Hwang, J. K., Wang, C., Du, Z., Meyers, R. M., Kepler, T. B., Neuberg, D., Kwong, P. D., Mascola, J. R., Joyce, M. G., Bonsignori, M., Haynes, B. F., Yeap, L. S., & Alt, F. W. (2017, Aug 8). Sequence intrinsic somatic mutation mechanisms contribute to affinity maturation of VRC01-class HIV-1 broadly neutralizing antibodies. *Proc Natl Acad Sci U S A, 114*(32), 8614-8619. https://doi.org/10.1073/pnas.1709203114

Ichikawa, H. T., Sowden, M. P., Torelli, A. T., Bachl, J., Huang, P., Dance, G. S., Marr, S. H., Robert, J., Wedekind, J. E., Smith, H. C., & Bottaro, A. (2006, Jul 01). Structural phylogenetic analysis of activation-induced deaminase function. *J Immunol, 177*(1), 355-361. http://www.ncbi.nlm.nih.gov/pubmed/16785531

Imkeller, K., & Wardemann, H. (2018, Jul). Assessing human B cell repertoire diversity and convergence. *Immunol Rev, 284*(1), 51-66. https://doi.org/10.1111/imr.12670

Ingles-Prieto, A., Ibarra-Molero, B., Delgado-Delgado, A., Perez-Jimenez, R., Fernandez, J. M., Gaucher, E. A., Sanchez-Ruiz, J. M., & Gavira, J. A. (2013, Sep 3). Conservation of protein structure over four billion years. *Structure, 21*(9), 1690-1697. https://doi.org/10.1016/j.str.2013.06.020

Inkpen, S. M., Hori, T. S., Gamperl, A. K., Nash, G. W., & Rise, M. L. (2015, May). Characterization and expression analyses of five interferon regulatory factor transcripts (Irf4a, Irf4b, Irf7, Irf8, Irf10) in Atlantic cod (Gadus morhua). *Fish Shellfish Immunol, 44*(1), 365-381. https://doi.org/10.1016/j.fsi.2015.02.032

Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., & Honjo, T. (2004, Feb 17). Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci U S A, 101*(7), 1975-1980. https://doi.org/10.1073/pnas.0307335101

Iyer, L. M., Zhang, D., Rogozin, I. B., & Aravind, L. (2011, Dec). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res, 39*(22), 9473-9497. https://doi.org/10.1093/nar/gkr691

Jain, A., & Pasare, C. (2017, May 15). Innate Control of Adaptive Immunity: Beyond the Three-Signal Paradigm. *J Immunol, 198*(10), 3791-3800. https://doi.org/10.4049/jimmunol.1602000

Jenne, C. N., Kennedy, L. J., McCullagh, P., & Reynolds, J. D. (2003, Apr 01). A new model of sheep Ig diversification: shifting the emphasis toward combinatorial mechanisms and away from hypermutation. *J Immunol, 170*(7), 3739-3750. http://www.ncbi.nlm.nih.gov/pubmed/12646640

Johnson, K., Reddy, K. L., & Singh, H. (2009). Molecular pathways and mechanisms regulating the recombination of immunoglobulin genes during B-lymphocyte development. *Adv Exp Med Biol, 650*, 133-147. https://doi.org/10.1007/978-1-4419-0296-2_11

Jolly, C. J., Wagner, S. D., Rada, C., Klix, N., Milstein, C., & Neuberger, M. S. (1996, Jun). The targeting of somatic hypermutation. *Semin Immunol, 8*(3), 159-168. https://doi.org/10.1006/smim.1996.0020

Joy, J. B., Liang, R. H., McCloskey, R. M., Nguyen, T., & Poon, A. F. Y. (2016). Ancestral Reconstruction. *PLOS Computational Biology, 12*(7), e1004763. https://doi.org/10.1371/journal.pcbi.1004763

Kaattari, S. L., Zhang, H. L., Khor, I. W., Kaattari, I. M., & Shapiro, D. A. (2002, Mar). Affinity maturation in trout: clonal dominance of high affinity antibodies late in the immune response. *Dev Comp Immunol, 26*(2), 191-200. http://www.ncbi.nlm.nih.gov/pubmed/11696384

Kasturi, S. P., Skountzou, I., Albrecht, R. A., Koutsonanos, D., Hua, T., Nakaya, H. I., Ravindran, R., Stewart, S., Alam, M., Kwissa, M., Villinger, F., Murthy, N., Steel, J., Jacob, J., Hogan, R. J., Garcia-Sastre, A., Compans, R., & Pulendran, B. (2011, Feb 24). Programming the magnitude and persistence of antibody responses with innate immunity. *Nature, 470*(7335), 543-547. https://doi.org/10.1038/nature09737

Kawai, T., & Akira, S. (2011, May 27). Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity, 34*(5), 637-650. https://doi.org/10.1016/j.immuni.2011.05.006

Kim, C., & Lee, B. (2007, Sep 20). Accuracy of structure-based sequence alignment of automatic methods. *BMC Bioinformatics, 8*, 355. https://doi.org/10.1186/1471-2105-8-355

Kim, N., & Jinks-Robertson, S. (2012, Feb 14). Transcription as a source of genome instability. *Nat Rev Genet, 13*(3), 204-214. https://doi.org/10.1038/nrg3152

Kim, Y. K., Shin, J. S., & Nahm, M. H. (2016, Jan). NOD-Like Receptors in Infection, Immunity, and Diseases. *Yonsei Med J, 57*(1), 5-14. https://doi.org/10.3349/ymj.2016.57.1.5

King, J. J., & Larijani, M. (2017). A Novel Regulator of Activation-Induced Cytidine Deaminase/APOBECs in Immunity and Cancer: Schrodinger's CATalytic Pocket. *Front Immunol, 8*, 351. https://doi.org/10.3389/fimmu.2017.00351

King, J. J., & Larijani, M. (2020, Oct 21). Structural plasticity of substrate selection by activation-induced cytidine deaminase as a regulator of its genome-wide mutagenic activity. *FEBS Lett*. https://doi.org/10.1002/1873-3468.13962

King, J. J., Manuel, C. A., Barrett, C. V., Raber, S., Lucas, H., Sutter, P., & Larijani, M. (2015, Apr 07). Catalytic pocket inaccessibility of activation-induced cytidine deaminase is a safeguard against excessive mutagenic activity. *Structure, 23*(4), 615-627. https://doi.org/10.1016/j.str.2015.01.016

Kitamura, S., Ode, H., Nakashima, M., Imahashi, M., Naganawa, Y., Kurosawa, T., Yokomaku, Y., Yamane, T., Watanabe, N., Suzuki, A., Sugiura, W., & Iwatani, Y. (2012, Oct). The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat Struct Mol Biol, 19*(10), 1005-1010. https://doi.org/10.1038/nsmb.2378

Klemm, L., Duy, C., Iacobucci, I., Kuchen, S., von Levetzow, G., Feldhahn, N., Henke, N., Li, Z., Hoffmann, T. K., Kim, Y. M., Hofmann, W. K., Jumaa, H., Groffen, J., Heisterkamp, N., Martinelli, G., Lieber, M. R., Casellas, R., & Muschen, M. (2009, Sep 8). The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. *Cancer Cell, 16*(3), 232-245. https://doi.org/10.1016/j.ccr.2009.07.030

Kmiec, Z., Cyman, M., & Slebioda, T. J. (2017, Mar). Cells of the innate and adaptive immunity and their interactions in inflammatory bowel disease. *Adv Med Sci, 62*(1), 1-16. https://doi.org/10.1016/j.advms.2016.09.001

Koblansky, A. A., Jankovic, D., Oh, H., Hieny, S., Sungnak, W., Mathur, R., Hayden, M. S., Akira, S., Sher, A., & Ghosh, S. (2013). Recognition of profilin by Toll-like receptor 12 is critical for host resistance to Toxoplasma gondii. *Immunity, 38*(1), 119-130. https://doi.org/10.1016/j.immuni.2012.09.016

Kodgire, P., Mukkawar, P., North, J. A., Poirier, M. G., & Storb, U. (2012, May). Nucleosome stability dramatically impacts the targeting of somatic hypermutation. *Mol Cell Biol, 32*(10), 2030-2040. https://doi.org/10.1128/mcb.06722-11

Kodgire, P., Mukkawar, P., Ratnam, S., Martin, T. E., & Storb, U. (2013, Jul 1). Changes in RNA polymerase II progression influence somatic hypermutation of Ig-related genes by AID. *J Exp Med, 210*(7), 1481-1492. https://doi.org/10.1084/jem.20121523

Kodinariya, T., & Makwana, P. R. (2013, 01/01). Review on Determining of Cluster in K-means Clustering. *International Journal of Advance Research in Computer Science and Management Studies, 1*, 90-95.

Kohli, R. M., Abrams, S. R., Gajula, K. S., Maul, R. W., Gearhart, P. J., & Stivers, J. T. (2009, Aug 21). A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *J Biol Chem, 284*(34), 22898-22904. https://doi.org/10.1074/jbc.M109.025536

Kojima, F., Frolov, A., Matnani, R., Woodward, J. G., & Crofford, L. J. (2013, Nov 15). Reduced T cell-dependent humoral immune response in microsomal prostaglandin E synthase-1 null mice is mediated by nonhematopoietic cells. *J Immunol, 191*(10), 4979-4988. https://doi.org/10.4049/jimmunol.1301942

Kolar, G. R., Mehta, D., Pelayo, R., & Capra, J. D. (2007, Mar 15). A novel human B cell subpopulation representing the initial germinal center population to express AID. *Blood, 109*(6), 2545-2552. https://doi.org/10.1182/blood-2006-07-037150

Kotas, M. E., & Locksley, R. M. (2018, Jun 19). Why Innate Lymphoid Cells? *Immunity, 48*(6), 1081-1090. https://doi.org/10.1016/j.immuni.2018.06.002

Kranich, J., & Krautler, N. J. (2016). How Follicular Dendritic Cells Shape the B-Cell Antigenome. *Front Immunol, 7*, 225. https://doi.org/10.3389/fimmu.2016.00225

Kumar, H., Kawai, T., & Akira, S. (2011, Feb). Pathogen recognition by the innate immune system. *Int Rev Immunol, 30*(1), 16-34. https://doi.org/10.3109/08830185.2010.529976

Kumar, K., Woo, S. M., Siu, T., Cortopassi, W. A., Duarte, F., & Paton, R. S. (2018, Mar 14). Cation-pi interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem Sci, 9*(10), 2655-2665. https://doi.org/10.1039/c7sc04905f

Kumar, R., DiMenna, L., Schrode, N., Liu, T. C., Franck, P., Munoz-Descalzo, S., Hadjantonakis, A. K., Zarrin, A. A., Chaudhuri, J., Elemento, O., & Evans, T. (2013, Aug 01). AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature, 500*(7460), 89-92. https://doi.org/10.1038/nature12299

Kumar, S., Ingle, H., Prasad, D. V., & Kumar, H. (2013, Aug). Recognition of bacterial infection by innate immune sensors. *Crit Rev Microbiol, 39*(3), 229-246. https://doi.org/10.3109/1040841X.2012.706249

Kumar Singh, A., Tamrakar, A., Jaiswal, A., Kanayama, N., Agarwal, A., Tripathi, P., & Kodgire, P. (2019, Dec). Splicing regulator SRSF1-3 that controls somatic hypermutation of IgV genes interacts with topoisomerase 1 and AID. *Mol Immunol, 116*, 63-72. https://doi.org/10.1016/j.molimm.2019.10.002

Kunimoto, H., McKenney, A. S., Meydan, C., Shank, K., Nazir, A., Rapaport, F., Durham, B., Garrett-Bakelman, F. E., Pronier, E., Shih, A. H., Melnick, A., Chaudhuri, J., & Levine, R. L. (2017, Jan 11). Aid is a key regulator of myeloid/erythroid differentiation and DNA methylation in hematopoietic stem/progenitor cells. *Blood*. https://doi.org/10.1182/blood-2016-06-721977

Kuraoka, M., Liao, D., Yang, K., Allgood, S. D., Levesque, M. C., Kelsoe, G., & Ueda, Y. (2009, Sep 01). Activation-induced cytidine deaminase expression and activity in the absence of germinal centers: insights into hyper-IgM syndrome. *J Immunol, 183*(5), 3237-3248. https://doi.org/10.4049/jimmunol.0901548

Lamers, C. H. J. (1986, 1986/04/01). Histophysiology of a primary immune response against Aeromonas hydrophila in carp (Cyprinus carpio L.) [https://doi.org/10.1002/jez.1402380109]. *Journal of Experimental Zoology, 238*(1), 71-80. https://doi.org/https://doi.org/10.1002/jez.1402380109

Larijani, M., Frieder, D., Basit, W., & Martin, A. (2005, Feb). The mutation spectrum of purified AID is similar to the mutability index in Ramos cells and in ung(-/-)msh2(-/-) mice. *Immunogenetics, 56*(11), 840-845. https://doi.org/10.1007/s00251-004-0748-0

Larijani, M., Frieder, D., Sonbuchner, T. M., Bransteitter, R., Goodman, M. F., Bouhassira, E. E., Scharff, M. D., & Martin, A. (2005, Mar). Methylation protects cytidines from AID-mediated deamination. *Mol Immunol, 42*(5), 599-604. https://doi.org/10.1016/j.molimm.2004.09.007

Larijani, M., & Martin, A. (2007, Dec). Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. *Mol Cell Biol, 27*(23), 8038-8048. https://doi.org/10.1128/MCB.01046-07

Larijani, M., & Martin, A. (2012, Aug). The biochemistry of activation-induced deaminase and its physiological functions. *Semin Immunol, 24*(4), 255-263. https://doi.org/10.1016/j.smim.2012.05.003

Larijani, M., Petrov, A. P., Kolenchenko, O., Berru, M., Krylov, S. N., & Martin, A. (2007, Jan). AID associates with single-stranded DNA with high affinity and a long complex half-

life in a sequence-independent manner. *Mol Cell Biol, 27*(1), 20-30. https://doi.org/10.1128/MCB.00824-06

Laursen, L., Čalyševa, J., Gibson, T. J., & Jemth, P. (2020). Divergent evolution of a protein-protein interaction revealed through ancestral sequence reconstruction and resurrection. *Mol Biol Evol*. https://doi.org/10.1093/molbev/msaa198

Lee, G. R. (2018, Mar 3). The Balance of Th17 versus Treg Cells in Autoimmunity. *Int J Mol Sci, 19*(3). https://doi.org/10.3390/ijms19030730

Lee, S. S., Tranchina, D., Ohta, Y., Flajnik, M. F., & Hsu, E. (2002, Apr). Hypermutation in shark immunoglobulin light chain genes results in contiguous substitutions. *Immunity, 16*(4), 571-582. http://www.ncbi.nlm.nih.gov/pubmed/11970880

Lefranc, M. P. (2001, Jan 1). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res, 29*(1), 207-209. http://www.ncbi.nlm.nih.gov/pubmed/11125093

Lefranc, M. P. (2003, Jan 1). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res, 31*(1), 307-310. http://www.ncbi.nlm.nih.gov/pubmed/12520009

Lefranc, M. P., Clement, O., Kaas, Q., Duprat, E., Chastellan, P., Coelho, I., Combres, K., Ginestoux, C., Giudicelli, V., Chaume, D., & Lefranc, G. (2005). IMGT-Choreography for immunogenetics and immunoinformatics. *In Silico Biol, 5*(1), 45-60. http://www.ncbi.nlm.nih.gov/pubmed/15972004

Lefranc, M. P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., Hadi-Saljoqi, S., Sasorith, S., Lefranc, G., & Kossida, S. (2015, Jan). IMGT(R), the international ImMunoGeneTics information system(R) 25 years on. *Nucleic Acids Res, 43*(Database issue), D413-422. https://doi.org/10.1093/nar/gku1056

Lefranc, M. P., Giudicelli, V., Ginestoux, C., Bodmer, J., Muller, W., Bontrop, R., Lemaitre, M., Malik, A., Barbie, V., & Chaume, D. (1999, Jan 1). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res, 27*(1), 209-212. http://www.ncbi.nlm.nih.gov/pubmed/9847182

Lefranc, M. P., Giudicelli, V., Ginestoux, C., Jabado-Michaloud, J., Folch, G., Bellahcene, F., Wu, Y., Gemrot, E., Brochet, X., Lane, J., Regnier, L., Ehrenmann, F., Lefranc, G., & Duroux, P. (2009, Jan). IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res, 37*(Database issue), D1006-1012. https://doi.org/10.1093/nar/gkn838

Lefranc, M. P., Giudicelli, V., Kaas, Q., Duprat, E., Jabado-Michaloud, J., Scaviner, D., Ginestoux, C., Clement, O., Chaume, D., & Lefranc, G. (2005, Jan 1). IMGT, the

international ImMunoGeneTics information system. *Nucleic Acids Res, 33*(Database issue), D593-597. https://doi.org/10.1093/nar/gki065

Leighton, P. A., Morales, J., Harriman, W. D., & Ching, K. H. (2018). V(D)J Rearrangement Is Dispensable for Producing CDR-H3 Sequence Diversity in a Gene Converting Species. *Front Immunol, 9*, 1317. https://doi.org/10.3389/fimmu.2018.01317

Leonard, B., Hart, S. N., Burns, M. B., Carpenter, M. A., Temiz, N. A., Rathore, A., Vogel, R. I., Nikas, J. B., Law, E. K., Brown, W. L., Li, Y., Zhang, Y., Maurer, M. J., Oberg, A. L., Cunningham, J. M., Shridhar, V., Bell, D. A., April, C., Bentley, D., Bibikova, M., Cheetham, R. K., Fan, J. B., Grocock, R., Humphray, S., Kingsbury, Z., Peden, J., Chien, J., Swisher, E. M., Hartmann, L. C., Kalli, K. R., Goode, E. L., Sicotte, H., Kaufmann, S. H., & Harris, R. S. (2013, Dec 15). APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Res, 73*(24), 7222-7231. https://doi.org/10.1158/0008-5472.CAN-13-1753

Lim, S. A., Bolin, E. R., & Marqusee, S. (2018, Sep 11). Tracing a protein's folding pathway over evolutionary time using ancestral sequence reconstruction and hydrogen exchange. *Elife, 7*. https://doi.org/10.7554/eLife.38369

Lindley, R. A., Humbert, P., Larner, C., Akmeemana, E. H., & Pendlebury, C. R. (2016, Sep). Association between targeted somatic mutation (TSM) signatures and HGS-OvCa progression. *Cancer Med, 5*(9), 2629-2640. https://doi.org/10.1002/cam4.825

Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003, Feb 15). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins, 50*(3), 437-450. https://doi.org/10.1002/prot.10286

Lucas, X., Bauza, A., Frontera, A., & Quinonero, D. (2016, Feb 1). A thorough anion-pi interaction study in biomolecules: on the importance of cooperativity effects. *Chem Sci, 7*(2), 1038-1050. https://doi.org/10.1039/c5sc01386k

Lund, V., Arnesen, J. A., Mikkelsen, H., Gravningen, K., Brown, L., & Schroder, M. B. (2008, Dec 09). Atypical furunculosis vaccines for Atlantic cod (Gadus morhua); vaccine efficacy and antibody responses. *Vaccine, 26*(52), 6791-6799. https://doi.org/10.1016/j.vaccine.2008.10.012

Lund, V., Bordal, S., Kjellsen, O., Mikkelsen, H., & Schroder, M. B. (2006). Comparison of antibody responses in Atlantic cod (Gadus morhua L.) to Aeromonas salmonicida and Vibrio anguillarum. *Dev Comp Immunol, 30*(12), 1145-1155. https://doi.org/10.1016/j.dci.2006.02.004

Lund, V., Bordal, S., & Schroder, M. B. (2007, Oct). Specificity and durability of antibody responses in Atlantic cod (Gadus morhua L.) immunised with Vibrio anguillarum O2b. *Fish Shellfish Immunol, 23*(4), 906-910. https://doi.org/10.1016/j.fsi.2007.04.006

Lundqvist, M. L., Middleton, D. L., Radford, C., Warr, G. W., & Magor, K. E. (2006). Immunoglobulins of the non-galliform birds: antibody expression and repertoire in the duck. *Dev Comp Immunol, 30*(1-2), 93-100. https://doi.org/10.1016/j.dci.2005.06.019

Magnadottir, B., Jonsdottir, H., Helgason, S., Bjornsson, B., Jorgensen, T. O., & Pilstrom, L. (1999, Feb). Humoral immune parameters in Atlantic cod (Gadus morhua L.) I. The effects of environmental temperature. *Comp Biochem Physiol B Biochem Mol Biol, 122*(2), 173-180. http://www.ncbi.nlm.nih.gov/pubmed/10327607

Magnadottir, B., Jonsdottir, H., Helgason, S., Bjornsson, B., Solem, S. T., & Pilstrom, L. (2001, Jan). Immune parameters of immunised cod (Gadus morhua L.). *Fish Shellfish Immunol, 11*(1), 75-89. https://doi.org/10.1006/fsim.2000.0296

Magor, B. G. (2015, Jul 31). Antibody Affinity Maturation in Fishes-Our Current Understanding. *Biology (Basel), 4*(3), 512-524. https://doi.org/10.3390/biology4030512

Mahdaviani, S. A., Hirbod-Mobarakeh, A., Wang, N., Aghamohammadi, A., Hammarström, L., Masjedi, M. R., Pan-Hammarström, Q., & Rezaei, N. (2012, Aug). Novel mutation of the activation-induced cytidine deaminase gene in a Tajik family: special review on hyper-immunoglobulin M syndrome. *Expert Rev Clin Immunol, 8*(6), 539-546. https://doi.org/10.1586/eci.12.46

Mai, T., Zan, H., Zhang, J., Hawkins, J. S., Xu, Z., & Casali, P. (2010, Nov 26). Estrogen receptors bind to and activate the HOXC4/HoxC4 promoter to potentiate HoxC4-mediated activation-induced cytosine deaminase induction, immunoglobulin class switch DNA recombination, and somatic hypermutation. *J Biol Chem, 285*(48), 37797-37810. https://doi.org/10.1074/jbc.M110.169086

Mak, C. H., Pham, P., Afif, S. A., & Goodman, M. F. (2013, Oct 11). A mathematical model for scanning and catalysis on single-stranded DNA, illustrated with activation-induced deoxycytidine deaminase. *J Biol Chem, 288*(41), 29786-29795. https://doi.org/10.1074/jbc.M113.506550

Malecek, K., Brandman, J., Brodsky, J. E., Ohta, Y., Flajnik, M. F., & Hsu, E. (2005, Dec 15). Somatic hypermutation and junctional diversification at Ig heavy chain loci in the nurse shark. *J Immunol, 175*(12), 8105-8115. http://www.ncbi.nlm.nih.gov/pubmed/16339548

Malmstrom, M., Jentoft, S., Gregers, T. F., & Jakobsen, K. S. (2013). Unraveling the evolution of the Atlantic cod's (Gadus morhua L.) alternative immune strategy. *PLoS One, 8*(9), e74004. https://doi.org/10.1371/journal.pone.0074004

Malmstrom, M., Matschiner, M., Torresen, O. K., Star, B., Snipen, L. G., Hansen, T. F., Baalsrud, H. T., Nederbragt, A. J., Hanel, R., Salzburger, W., Stenseth, N. C., Jakobsen, K. S., & Jentoft, S. (2016, Oct). Evolution of the immune system influences speciation rates in teleost fishes. *Nat Genet, 48*(10), 1204-1210. https://doi.org/10.1038/ng.3645

Malu, S., Malshetty, V., Francis, D., & Cortes, P. (2012, Dec). Role of non-homologous end joining in V(D)J recombination. *Immunol Res, 54*(1-3), 233-246. https://doi.org/10.1007/s12026-012-8329-z

Mandal, A., & Viswanathan, C. (2015, Jun). Natural killer cells: In health and disease. *Hematol Oncol Stem Cell Ther, 8*(2), 47-55. https://doi.org/10.1016/j.hemonc.2014.11.006

Mandler, R., Chu, C. C., Paul, W. E., Max, E. E., & Snapper, C. M. (1993, Nov 1). Interleukin 5 induces S mu-S gamma 1 DNA rearrangement in B cells activated with dextran-anti-IgD antibodies and interleukin 4: a three component model for Ig class switching. *J Exp Med, 178*(5), 1577-1586. https://doi.org/10.1084/jem.178.5.1577

Marianes, A. E., & Zimmerman, A. M. (2011, Feb). Targets of somatic hypermutation within immunoglobulin light chain genes in zebrafish. *Immunology, 132*(2), 240-255. https://doi.org/10.1111/j.1365-2567.2010.03358.x

Marino, D., Perkovic, M., Hain, A., Jaguva Vasudevan, A. A., Hofmann, H., Hanschmann, K. M., Muhlebach, M. D., Schumann, G. G., Konig, R., Cichutek, K., Haussinger, D., & Munk, C. (2016). APOBEC4 Enhances the Replication of HIV-1. *PLoS One, 11*(6), e0155422. https://doi.org/10.1371/journal.pone.0155422

Marr, S., Morales, H., Bottaro, A., Cooper, M., Flajnik, M., & Robert, J. (2007, Nov 15). Localization and differential expression of activation-induced cytidine deaminase in the amphibian Xenopus upon antigen stimulation and during early development. *J Immunol, 179*(10), 6783-6789. http://www.ncbi.nlm.nih.gov/pubmed/17982068

Marshall, C. J. (1997, Sep). Cold-adapted enzymes. *Trends Biotechnol, 15*(9), 359-364. https://doi.org/10.1016/s0167-7799(97)01086-x

Martinez, P. (2018, 2018-August-28). The Comparative Method in Biology and the Essentialist Trap [Perspective]. *Frontiers in Ecology and Evolution, 6*(130). https://doi.org/10.3389/fevo.2018.00130

Maslo, B., Gignoux-Wolfsohn, S. A., & Fefferman, N. H. (2017). Success of Wildlife Disease Treatment Depends on Host Immune Response [10.3389/fevo.2017.00028].

Frontiers in Ecology and Evolution, 5, 28. https://www.frontiersin.org/article/10.3389/fevo.2017.00028

Maul, R. W., & Gearhart, P. J. (2010). Chapter six - AID and Somatic Hypermutation. In F. W. Alt (Ed.), *Adv Immunol* (Vol. 105, pp. 159-191). Academic Press. https://doi.org/https://doi.org/10.1016/S0065-2776(10)05006-6

Maul, R. W., MacCarthy, T., Frank, E. G., Donigan, K. A., McLenigan, M. P., Yang, W., Saribasak, H., Huston, D. E., Lange, S. S., Woodgate, R., & Gearhart, P. J. (2016, Aug 22). DNA polymerase ι functions in the generation of tandem mutations during somatic hypermutation of antibody genes. *J Exp Med, 213*(9), 1675-1683. https://doi.org/10.1084/jem.20151227

Maul, R. W., Saribasak, H., Martomo, S. A., McClure, R. L., Yang, W., Vaisman, A., Gramlich, H. S., Schatz, D. G., Woodgate, R., Wilson, D. M., 3rd, & Gearhart, P. J. (2011, Jan). Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat Immunol, 12*(1), 70-76. https://doi.org/10.1038/ni.1970

Mayer, S., Raulf, M. K., & Lepenies, B. (2017, Feb). C-type lectins: their network and roles in pathogen recognition and immunity. *Histochem Cell Biol, 147*(2), 223-237. https://doi.org/10.1007/s00418-016-1523-7

McBride, K. M., Barreto, V., Ramiro, A. R., Stavropoulos, P., & Nussenzweig, M. C. (2004, May 03). Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J Exp Med, 199*(9), 1235-1244. https://doi.org/10.1084/jem.20040373

McCarthy, H., Wierda, W. G., Barron, L. L., Cromwell, C. C., Wang, J., Coombes, K. R., Rangel, R., Elenitoba-Johnson, K. S., Keating, M. J., & Abruzzo, L. V. (2003, Jun 15). High expression of activation-induced cytidine deaminase (AID) and splice variants is a distinctive feature of poor-prognosis chronic lymphocytic leukemia. *Blood, 101*(12), 4903-4908. https://doi.org/10.1182/blood-2002-09-2906

Meffre, E., Catalan, N., Seltz, F., Fischer, A., Nussenzweig, M. C., & Durandy, A. (2001, Aug 06). Somatic hypermutation shapes the antibody repertoire of memory B cells in humans. *J Exp Med, 194*(3), 375-378. http://www.ncbi.nlm.nih.gov/pubmed/11489956

Mehr, R., Edelman, H., Sehgal, D., & Mage, R. (2004, Apr 15). Analysis of mutational lineage trees from sites of primary and secondary Ig gene diversification in rabbits and chickens. *J Immunol, 172*(8), 4790-4796. http://www.ncbi.nlm.nih.gov/pubmed/15067055

Melchers, F. (2015, Jun). Checkpoints that control B cell development. *J Clin Invest, 125*(6), 2203-2210. https://doi.org/10.1172/JCI78083

Meng, F. L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K. R., Meyers, R. M., Amor, C., Wasserman, C. R., Neuberg, D., Casellas, R., Nussenzweig, M. C., Bradner, J. E., Liu, X. S., & Alt, F. W. (2014, Dec 18). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell, 159*(7), 1538-1548. https://doi.org/10.1016/j.cell.2014.11.014

Merkl, R., & Sterner, R. (2016, Jan). Ancestral protein reconstruction: techniques and applications. *Biol Chem, 397*(1), 1-21. https://doi.org/10.1515/hsz-2015-0158

Merkl, R., & Sterner, R. (2016, 2016/12/01/). Reconstruction of ancestral enzymes. *Perspectives in Science, 9*, 17-23. https://doi.org/https://doi.org/10.1016/j.pisc.2016.08.002

Mesin, L., Ersching, J., & Victora, G. D. (2016, Sep 20). Germinal Center B Cell Dynamics. *Immunity, 45*(3), 471-482. https://doi.org/10.1016/j.immuni.2016.09.001

Methot, S. P., & Di Noia, J. M. (2017). Molecular Mechanisms of Somatic Hypermutation and Class Switch Recombination. *Adv Immunol, 133*, 37-87. https://doi.org/10.1016/bs.ai.2016.11.002

Mikkelsen, H., Lund, V., Larsen, R., & Seppola, M. (2011, Jan). Vibriosis vaccines based on various sero-subgroups of Vibrio anguillarum O2 induce specific protection in Atlantic cod (Gadus morhua L.) juveniles. *Fish Shellfish Immunol, 30*(1), 330-339. https://doi.org/10.1016/j.fsi.2010.11.007

Minegishi, Y., Lavoie, A., Cunningham-Rundles, C., Bedard, P. M., Hebert, J., Cote, L., Dan, K., Sedlak, D., Buckley, R. H., Fischer, A., Durandy, A., & Conley, M. E. (2000, Dec). Mutations in activation-induced cytidine deaminase in patients with hyper IgM syndrome. *Clin Immunol, 97*(3), 203-210. https://doi.org/10.1006/clim.2000.4956

Mirete-Bachiller, S., Olivieri, D., & Gambón Deza, F. (2019). *Gouania willdenowi is a teleost fish without immunoglobulin genes*. https://doi.org/10.1101/793695

Mix, E., Goertsches, R., & Zett, U. K. (2006, Sep). Immunoglobulins--basic considerations. *J Neurol, 253 Suppl 5*, V9-17. https://doi.org/10.1007/s00415-006-5002-2

Mjosberg, J., & Spits, H. (2016, Nov). Human innate lymphoid cells. *J Allergy Clin Immunol, 138*(5), 1265-1276. https://doi.org/10.1016/j.jaci.2016.09.009

Mohri, T., Nagata, K., Kuwamoto, S., Matsushita, M., Sugihara, H., Kato, M., Horie, Y., Murakami, I., & Hayashi, K. (2017, Jun). Aberrant expression of AID and AID activators of NF-kappaB and PAX5 is irrelevant to EBV-associated gastric cancers, but is associated with carcinogenesis in certain EBV-non-associated gastric cancers. *Oncol Lett, 13*(6), 4133-4140. https://doi.org/10.3892/ol.2017.5978

Mond, J. J., Lees, A., & Snapper, C. M. (1995). T cell-independent antigens type 2. *Annu Rev Immunol, 13*, 655-692. https://doi.org/10.1146/annurev.iy.13.040195.003255

Montecino-Rodriguez, E., Leathers, H., & Dorshkind, K. (2006, Mar). Identification of a B-1 B cell-specified progenitor. *Nat Immunol, 7*(3), 293-301. https://doi.org/10.1038/ni1301

Moon, S. Y., Eun, H. J., Baek, S. K., Jin, S. J., Kim, T. S., Kim, S. W., Seong, H. H., Choi, I. C., & Lee, J. H. (2016, Oct). Activation-Induced Cytidine Deaminase Induces DNA Demethylation of Pluripotency Genes in Bovine Differentiated Cells. *Cell Reprogram, 18*(5), 298-308. https://doi.org/10.1089/cell.2015.0076

Morgan, H. D., Dean, W., Coker, H. A., Reik, W., & Petersen-Mahrt, S. K. (2004, Dec 10). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem, 279*(50), 52353-52360. https://doi.org/10.1074/jbc.M407695200

Munoz, D. P., Lee, E. L., Takayama, S., Coppe, J. P., Heo, S. J., Boffelli, D., Di Noia, J. M., & Martin, D. I. (2013, Aug 06). Activation-induced cytidine deaminase (AID) is necessary for the epithelial-mesenchymal transition in mammary epithelial cells. *Proc Natl Acad Sci U S A, 110*(32), E2977-2986. https://doi.org/10.1073/pnas.1301021110

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. (2000, Sep 01). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell, 102*(5), 553-563. http://www.ncbi.nlm.nih.gov/pubmed/11007474

Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., & Honjo, T. (1999, Jun 25). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem, 274*(26), 18470-18476. http://www.ncbi.nlm.nih.gov/pubmed/10373455

Musat, M. G., Nitulescu, G. M., Surleac, M., Tsatsakis, A., Spandidos, D. A., & Margina, D. (2019, Dec). HIV1 integrase inhibitors targeting various DDE transposases: Retroviral integration versus RAGmediated recombination (Review). *Mol Med Rep, 20*(6), 4749-4762. https://doi.org/10.3892/mmr.2019.10777

Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K., & Honjo, T. (2000, Aug 15). Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene. *Genomics, 68*(1), 85-88. https://doi.org/10.1006/geno.2000.6268

Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T., Zhang, Y., & Kohli, R. M. (2012, Sep). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol, 8*(9), 751-758. https://doi.org/10.1038/nchembio.1042

Nagaoka, H., Muramatsu, M., Yamamura, N., Kinoshita, K., & Honjo, T. (2002, Feb 18). Activation-induced deaminase (AID)-directed hypermutation in the immunoglobulin Smu region: implication of AID involvement in a common step of class switch recombination and somatic hypermutation. *J Exp Med, 195*(4), 529-534. http://www.ncbi.nlm.nih.gov/pubmed/11854365

Nambu, Y., Sugai, M., Gonda, H., Lee, C. G., Katakai, T., Agata, Y., Yokota, Y., & Shimizu, A. (2003, Dec 19). Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science, 302*(5653), 2137-2140. https://doi.org/10.1126/science.1092481

Newberry, R. W., & Raines, R. T. (2019, Aug 16). Secondary Forces in Protein Folding. *ACS Chem Biol, 14*(8), 1677-1686. https://doi.org/10.1021/acschembio.9b00339

Newton, M. S., Arcus, V. L., & Patrick, W. M. (2015, Jun 6). Rapid bursts and slow declines: on the possible evolutionary trajectories of enzymes. *J R Soc Interface, 12*(107). https://doi.org/10.1098/rsif.2015.0036

Nguyen, T. T., Elsner, R. A., & Baumgarth, N. (2015, Feb 15). Natural IgM prevents autoimmunity by enforcing B cell central tolerance induction. *J Immunol, 194*(4), 1489-1502. https://doi.org/10.4049/jimmunol.1401880

Nishikawa, T., Ota, T., & Isogai, T. (2000, Nov). Prediction whether a human cDNA sequence contains initiation codon by combining statistical information and similarity with protein sequences. *Bioinformatics, 16*(11), 960-967. https://doi.org/10.1093/bioinformatics/16.11.960

Noguchi, E., Shibasaki, M., Inudou, M., Kamioka, M., Yokouchi, Y., Yamakawa-Kobayashi, K., Hamaguchi, H., Matsui, A., & Arinami, T. (2001, Sep). Association between a new polymorphism in the activation-induced cytidine deaminase gene and atopic asthma and the regulation of total serum IgE levels. *J Allergy Clin Immunol, 108*(3), 382-386. https://doi.org/10.1067/mai.2001.117456

Nojima, H., Hon-Nami, K., Oshima, T., & Noda, H. (1978, Jun 15). Reversible thermal unfolding of thermostable cytochrome c-552. *J Mol Biol, 122*(1), 33-42. https://doi.org/10.1016/0022-2836(78)90106-7

Nojima, H., Ikai, A., Oshima, T., & Noda, H. (1977, Nov 5). Reversible thermal unfolding of thermostable phosphoglycerate kinase. Thermostability associated with mean zero

enthalpy change. *J Mol Biol, 116*(3), 429-442. https://doi.org/10.1016/0022-2836(77)90078-x

Nowak, U., Matthews, A. J., Zheng, S., & Chaudhuri, J. (2011, Feb). The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. *Nat Immunol, 12*(2), 160-166. https://doi.org/10.1038/ni.1977

Nymo, I. H., Seppola, M., Al Dahouk, S., Bakkemo, K. R., Jimenez de Bagues, M. P., Godfroid, J., & Larsen, A. K. (2016). Experimental Challenge of Atlantic Cod (Gadus morhua) with a Brucella pinnipedialis Strain from Hooded Seal (Cystophora cristata). *PLoS One, 11*(7), e0159272. https://doi.org/10.1371/journal.pone.0159272

Ohm-Laursen, L., & Barington, T. (2007, Apr 1). Analysis of 6912 unselected somatic hypermutations in human VDJ rearrangements reveals lack of strand specificity and correlation between phase II substitution rates and distance to the nearest 3' activation-induced cytidine deaminase target. *J Immunol, 178*(7), 4322-4334. https://doi.org/10.4049/jimmunol.178.7.4322

Ohmori, K., Maeda, S., Okayama, T., Masuda, K., Ohno, K., & Tsujimoto, H. (2004, Jun). Molecular cloning of canine activation-induced cytidine deaminase (AID) cDNA and its expression in normal tissues. *J Vet Med Sci, 66*(6), 739-741. http://www.ncbi.nlm.nih.gov/pubmed/15240955

Olsson, M. H., Sondergaard, C. R., Rostkowski, M., & Jensen, J. H. (2011, Feb 8). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput, 7*(2), 525-537. https://doi.org/10.1021/ct100578z

Oppezzo, P., Vuillier, F., Vasconcelos, Y., Dumas, G., Magnac, C., Payelle-Brogard, B., Pritsch, O., & Dighiero, G. (2003, May 15). Chronic lymphocytic leukemia B cells expressing AID display dissociation between class switch recombination and somatic hypermutation. *Blood, 101*(10), 4029-4032. https://doi.org/10.1182/blood-2002-10-3175

Oreste, U., & Coscia, M. (2002, Aug 07). Specific features of immunoglobulin VH genes of the Antarctic teleost Trematomus bernacchii. *Gene, 295*(2), 199-204. http://www.ncbi.nlm.nih.gov/pubmed/12354654

Outters, P., Jaeger, S., Zaarour, N., & Ferrier, P. (2015). Long-Range Control of V(D)J Recombination & Allelic Exclusion: Modeling Views. *Adv Immunol, 128*, 363-413. https://doi.org/10.1016/bs.ai.2015.08.002

Owen, D. L., Sjaastad, L. E., & Farrar, M. A. (2019, Oct 15). Regulatory T Cell Development in the Thymus. *J Immunol, 203*(8), 2031-2041. https://doi.org/10.4049/jimmunol.1900662

Owen, J. A. (2019). *Kuby immunology* (Eighth edition.. ed.). New York : W.H. Freeman, Macmillan Learning.

Pagel, M., Meade, A., & Barker, D. (2004, Oct). Bayesian estimation of ancestral character states on phylogenies. *Syst Biol, 53*(5), 673-684. https://doi.org/10.1080/10635150490522232

Palgen, J. L., Tchitchek, N., Elhmouzi-Younes, J., Delandre, S., Namet, I., Rosenbaum, P., Dereuddre-Bosquet, N., Martinon, F., Cosma, A., Levy, Y., Le Grand, R., & Beignon, A. S. (2018, Feb 15). Prime and Boost Vaccination Elicit a Distinct Innate Myeloid Cell Immune Response. *Sci Rep, 8*(1), 3087. https://doi.org/10.1038/s41598-018-21222-2

Palma, J., Tokarz-Deptula, B., Deptula, J., & Deptula, W. (2018). Natural antibodies - facts known and unknown. *Cent Eur J Immunol, 43*(4), 466-475. https://doi.org/10.5114/ceji.2018.81354

Parham, P. (2015, Sep). Co-evolution of lymphocyte receptors with MHC class I. *Immunol Rev, 267*(1), 1-5. https://doi.org/10.1111/imr.12338

Parham, P. (2016, Sep 28). How the codfish changed its immune system. *Nat Genet, 48*(10), 1103-1104. https://doi.org/10.1038/ng.3684

Patel, B., Banerjee, R., Samanta, M., & Das, S. (2018, Jun). Diversity of Immunoglobulin (Ig) Isotypes and the Role of Activation-Induced Cytidine Deaminase (AID) in Fish. *Mol Biotechnol, 60*(6), 435-453. https://doi.org/10.1007/s12033-018-0081-8

Patenaude, A. M., Orthwein, A., Hu, Y., Campo, V. A., Kavli, B., Buschiazzo, A., & Di Noia, J. M. (2009, May). Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat Struct Mol Biol, 16*(5), 517-527. https://doi.org/10.1038/nsmb.1598

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., Nieland, T. J., Root, D. E., Casellas, R., & Nussenzweig, M. C. (2010, Oct 1). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell, 143*(1), 122-133. https://doi.org/10.1016/j.cell.2010.09.017

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., & Louppe, G. (2012, 01/02). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*.

Pei, J., Kim, B. H., & Grishin, N. V. (2008, Apr). PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res, 36*(7), 2295-2300. https://doi.org/10.1093/nar/gkn072

Peters, A., & Storb, U. (1996, Jan). Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity, 4*(1), 57-65. https://doi.org/10.1016/s1074-7613(00)80298-8

Petersen, M. F., & Steffensen, J. F. (2003, Jan). Preferred temperature of juvenile Atlantic cod Gadus morhua with different haemoglobin genotypes at normoxia and moderate hypoxia. *J Exp Biol, 206*(Pt 2), 359-364. http://www.ncbi.nlm.nih.gov/pubmed/12477905

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004, Oct). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem, 25*(13), 1605-1612. https://doi.org/10.1002/jcc.20084

Pham, P., Afif, S. A., Shimoda, M., Maeda, K., Sakaguchi, N., Pedersen, L. C., & Goodman, M. F. (2016, Jul). Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification. *DNA Repair (Amst), 43*, 48-56. https://doi.org/10.1016/j.dnarep.2016.05.029

Pham, P., Landolph, A., Mendez, C., Li, N., & Goodman, M. F. (2013, Oct 11). A biochemical analysis linking APOBEC3A to disparate HIV-1 restriction and skin cancer. *J Biol Chem, 288*(41), 29294-29304. https://doi.org/10.1074/jbc.M113.504175

Philip, V., Harris, J., Adams, R., Nguyen, D., Spiers, J., Baudry, J., Howell, E. E., & Hinde, R. J. (2011, Apr 12). A survey of aspartate-phenylalanine and glutamate-phenylalanine interactions in the protein data bank: searching for anion-pi pairs. *Biochemistry, 50*(14), 2939-2950. https://doi.org/10.1021/bi200066k

Pieper, K., Grimbacher, B., & Eibel, H. (2013, Apr). B-cell biology and development. *J Allergy Clin Immunol, 131*(4), 959-971. https://doi.org/10.1016/j.jaci.2013.01.046

Pinaud, E., Khamlichi, A. A., Le Morvan, C., Drouet, M., Nalesso, V., Le Bert, M., & Cogné, M. (2001, Aug). Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. *Immunity, 15*(2), 187-199. https://doi.org/10.1016/s1074-7613(01)00181-9

Pone, E. J., Zhang, J., Mai, T., White, C. A., Li, G., Sakakura, J. K., Patel, P. J., Al-Qahtani, A., Zan, H., Xu, Z., & Casali, P. (2012, Apr 03). BCR-signalling synergizes with TLR-signalling for induction of AID and immunoglobulin class-switching through the non-canonical NF-kappaB pathway. *Nat Commun, 3*, 767. https://doi.org/10.1038/ncomms1769

Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., & Reik, W. (2010, Feb 25). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature, 463*(7284), 1101-1105. https://doi.org/10.1038/nature08829

Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F., & Chen, X. S. (2007, Jan 25). The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature, 445*(7126), 447-451. https://doi.org/10.1038/nature05492

Pucci, F., Dhanani, M., Dehouck, Y., & Rooman, M. (2014). Protein thermostability prediction within homologous families using temperature-dependent statistical potentials. *PLoS One, 9*(3), e91659. https://doi.org/10.1371/journal.pone.0091659

Pucci, F., Kwasigroch, J. M., & Rooman, M. (2017). SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. *Bioinformatics, 33*(21), 3415-3422. https://doi.org/10.1093/bioinformatics/btx417

Pucci, F., & Rooman, M. (2014, Jul). Stability curve prediction of homologous proteins using temperature-dependent statistical potentials. *PLoS Comput Biol, 10*(7), e1003689. https://doi.org/10.1371/journal.pcbi.1003689

Pucci, F., & Rooman, M. (2016, May). Towards an accurate prediction of the thermal stability of homologous proteins. *J Biomol Struct Dyn, 34*(5), 1132-1142. https://doi.org/10.1080/07391102.2015.1073631

Pucci, F., & Rooman, M. (2017, 2017/02/01/). Physical and molecular bases of protein thermal stability and cold adaptation. *Curr Opin Struct Biol, 42*, 117-128. https://doi.org/https://doi.org/10.1016/j.sbi.2016.12.007

Qamar, N., & Fuleihan, R. L. (2014, Apr). The hyper IgM syndromes. *Clin Rev Allergy Immunol, 46*(2), 120-130. https://doi.org/10.1007/s12016-013-8378-7

Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K. R., Resch, W., Liang, G., Tang, Z., Mathé, E., Benner, C., Dubois, W., Nelson, S., Vian, L., Oliveira, T. Y., Jankovic, M., Hakim, O., Gazumyan, A., Pavri, R., Awasthi, P., Song, B., Liu, G., Chen, L., Zhu, S., Feigenbaum, L., Staudt, L., Murre, C., Ruan, Y., Robbiani, D. F., Pan-Hammarström, Q., Nussenzweig, M. C., & Casellas, R. (2014, Dec 18). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell, 159*(7), 1524-1537. https://doi.org/10.1016/j.cell.2014.11.013

Qiao, Q., Wang, L., Meng, F. L., Hwang, J. K., Alt, F. W., & Wu, H. (2017, Aug 03). AID Recognizes Structured DNA for Class Switch Recombination. *Mol Cell, 67*(3), 361-373 e364. https://doi.org/10.1016/j.molcel.2017.06.034

Qin, T., Liu, Z., & Zhao, H. (2015, Dec). Organization and genomic complexity of sheep immunoglobulin light chain gene loci. *Immunol Lett, 168*(2), 313-318. https://doi.org/10.1016/j.imlet.2015.10.010

Qin, T., Zhao, H., Zhu, H., Wang, D., Du, W., & Hao, H. (2015, Aug). Immunoglobulin genomics in the prairie vole (Microtus ochrogaster). *Immunol Lett, 166*(2), 79-86. https://doi.org/10.1016/j.imlet.2015.06.001

Qiu, X., Duvvuri, V. R., & Bahl, J. (2019, May 28). Computational Approaches and Challenges to Developing Universal Influenza Vaccines. *Vaccines (Basel), 7*(2). https://doi.org/10.3390/vaccines7020045

Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., & Cairns, B. R. (2008, Dec 26). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell, 135*(7), 1201-1212. https://doi.org/10.1016/j.cell.2008.11.042

Rajaee, A., Barnett, R., & Cheadle, W. G. (2018, Feb/Mar). Pathogen- and Danger-Associated Molecular Patterns and the Cytokine Response in Sepsis. *Surg Infect (Larchmt), 19*(2), 107-116. https://doi.org/10.1089/sur.2017.264

Rajewsky, K., Forster, I., & Cumano, A. (1987, Nov 20). Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science, 238*(4830), 1088-1094. http://www.ncbi.nlm.nih.gov/pubmed/3317826

Ramiro, A. R., & Barreto, V. M. (2016, Apr). Correction to: 'Activation-induced cytidine deaminase and active cytidine demethylation': [Trends in Biochemical Sciences, 40 (2015), 172-181]. *Trends Biochem Sci, 41*(4), 387. https://doi.org/10.1016/j.tibs.2015.12.005

Randall, R. N., Radford, C. E., Roof, K. A., Natarajan, D. K., & Gaucher, E. A. (2016, Sep 15). An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat Commun, 7*, 12847. https://doi.org/10.1038/ncomms12847

Razvi, A., & Scholtz, J. M. (2006, Jul). Lessons in stability from thermophilic proteins. *Protein Sci, 15*(7), 1569-1578. https://doi.org/10.1110/ps.062130306

Reis e Sousa, C. (2004, Feb). Activation of dendritic cells: translating innate into adaptive immunity. *Curr Opin Immunol, 16*(1), 21-25. https://doi.org/10.1016/j.coi.2003.11.007

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A., & Durandy, A. (2000, Sep 01). Activation-induced cytidine deaminase (AID) deficiency

causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell, 102*(5), 565-575. http://www.ncbi.nlm.nih.gov/pubmed/11007475

Reynaud, C. A., Mackay, C. R., Müller, R. G., & Weill, J. C. (1991, Mar 8). Somatic generation of diversity in a mammalian primary lymphoid organ: the sheep ileal Peyer's patches. *Cell, 64*(5), 995-1005. https://doi.org/10.1016/0092-8674(91)90323-q

Rise, M. L., Hall, J. R., Alcock, B. P., & Hori, T. S. (2012, Nov 10). Dynamic expression profiles of virus-responsive and putative antimicrobial peptide-encoding transcripts during Atlantic cod (Gadus morhua) embryonic and early larval development. *Gene, 509*(2), 232-246. https://doi.org/10.1016/j.gene.2012.08.017

Rodríguez-Cortez, V. C., Martínez-Redondo, P., Català-Moll, F., Rodríguez-Ubreva, J., Garcia-Gomez, A., Poorani-Subramani, G., Ciudad, L., Hernando, H., Pérez-García, A., Company, C., Urquiza, J. M., Ramiro, A. R., Di Noia, J. M., Vaquero, A., & Ballestar, E. (2017, Aug 8). Activation-induced cytidine deaminase targets SUV4-20-mediated histone H4K20 trimethylation to class-switch recombination sites. *Sci Rep, 7*(1), 7594. https://doi.org/10.1038/s41598-017-07380-9

Rogozin, I. B., Basu, M. K., Jordan, I. K., Pavlov, Y. I., & Koonin, E. V. (2005, Sep). APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle, 4*(9), 1281-1285. https://doi.org/10.4161/cc.4.9.1994

Roh, J. S., & Sohn, D. H. (2018, Aug). Damage-Associated Molecular Patterns in Inflammatory Diseases. *Immune Netw, 18*(4), e27. https://doi.org/10.4110/in.2018.18.e27

Ron Fricke, W. E. a. J. D. F. *Eschimeyer's catalog of fishes: species by family/subfamily. (2019). (http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp). Electronic version accessed 07 Oct 2019.* http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics, 19*(12), 1572-1574. https://doi.org/10.1093/bioinformatics/btg180

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012, May). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol, 61*(3), 539-542. https://doi.org/10.1093/sysbio/sys029

Roskoski, R. (2015). Michaelis-Menten Kinetics. In. https://doi.org/10.1016/B978-0-12-801238-3.05143-6

Roth, D. B. (2000, 2000/08/02). From lymphocytes to sharks: V(D)J recombinase moves to the germline. *Genome Biol, 1*(2), reviews1014.1011. https://doi.org/10.1186/gb-2000-1-2-reviews1014

Roth, D. B. (2014, Dec). V(D)J Recombination: Mechanism, Errors, and Fidelity. *Microbiol Spectr, 2*(6). https://doi.org/10.1128/microbiolspec.MDNA3-0041-2014

Rothenfluh, H. S., Taylor, L., Bothwell, A. L., Both, G. W., & Steele, E. J. (1993, Sep). Somatic hypermutation in 5' flanking regions of heavy chain antibody variable regions. *Eur J Immunol, 23*(9), 2152-2159. https://doi.org/10.1002/eji.1830230916

Roy, A., Kucukural, A., & Zhang, Y. (2010, Apr). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc, 5*(4), 725-738. https://doi.org/10.1038/nprot.2010.5

Roy, D., Yu, K., & Lieber, M. R. (2008, Jan). Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol Cell Biol, 28*(1), 50-60. https://doi.org/10.1128/mcb.01251-07

Ruder, U., Denkert, C., Kunze, C. A., Jank, P., Lindner, J., Johrens, K., Kulbe, H., Sehouli, J., Dietel, M., Braicu, E., & Darb-Esfahani, S. (2019, Apr). APOBEC3B protein expression and mRNA analyses in patients with high-grade serous ovarian carcinoma. *Histol Histopathol, 34*(4), 405-417. https://doi.org/10.14670/HH-18-050

Ruiz, M., Giudicelli, V., Ginestoux, C., Stoehr, P., Robinson, J., Bodmer, J., Marsh, S. G., Bontrop, R., Lemaitre, M., Lefranc, G., Chaume, D., & Lefranc, M. P. (2000, Jan 1). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res, 28*(1), 219-221. http://www.ncbi.nlm.nih.gov/pubmed/10592230

Salinas, I. (2015, Aug 12). The Mucosal Immune System of Teleost Fish. *Biology (Basel), 4*(3), 525-539. https://doi.org/10.3390/biology4030525

Sallan, L. C. (2014, Nov). Major issues in the origins of ray-finned fish (Actinopterygii) biodiversity. *Biol Rev Camb Philos Soc, 89*(4), 950-971. https://doi.org/10.1111/brv.12086

Samuelsen, O. B., Nerland, A. H., Jorgensen, T., Schroder, M. B., Svasand, T., & Bergh, O. (2006, Aug 30). Viral and bacterial diseases of Atlantic cod Gadus morhua, their prophylaxis and treatment: a review. *Dis Aquat Organ, 71*(3), 239-254. https://doi.org/10.3354/dao071239

Saribasak, H., Maul, R. W., Cao, Z., Yang, W. W., Schenten, D., Kracker, S., & Gearhart, P. J. (2012, Jun 4). DNA polymerase zeta generates tandem mutations in immunoglobulin variable regions. *J Exp Med, 209*(6), 1075-1081. https://doi.org/10.1084/jem.20112234

Sasaki, H., Suzuki, A., Tatematsu, T., Shitara, M., Hikosaka, Y., Okuda, K., Moriyama, S., Yano, M., & Fujii, Y. (2014, May). APOBEC3B gene overexpression in non-small-cell lung cancer. *Biomed Rep, 2*(3), 392-395. https://doi.org/10.3892/br.2014.256

Saunders, H. L., & Magor, B. G. (2004, Jun). Cloning and expression of the AID gene in the channel catfish. *Dev Comp Immunol, 28*(7-8), 657-663. https://doi.org/10.1016/j.dci.2004.01.002

Saunders, H. L., Oko, A. L., Scott, A. N., Fan, C. W., & Magor, B. G. (2010, Jun). The cellular context of AID expressing cells in fish lymphoid tissues. *Dev Comp Immunol, 34*(6), 669-676. https://doi.org/10.1016/j.dci.2010.01.013

Schaeffer, L. (2008). Chapter 14 - The Role of Functional Groups in Drug–Receptor Interactions. In C. G. Wermuth, D. Aldous, P. Raboisson, & D. Rognan (Eds.), *The Practice of Medicinal Chemistry (Fourth Edition)* (pp. 359-378). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-417205-0.00014-6

Schreck, S., Buettner, M., Kremmer, E., Bogdan, M., Herbst, H., & Niedobitek, G. (2006, Sep). Activation-induced cytidine deaminase (AID) is expressed in normal spermatogenesis but only infrequently in testicular germ cell tumours. *J Pathol, 210*(1), 26-31. https://doi.org/10.1002/path.2014

Schroder, M. B., Ellingsen, T., Mikkelsen, H., Norderhus, E. A., & Lund, V. (2009, Aug). Comparison of antibody responses in Atlantic cod (Gadus morhua L.) to Vibrio anguillarum, Aeromonas salmonicida and Francisella sp. *Fish Shellfish Immunol, 27*(2), 112-119. https://doi.org/10.1016/j.fsi.2008.11.016

Schroeder, H. W., Jr., & Cavacini, L. (2010, Feb). Structure and function of immunoglobulins. *J Allergy Clin Immunol, 125*(2 Suppl 2), S41-52. https://doi.org/10.1016/j.jaci.2009.09.046

Seifert, M., Scholtysik, R., & Kuppers, R. (2019). Origin and Pathogenesis of B Cell Lymphomas. *Methods Mol Biol, 1956*, 1-33. https://doi.org/10.1007/978-1-4939-9151-8_1

Senavirathne, G., Bertram, J. G., Jaszczur, M., Chaurasiya, K. R., Pham, P., Mak, C. H., Goodman, M. F., & Rueda, D. (2015, Dec 18). Activation-induced deoxycytidine deaminase (AID) co-transcriptional scanning at single-molecule resolution. *Nat Commun, 6*, 10209. https://doi.org/10.1038/ncomms10209

Sernandez, I. V., de Yebenes, V. G., Dorsett, Y., & Ramiro, A. R. (2008). Haploinsufficiency of activation-induced deaminase for antibody diversification and chromosome translocations both in vitro and in vivo. *PLoS One, 3*(12), e3927. https://doi.org/10.1371/journal.pone.0003927

Shen, H. M., Poirier, M. G., Allen, M. J., North, J., Lal, R., Widom, J., & Storb, U. (2009, May 11). The activation-induced cytidine deaminase (AID) efficiently targets DNA in nucleosomes but only during transcription. *J Exp Med, 206*(5), 1057-1071. https://doi.org/10.1084/jem.20082678

Shi, K., Carpenter, M. A., Banerjee, S., Shaban, N. M., Kurahashi, K., Salamango, D. J., McCann, J. L., Starrett, G. J., Duffy, J. V., Demir, O., Amaro, R. E., Harki, D. A., Harris, R. S., & Aihara, H. (2017, Feb). Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol, 24*(2), 131-139. https://doi.org/10.1038/nsmb.3344

Shi, Y., & Yokoyama, S. (2003, Jul 8). Molecular analysis of the evolutionary significance of ultraviolet vision in vertebrates. *Proc Natl Acad Sci U S A, 100*(14), 8308-8313. https://doi.org/10.1073/pnas.1532535100

Shi, Z., Cai, Z., Sanchez, A., Zhang, T., Wen, S., Wang, J., Yang, J., Fu, S., & Zhang, D. (2011). A novel Toll-like receptor that recognizes vesicular stomatitis virus. *J Biol Chem, 286*(6), 4517-4524. https://doi.org/10.1074/jbc.M110.159590

Shimamoto, R., Amano, N., Ichisaka, T., Watanabe, A., Yamanaka, S., & Okita, K. (2014). Generation and characterization of induced pluripotent stem cells from Aid-deficient mice. *PLoS One, 9*(4), e94735. https://doi.org/10.1371/journal.pone.0094735

Shimoda, N., Hirose, K., Kaneto, R., Izawa, T., Yokoi, H., Hashimoto, N., & Kikuchi, Y. (2014). No evidence for AID/MBD4-coupled DNA demethylation in zebrafish embryos. *PLoS One, 9*(12), e114816. https://doi.org/10.1371/journal.pone.0114816

Shiraki, K., Nishikori, S., Fujiwara, S., Hashimoto, H., Kai, Y., Takagi, M., & Imanaka, T. (2001, Aug). Comparative analyses of the conformational stability of a hyperthermophilic protein and its mesophilic counterpart. *Eur J Biochem, 268*(15), 4144-4150. https://doi.org/10.1046/j.1432-1327.2001.02324.x

Siddiqui, K. S., & Cavicchioli, R. (2006). Cold-adapted enzymes. *Annu Rev Biochem, 75*, 403-433. https://doi.org/10.1146/annurev.biochem.75.103004.142723

Silvas, T. V., & Schiffer, C. A. (2019, Sep). APOBEC3s: DNA-editing human cytidine deaminases. *Protein Sci, 28*(9), 1552-1566. https://doi.org/10.1002/pro.3670

Smalas, A. O., Leiros, H. K., Os, V., & Willassen, N. P. (2000). Cold adapted enzymes. *Biotechnol Annu Rev, 6*, 1-57. http://www.ncbi.nlm.nih.gov/pubmed/11193291

Smith, N. C., Rise, M. L., & Christian, S. L. (2019). A Comparison of the Innate and Adaptive Immune Systems in Cartilaginous Fish, Ray-Finned Fish, and Lobe-Finned Fish. *Front Immunol, 10*, 2292. https://doi.org/10.3389/fimmu.2019.02292

Sohail, A., Klapacz, J., Samaranayake, M., Ullah, A., & Bhagwat, A. S. (2003, Jun 15). Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res, 31*(12), 2990-2994. https://doi.org/10.1093/nar/gkg464

Solbakken, M. H. (2016). Evolutionary and functional insight into the teleost immune system- lessons learned from Atlantic cod and other teleosts [PhD thesis].

Solbakken, M. H., Jentoft, S., Reitan, T., Mikkelsen, H., Gregers, T. F., Bakke, O., Jakobsen, K. S., & Seppola, M. (2019, Jun). Disentangling the immune response and host-pathogen interactions in Francisella noatunensis infected Atlantic cod. *Comp Biochem Physiol Part D Genomics Proteomics, 30*, 333-346. https://doi.org/10.1016/j.cbd.2019.04.004

Solbakken, M. H., Jentoft, S., Reitan, T., Mikkelsen, H., Jakobsen, K. S., & Seppola, M. (2019, Sep). Whole transcriptome analysis of the Atlantic cod vaccine response reveals subtle changes in adaptive immunity. *Comp Biochem Physiol Part D Genomics Proteomics, 31*, 100597. https://doi.org/10.1016/j.cbd.2019.100597

Solbakken, M. H., Rise, M. L., Jakobsen, K. S., & Jentoft, S. (2016, Dec 31). Successive Losses of Central Immune Genes Characterize the Gadiformes' Alternate Immunity. *Genome Biol Evol, 8*(11), 3508-3515. https://doi.org/10.1093/gbe/evw250

Solbakken, M. H., Torresen, O. K., Nederbragt, A. J., Seppola, M., Gregers, T. F., Jakobsen, K. S., & Jentoft, S. (2016, Apr 29). Evolutionary redesign of the Atlantic cod (Gadus morhua L.) Toll-like receptor repertoire by gene losses and expansions. *Sci Rep, 6*, 25211. https://doi.org/10.1038/srep25211

Solbakken, M. H., Voje, K. L., Jakobsen, K. S., & Jentoft, S. (2017, Apr 26). Linking species habitat and past palaeoclimatic events to evolution of the teleost innate immune system. *Proc Biol Sci, 284*(1853). https://doi.org/10.1098/rspb.2016.2810

Solem, S. T., & Stenvik, J. (2006). Antibody repertoire development in teleosts--a review with emphasis on salmonids and Gadus morhua L. *Dev Comp Immunol, 30*(1-2), 57-76. https://doi.org/10.1016/j.dci.2005.06.007

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. https://doi.org/10.1093/bioinformatics/btu033

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res, 34*(suppl_2), W435-W439. https://doi.org/10.1093/nar/gkl200

Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrom, M., Gregers, T. F., Rounge, T. B., Paulsen, J., Solbakken, M. H., Sharma, A., Wetten, O. F., Lanzen, A., Winer, R., Knight, J., Vogel, J. H., Aken, B., Andersen, O., Lagesen, K., Tooming-Klunderud, A., Edvardsen, R. B., Tina, K. G., Espelund, M., Nepal, C., Previti, C., Karlsen, B. O., Moum, T., Skage, M., Berg, P. R., Gjoen, T., Kuhl, H., Thorsen, J., Malde, K., Reinhardt, R., Du, L., Johansen, S. D., Searle, S., Lien, S., Nilsen, F., Jonassen, I., Omholt, S. W., Stenseth, N. C., & Jakobsen, K. S. (2011, Aug 10). The genome sequence of Atlantic cod reveals a unique immune system. *Nature, 477*(7363), 207-210. https://doi.org/10.1038/nature10342

Star, B., Torresen, O. K., Nederbragt, A. J., Jakobsen, K. S., Pampoulie, C., & Jentoft, S. (2016, Aug 08). Genomic characterization of the Atlantic cod sex-locus. *Sci Rep, 6*, 31235. https://doi.org/10.1038/srep31235

Stavnezer, J., & Schrader, C. E. (2014, Dec 1). IgH chain class switch recombination: mechanism and regulation. *J Immunol, 193*(11), 5370-5378. https://doi.org/10.4049/jimmunol.1401849

Steele, E. J. (2016, Sep). Somatic hypermutation in immunity and cancer: Critical analysis of strand-biased and codon-context mutation signatures. *DNA Repair (Amst), 45*, 1-24. https://doi.org/10.1016/j.dnarep.2016.07.001

Storb, U., Peters, A., Klotz, E., Kim, N., Shen, H. M., Kage, K., Rogerson, B., & Martin, T. E. (1998). Somatic hypermutation of immunoglobulin genes is linked to transcription. *Curr Top Microbiol Immunol, 229*, 11-19. https://doi.org/10.1007/978-3-642-71984-4_2

Storey, K. B. (2016, 2016/09/01/). Comparative enzymology—new insights from studies of an "old" enzyme, lactate dehydrogenase. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 199*, 13-20. https://doi.org/https://doi.org/10.1016/j.cbpb.2015.12.004

Sun, H., Merugu, S., Gu, X., Kang, Y. Y., Dickinson, D. P., Callaerts, P., & Li, W. H. (2002, Sep). Identification of essential amino acid changes in paired domain evolution using a novel combination of evolutionary analysis and in vitro and in vivo studies. *Mol Biol Evol, 19*(9), 1490-1500. https://doi.org/10.1093/oxfordjournals.molbev.a004212

Swaminathan, S., Klemm, L., Park, E., Papaemmanuil, E., Ford, A., Kweon, S. M., Trageser, D., Hasselfeld, B., Henke, N., Mooster, J., Geng, H., Schwarz, K., Kogan, S. C., Casellas, R., Schatz, D. G., Lieber, M. R., Greaves, M. F., & Muschen, M. (2015, Jul).

Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. *Nat Immunol, 16*(7), 766-774. https://doi.org/10.1038/ni.3160

Tacchi, L., Larragoite, E. T., Munoz, P., Amemiya, C. T., & Salinas, I. (2015, Sep 21). African Lungfish Reveal the Evolutionary Origins of Organized Mucosal Lymphoid Tissue in Vertebrates. *Curr Biol, 25*(18), 2417-2424. https://doi.org/10.1016/j.cub.2015.07.066

Takeda, K., & Akira, S. (2015, Apr 1). Toll-like receptors. *Curr Protoc Immunol, 109*, 14 12 11-10. https://doi.org/10.1002/0471142735.im1412s109

Takeuchi, A., & Saito, T. (2017). CD4 CTL, a Cytotoxic Subset of CD4(+) T Cells, Their Differentiation and Function. *Front Immunol, 8*, 194. https://doi.org/10.3389/fimmu.2017.00194

Takizawa, M., Tolarova, H., Li, Z., Dubois, W., Lim, S., Callen, E., Franco, S., Mosaico, M., Feigenbaum, L., Alt, F. W., Nussenzweig, A., Potter, M., & Casellas, R. (2008, Sep 01). AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. *J Exp Med, 205*(9), 1949-1957. https://doi.org/10.1084/jem.20081007

Tang, C., Bagnara, D., Chiorazzi, N., Scharff, M. D., & MacCarthy, T. (2020). AID Overlapping and Polη Hotspots Are Key Features of Evolutionary Variation Within the Human Antibody Heavy Chain (IGHV) Genes. *Front Immunol, 11*, 788. https://doi.org/10.3389/fimmu.2020.00788

Tang, E. S., & Martin, A. (2007, Nov). Immunoglobulin gene conversion: synthesizing antibody diversification and DNA repair. *DNA Repair (Amst), 6*(11), 1557-1571. https://doi.org/10.1016/j.dnarep.2007.05.002

Thompson, M. R., Kaminski, J. J., Kurt-Jones, E. A., & Fitzgerald, K. A. (2011, Jun). Pattern recognition receptors and the innate immune response to viral infection. *Viruses, 3*(6), 920-940. https://doi.org/10.3390/v3060920

Thomson, D. W., Shahrin, N. H., Wang, P. P. S., Wadham, C., Shanmuganathan, N., Scott, H. S., Dinger, M. E., Hughes, T. P., Schreiber, A. W., & Branford, S. (2020, Feb 19). Aberrant RAG-mediated recombination contributes to multiple structural rearrangements in lymphoid blast crisis of chronic myeloid leukemia. *Leukemia*. https://doi.org/10.1038/s41375-020-0751-y

Thornton, J. W. (2001, May 8). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc Natl Acad Sci U S A, 98*(10), 5671-5676. https://doi.org/10.1073/pnas.091553298

Thornton, J. W., Need, E., & Crews, D. (2003, Sep 19). Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science, 301*(5640), 1714-1717. https://doi.org/10.1126/science.1086185

Tomlinson, I. M. (1998). Immunoglobulin Genes. In P. J. Delves (Ed.), *Encyclopedia of Immunology (Second Edition)* (pp. 1323-1328). Elsevier. https://doi.org/https://doi.org/10.1006/rwei.1999.0339

Torres, A. G., Pineyro, D., Filonava, L., Stracker, T. H., Batlle, E., & Ribas de Pouplana, L. (2014, Nov 28). A-to-I editing on tRNAs: biochemical, biological and evolutionary implications. *FEBS Lett, 588*(23), 4279-4286. https://doi.org/10.1016/j.febslet.2014.09.025

Torresen, O. K., Star, B., Jentoft, S., Reinar, W. B., Grove, H., Miller, J. R., Walenz, B. P., Knight, J., Ekholm, J. M., Peluso, P., Edvardsen, R. B., Tooming-Klunderud, A., Skage, M., Lien, S., Jakobsen, K. S., & Nederbragt, A. J. (2017, Jan 18). An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics, 18*(1), 95. https://doi.org/10.1186/s12864-016-3448-x

Trachsel, M. A., Ottiger, P., Frey, H.-M., Pfaffen, C., Bihlmeier, A., Klopper, W., & Leutwyler, S. (2015, 2015/06/25). Modeling the Histidine–Phenylalanine Interaction: The NH···π Hydrogen Bond of Imidazole·Benzene. *The Journal of Physical Chemistry B, 119*(25), 7778-7790. https://doi.org/10.1021/jp512766r

Trancoso, I., Morimoto, R., & Boehm, T. (2020, Apr 27). Co-evolution of mutagenic genome editors and vertebrate adaptive immunity. *Curr Opin Immunol, 65*, 32-41. https://doi.org/10.1016/j.coi.2020.03.001

Trott, O., & Olson, A. J. (2010, Jan 30). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem, 31*(2), 455-461. https://doi.org/10.1002/jcc.21334

Tsay, G. J., & Zouali, M. (2018). The Interplay Between Innate-Like B Cells and Other Cell Types in Autoimmunity. *Front Immunol, 9*, 1064. https://doi.org/10.3389/fimmu.2018.01064

Uchiyama, T. (1982). Modulation of immune response by bacterial lipopolysaccharide (LPS): roles of macrophages and T cells in vitro adjuvant effect of LPS on antibody response to T cell-dependent and T cell-independent antigens. *Microbiol Immunol, 26*(3), 213-225. https://doi.org/10.1111/j.1348-0421.1982.tb00173.x

Vaidyanathan, B., Yen, W. F., Pucella, J. N., & Chaudhuri, J. (2014). AIDing Chromatin and Transcription-Coupled Orchestration of Immunoglobulin Class-Switch Recombination. *Front Immunol, 5*, 120. https://doi.org/10.3389/fimmu.2014.00120

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data* (1st ed.). O'Reilly Media Inc.

Varshney, P., Yadav, V., & Saini, N. (2016, Sep). Lipid rafts in immune signalling: current progress and future perspective. *Immunology, 149*(1), 13-24. https://doi.org/10.1111/imm.12617

Vaughn, J. P., Creacy, S. D., Routh, E. D., Joyner-Butt, C., Jenkins, G. S., Pauli, S., Nagamine, Y., & Akman, S. A. (2005, Nov 18). The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J Biol Chem, 280*(46), 38117-38120. https://doi.org/10.1074/jbc.C500348200

Verma, S., Goldammer, T., & Aitken, R. (2010, Apr 15). Cloning and expression of activation induced cytidine deaminase from Bos taurus. *Vet Immunol Immunopathol, 134*(3-4), 151-159. https://doi.org/10.1016/j.vetimm.2009.08.016

Vialle, R. A., Tamuri, A. U., & Goldman, N. (2018, Jul 1). Alignment Modulates Ancestral Sequence Reconstruction Accuracy. *Mol Biol Evol, 35*(7), 1783-1797. https://doi.org/10.1093/molbev/msy055

Vivier, E., Artis, D., Colonna, M., Diefenbach, A., Di Santo, J. P., Eberl, G., Koyasu, S., Locksley, R. M., McKenzie, A. N. J., Mebius, R. E., Powrie, F., & Spits, H. (2018, Aug 23). Innate Lymphoid Cells: 10 Years On. *Cell, 174*(5), 1054-1066. https://doi.org/10.1016/j.cell.2018.07.017

Voss, E. W., Jr., & Sigel, M. M. (1972, Oct). Valence and temporal change in affinity of purified 7S and 18S nurse shark anti-2,4 dinitrophenyl antibodies. *J Immunol, 109*(4), 665-673. http://www.ncbi.nlm.nih.gov/pubmed/4561469

Vuong, B. Q., Herrick-Reynolds, K., Vaidyanathan, B., Pucella, J. N., Ucher, A. J., Donghia, N. M., Gu, X., Nicolas, L., Nowak, U., Rahman, N., Strout, M. P., Mills, K. D., Stavnezer, J., & Chaudhuri, J. (2013, Nov). A DNA break- and phosphorylation-dependent positive feedback loop promotes immunoglobulin class-switch recombination. *Nat Immunol, 14*(11), 1183-1189. https://doi.org/10.1038/ni.2732

Wagner, S. D., Milstein, C., & Neuberger, M. S. (1995, Aug 31). Codon bias targets mutation. *Nature, 376*(6543), 732. https://doi.org/10.1038/376732a0

Wakae, K., Magor, B. G., Saunders, H., Nagaoka, H., Kawamura, A., Kinoshita, K., Honjo, T., & Muramatsu, M. (2006, Jan). Evolution of class switch recombination function in fish

activation-induced cytidine deaminase, AID. *Int Immunol, 18*(1), 41-47. https://doi.org/10.1093/intimm/dxh347

Walther, S., Tietze, M., Czerny, C. P., Konig, S., & Diesterbeck, U. S. (2016). Development of a Bioinformatics Framework for the Detection of Gene Conversion and the Analysis of Combinatorial Diversity in Immunoglobulin Heavy Chains in Four Cattle Breeds. *PLoS One, 11*(11), e0164567. https://doi.org/10.1371/journal.pone.0164567

Wang, B., Tian, Y., & Yin, Q. (2019). AIM2 Inflammasome Assembly and Signaling. *Adv Exp Med Biol, 1172*, 143-155. https://doi.org/10.1007/978-981-13-9367-9_7

Wang, M., Rada, C., & Neuberger, M. S. (2010, Jan 18). Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *J Exp Med, 207*(1), 141-153. https://doi.org/10.1084/jem.20092238

Wang, M., Yang, Z., Rada, C., & Neuberger, M. S. (2009, Jul). AID upmutants isolated using a high-throughput screen highlight the immunity/cancer balance limiting DNA deaminase activity. *Nat Struct Mol Biol, 16*(7), 769-776. https://doi.org/10.1038/nsmb.1623

Wei, L., Chahwan, R., Wang, S., Wang, X., Pham, P. T., Goodman, M. F., Bergman, A., Scharff, M. D., & MacCarthy, T. (2015, Feb 17). Overlapping hotspots in CDRs are critical sites for V region diversification. *Proc Natl Acad Sci U S A, 112*(7), E728-737. https://doi.org/10.1073/pnas.1500788112

Wheeler, L. C., Lim, S. A., Marqusee, S., & Harms, M. J. (2016, Jun). The thermostability and specificity of ancient proteins. *Curr Opin Struct Biol, 38*, 37-43. https://doi.org/10.1016/j.sbi.2016.05.015

Wiens, G. D., Brown, M., & Rittenberg, M. B. (2003, May 15). Repertoire shift in the humoral response to phosphocholine-keyhole limpet hemocyanin: VH somatic mutation in germinal center B cells impairs T15 Ig function. *J Immunol, 170*(10), 5095-5102. http://www.ncbi.nlm.nih.gov/pubmed/12734355

Wiens, G. D., Lekkerkerker, A., Veltman, I., & Rittenberg, M. B. (2001, Aug 15). Mutation of a single conserved residue in VH complementarity-determining region 2 results in a severe Ig secretion defect. *J Immunol, 167*(4), 2179-2186. http://www.ncbi.nlm.nih.gov/pubmed/11490003

Wiens, G. D., Roberts, V. A., Whitcomb, E. A., O'Hare, T., Stenzel-Poore, M. P., & Rittenberg, M. B. (1998, Apr). Harmful somatic mutations: lessons from the dark side. *Immunol Rev, 162*, 197-209. http://www.ncbi.nlm.nih.gov/pubmed/9602365

Wijesinghe, P., & Bhagwat, A. S. (2012, Oct). Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res, 40*(18), 9206-9217. https://doi.org/10.1093/nar/gks685

Williams, M. A., & Bevan, M. J. (2007). Effector and memory CTL differentiation. *Annu Rev Immunol, 25*, 171-192. https://doi.org/10.1146/annurev.immunol.25.022106.141548

Willmann, K. L., Milosevic, S., Pauklin, S., Schmitz, K. M., Rangam, G., Simon, M. T., Maslen, S., Skehel, M., Robert, I., Heyer, V., Schiavo, E., Reina-San-Martin, B., & Petersen-Mahrt, S. K. (2012, Oct 22). A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. *J Exp Med, 209*(11), 2099-2111. https://doi.org/10.1084/jem.20112145

Wilson, M., Hsu, E., Marcuz, A., Courtet, M., Du Pasquier, L., & Steinberg, C. (1992, Dec). What limits affinity maturation of antibodies in Xenopus--the rate of somatic mutation or the ability to select mutants? *EMBO J, 11*(12), 4337-4347. http://www.ncbi.nlm.nih.gov/pubmed/1425571

Wilson, T. M., Vaisman, A., Martomo, S. A., Sullivan, P., Lan, L., Hanaoka, F., Yasui, A., Woodgate, R., & Gearhart, P. J. (2005, Feb 21). MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. *J Exp Med, 201*(4), 637-645. https://doi.org/10.1084/jem.20042066

Winkler, T. H., & Martensson, I. L. (2018). The Role of the Pre-B Cell Receptor in B Cell Development, Repertoire Selection, and Tolerance. *Front Immunol, 9*, 2423. https://doi.org/10.3389/fimmu.2018.02423

Wong, H. S., & Germain, R. N. (2018, Apr). Robust control of the adaptive immune system. *Semin Immunol, 36*, 17-27. https://doi.org/10.1016/j.smim.2017.12.009

Wong, J. B., Hewitt, S. L., Heltemes-Harris, L. M., Mandal, M., Johnson, K., Rajewsky, K., Koralov, S. B., Clark, M. R., Farrar, M. A., & Skok, J. A. (2019, Oct 18). B-1a cells acquire their unique characteristics by bypassing the pre-BCR selection stage. *Nat Commun, 10*(1), 4768. https://doi.org/10.1038/s41467-019-12824-z

Wu, X., Darce, J. R., Chang, S. K., Nowakowski, G. S., & Jelinek, D. F. (2008, Dec 01). Alternative splicing regulates activation-induced cytidine deaminase (AID): implications for suppression of AID mutagenic activity in normal and malignant B cells. *Blood, 112*(12), 4675-4682. https://doi.org/10.1182/blood-2008-03-145995

Wu, X., Tian, J., & Wang, S. (2018). Insight Into Non-Pathogenic Th17 Cells in Autoimmune Diseases. *Front Immunol, 9*, 1112. https://doi.org/10.3389/fimmu.2018.01112

Xu, L., Gorham, B., Li, S. C., Bottaro, A., Alt, F. W., & Rothman, P. (1993, Apr 15). Replacement of germ-line epsilon promoter by gene targeting alters control of immunoglobulin heavy chain class switching. *Proc Natl Acad Sci U S A, 90*(8), 3705-3709. https://doi.org/10.1073/pnas.90.8.3705

Yam-Puc, J. C., Zhang, L., Zhang, Y., & Toellner, K. M. (2018). Role of B-cell receptors for B-cell development and antigen-induced differentiation. *F1000Res, 7*, 429. https://doi.org/10.12688/f1000research.13567.1

Yang, F., Waldbieser, G. C., & Lobb, C. J. (2006, Feb 01). The nucleotide targets of somatic mutation and the role of selection in immunoglobulin heavy chains of a teleost fish. *J Immunol, 176*(3), 1655-1667. http://www.ncbi.nlm.nih.gov/pubmed/16424195

Yang, G., Miton, C. M., & Tokuriki, N. (2020, Aug). A mechanistic view of enzyme evolution. *Protein Sci, 29*(8), 1724-1747. https://doi.org/10.1002/pro.3901

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015, Jan). The I-TASSER Suite: protein structure and function prediction. *Nat Methods, 12*(1), 7-8. https://doi.org/10.1038/nmeth.3213

Yang, X., Lin, G., Han, Z., & Chai, J. (2019). Structural Biology of NOD-Like Receptors. *Adv Exp Med Biol, 1172*, 119-141. https://doi.org/10.1007/978-981-13-9367-9_6

Yang, Z. (2006, 01/01). Computational Molecular Evolution. *Yang, Z. (2006) Computational molecular evolution. Oxford Series in Ecology and Evolution . Oxford University Press, Oxford, UK. ISBN 9780198566991*. https://doi.org/10.1093/acprof:oso/9780198567028.001.0001

Yazdani, R., Fekrvand, S., Shahkarami, S., Azizi, G., Moazzami, B., Abolhassani, H., & Aghamohammadi, A. (2019, Jan). The hyper IgM syndromes: Epidemiology, pathogenesis, clinical manifestations, diagnosis and management. *Clin Immunol, 198*, 19-30. https://doi.org/10.1016/j.clim.2018.11.007

Ye, J., Bromage, E. S., & Kaattari, S. L. (2010, Jan 15). The strength of B cell interaction with antigen determines the degree of IgM polymerization. *J Immunol, 184*(2), 844-850. https://doi.org/10.4049/jimmunol.0902364

Yeap, L. S., Hwang, J. K., Du, Z., Meyers, R. M., Meng, F. L., Jakubauskaitė, A., Liu, M., Mani, V., Neuberg, D., Kepler, T. B., Wang, J. H., & Alt, F. W. (2015, Nov 19). Sequence-Intrinsic Mechanisms that Target AID Mutational Outcomes on Antibody Genes. *Cell, 163*(5), 1124-1137. https://doi.org/10.1016/j.cell.2015.10.042

Yeap, L. S., & Meng, F. L. (2019). Cis- and trans-factors affecting AID targeting and mutagenic outcomes in antibody diversification. *Adv Immunol, 141*, 51-103. https://doi.org/10.1016/bs.ai.2019.01.002

Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., & Lieber, M. R. (2003, May). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol, 4*(5), 442-451. https://doi.org/10.1038/ni919

Yu, K., & Lieber, M. R. (2019, Aug). Current insights into the mechanism of mammalian immunoglobulin class switch recombination. *Crit Rev Biochem Mol Biol, 54*(4), 333-351. https://doi.org/10.1080/10409238.2019.1659227

Yu, K., Roy, D., Bayramyan, M., Haworth, I. S., & Lieber, M. R. (2005, Mar). Fine-structure analysis of activation-induced deaminase accessibility to class switch region R-loops. *Mol Cell Biol, 25*(5), 1730-1736. https://doi.org/10.1128/mcb.25.5.1730-1736.2005

Zakas, P. M., Brown, H. C., Knight, K., Meeks, S. L., Spencer, H. T., Gaucher, E. A., & Doering, C. B. (2017, Jan). Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nat Biotechnol, 35*(1), 35-37. https://doi.org/10.1038/nbt.3677

Zan, H., & Casali, P. (2013, Mar). Regulation of Aicda expression and AID activity. *Autoimmunity, 46*(2), 83-101. https://doi.org/10.3109/08916934.2012.749244

Zandvoort, A., & Timens, W. (2002, Oct). The dual function of the splenic marginal zone: essential for initiation of anti-TI-2 responses but also vital in the general first-line defense against blood-borne antigens. *Clin Exp Immunol, 130*(1), 4-11. https://doi.org/10.1046/j.1365-2249.2002.01953.x

Zanotti, K. J., & Gearhart, P. J. (2016, Feb). Antibody diversification caused by disrupted mismatch repair and promiscuous DNA polymerases. *DNA Repair (Amst), 38*, 110-116. https://doi.org/10.1016/j.dnarep.2015.11.011

Zhang, J., Bottaro, A., Li, S., Stewart, V., & Alt, F. W. (1993, Sep). A selective defect in IgG2b switching as a result of targeted mutation of the I gamma 2b promoter and exon. *EMBO J, 12*(9), 3529-3537.

Zhang, Y. (2008, Jan 23). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics, 9*, 40. https://doi.org/10.1186/1471-2105-9-40

Zhang, Y., Garcia-Ibanez, L., & Toellner, K. M. (2016, Mar). Regulation of germinal center B-cell differentiation. *Immunol Rev, 270*(1), 8-19. https://doi.org/10.1111/imr.12396

Zhao, Y., Jackson, S. M., & Aitken, R. (2006). The bovine antibody repertoire. *Dev Comp Immunol, 30*(1-2), 175-186. https://doi.org/10.1016/j.dci.2005.06.012

Zhao, Y., Pan-Hammarstrom, Q., Zhao, Z., & Hammarstrom, L. (2005). Identification of the activation-induced cytidine deaminase gene from zebrafish: an evolutionary analysis. *Dev Comp Immunol, 29*(1), 61-71. https://doi.org/10.1016/j.dci.2004.05.005

Zheng, S., Vuong, B. Q., Vaidyanathan, B., Lin, J. Y., Huang, F. T., & Chaudhuri, J. (2015, May 7). Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA. *Cell, 161*(4), 762-773. https://doi.org/10.1016/j.cell.2015.03.020

Zhong, C., Zheng, M., & Zhu, J. (2018, Aug). Lymphoid tissue inducer-A divergent member of the ILC family. *Cytokine Growth Factor Rev, 42*, 5-12. https://doi.org/10.1016/j.cytogfr.2018.02.004

Zhu, C., & Hsu, E. (2010, Nov 01). Error-prone DNA repair activity during somatic hypermutation in shark B lymphocytes. *J Immunol, 185*(9), 5336-5347. https://doi.org/10.4049/jimmunol.1000779

Zhu, C., Tong, J., Yu, X., & Guo, W. (2015, Aug). Comparative mapping for bighead carp (Aristichthys nobilis) against model and non-model fishes provides insights into the genomic evolution of cyprinids. *Mol Genet Genomics, 290*(4), 1313-1326. https://doi.org/10.1007/s00438-015-0992-z

Zoete, V., Cuendet, M. A., Grosdidier, A., & Michielin, O. (2011, Aug). SwissParam: a fast force field generation tool for small organic molecules. *J Comput Chem, 32*(11), 2359-2368. https://doi.org/10.1002/jcc.21816

Zou, J., Wang, C., Ma, X., Wang, E., & Peng, G. (2017). APOBEC3B, a molecular driver of mutagenesis in human cancers. *Cell Biosci, 7*, 29. https://doi.org/10.1186/s13578-017-0156-4

Zwollo, P., Hennessey, E., Moore, C., Marancik, D. P., Wiens, G. D., & Epp, L. (2017, Sep). A BCWD-resistant line of rainbow trout exhibits higher abundance of IgT(+) B cells and heavy chain tau transcripts compared to a susceptible line following challenge with Flavobacterium psychrophilum. *Dev Comp Immunol, 74*, 190-199. https://doi.org/10.1016/j.dci.2017.04.019

Appendices

Appendix 1: GenBank accession number of the Ig genes used in this thesis to identify Atlantic cod IgH locus as well as WRC analysis

ID	Description	Species
Protein queries (full-length IgM and IgD):		
ACO88906.1	IgD	*Siniperca chuatsi*
BAD34541.1	IgD	*Takifugu rubripes*
AIC33830.1	IgD	*Lutjanus sanguineus*
AFI33218.1	IgD	*Epinephelus coioides*
AAX78205.1	IgM	*Epinephelus coioides*
BAB60868.1	IgM	*Paralichthys olivaceus*
A0A126CRL5	IgM	*Oreochromis niloticus*
A0A0G3VMZ6	IgM	*Gadus macrocephalus*
Full-length IgZ gene:		
ID	Description	Species
AIC33829.1	IgZ heavy chain transmembrane	*Lutjanus sanguineus*
AIC33828.1	immunoglobulin Z heavy chain	*Lutjanus sanguineus*
ADD82653.1	immunoglobulin Z heavy chain, partial	*Ctenopharyngodon idella*
ADD82655.1	secretory IgZ	*Ctenopharyngodon idella*
ABY76180.1	membrane bound IgZ, partial	*Ctenopharyngodon idella*
IgV$_H$ genes:		

ID	Description	Species
AJ274705.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997031136 (0936) Family I	*Gadus morhua*
AJ274706.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030733 (1233) Family I	*Gadus morhua*
AJ274707.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030741 (1241) Family I	*Gadus morhua*
AJ274708.1	partial mRNA for immunoglobulin heavy chain variable region clone 1997102105 (1905) Family II	*Gadus morhua*
AJ274709.1	partial mRNA for immunoglobulin heavy chain variable region clone 1997102107 (1907a) Family II	*Gadus morhua*
AJ274710.1	partial mRNA for immunoglobulin heavy chain variable region clone 1997111806 (1906) Family II	*Gadus morhua*
AJ274711.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997031139 (1139) Family IV	*Gadus morhua*
AJ274712.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297021302 (1202a) Family IV	*Gadus morhua*
AJ274713.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297021409 (1209) Family IV	*Gadus morhua*
AJ274714.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030705 (1205b) Family IV	*Gadus morhua*
AJ274715.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297021305 (1205a) Family III	*Gadus morhua*
AJ274716.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997021408 (0908) Family III	*Gadus morhua*
AJ274717.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297021402 (1202b) Family III	*Gadus morhua*
AJ274718.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297021408 (1208)	*Gadus morhua*
AJ274719.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297021411 (1211) Family III	*Gadus morhua*
AJ274720.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030702 (1202c) Family III	*Gadus morhua*
AJ274721.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030715 (1215) family III	*Gadus morhua*
AJ274722.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030722 (1222) Family III	*Gadus morhua*
AJ274723.1	partial mRNA for immunoglobulin heavy chain variable region clone 2096110714 (2014) Family III	*Gadus morhua*
AJ274724.1	partial mRNA for immunoglobulin heavy chain variable region clone 2096110629 (2029) Family III	*Gadus morhua*

AJ274725.1	partial mRNA for immunoglobulin heavy chain variable region clone 2096110631 (2031) Family III	*Gadus morhua*
AJ274726.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997021401 (0901) Family III	*Gadus morhua*
AJ274727.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997031130 (0930) Family III	*Gadus morhua*
AJ274728.1	partial mRNA for immunoglobulin heavy chain variable region clone 1998012302 (1902) Family III	*Gadus morhua*
AJ274729.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030732 (1232) Family III	*Gadus morhua*
AJ274730.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030745 (1245) Family III	*Gadus morhua*
AJ274731.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997031129 (0929) Family III	*Gadus morhua*
AJ274732.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297021304 (1204) Family III	*Gadus morhua*
AJ274733.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030714 (1214) Family III	*Gadus morhua*
AJ274734.1	partial mRNA for immunoglobulin heavy chain variable region clone 1997111807 (1907b) Family III	*Gadus morhua*
AJ274735.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997031143 (0943) Family III	*Gadus morhua*
AJ274736.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997031127 (0927) Family III	*Gadus morhua*
AJ274737.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030703 (1203b) family III	*Gadus morhua*
AJ274738.1	partial mRNA for immunoglobulin heavy chain variable region clone 1997102101 (1901) Family III	*Gadus morhua*
AJ274739.1	partial mRNA for immunoglobulin heavy chain variable region clone 1498012214 (1414) Family III	*Gadus morhua*
AJ274740.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997031134 (0934) Family III	*Gadus morhua*
AJ274741.1	partial mRNA for immunoglobulin heavy chain variable region clone 2098012010 (2010) Family III	*Gadus morhua*
AJ274742.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997031138 (0938) Family III	*Gadus morhua*
AJ274743.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997021404 (0904) Family III	*Gadus morhua*
AJ274744.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030706 (1206) Family III	*Gadus morhua*
AJ274745.1	partial mRNA for immunoglobulin heavy chain variable region clone 1497103004 (1404) Family III	*Gadus morhua*

AJ274746.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030728 (1228)	*Gadus morhua*
AJ274747.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297021303 (1203a) Family III	*Gadus morhua*
AJ274748.1	partial mRNA for immunoglobulin heavy chain variable region clone 2098011603 (2003) Family III	*Gadus morhua*
AJ274749.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030710 (1210b) Family III	*Gadus morhua*
AJ274750.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297030719 (1219) Family III	*Gadus morhua*
AJ274751.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997021407 (0907) Family III	*Gadus morhua*
AJ274752.1	partial mRNA for immunoglobulin heavy chain variable region clone 1297021310 (1210a) Family III	*Gadus morhua*
AJ274753.1	partial mRNA for immunoglobulin heavy chain variable region clone 0997031140 (0940) Family III	*Gadus morhua*
AJ274754.1	partial mRNA for immunoglobulin heavy chain variable region clone 1498020906 (1406) Family III	*Gadus morhua*
AJ274755.1	partial mRNA for immunoglobulin heavy chain variable region clone 2096110626 (2026) Family III	*Gadus morhua*
AJ274756.1	partial mRNA for immunoglobulin heavy chain variable region clone 2096102205 (2005) family III	*Gadus morhua*
AJ279353.1	partial mRNA for immunoglobulin heavy chain variable region clone 21	*Gadus morhua*
AJ279354.1	partial mRNA for immunoglobulin heavy chain variable region clone 34	*Gadus morhua*
AJ279355.1	partial mRNA for immunoglobulin heavy chain variable region clone 40	*Gadus morhua*
AJ279356.1	partial mRNA for immunoglobulin heavy chain variable region clone 49	*Gadus morhua*
AJ279357.1	partial mRNA for immunoglobulin heavy chain variable region clone 2	*Gadus morhua*
AJ279358.1	partial mRNA for immunoglobulin heavy chain variable region clone 14	*Gadus morhua*
AJ279359.1	partial mRNA for immunoglobulin heavy chain variable region clone 15	*Gadus morhua*
AJ279360.1	partial mRNA for immunoglobulin heavy chain variable region clone 29	*Gadus morhua*
AJ279361.1	partial mRNA for immunoglobulin heavy chain variable region clone 44	*Gadus morhua*
AJ279362.1	partial mRNA for immunoglobulin heavy chain variable region clone 19	*Gadus morhua*

AJ279363.1	partial mRNA for immunoglobulin heavy chain variable region clone 38	*Gadus morhua*
AJ279365.1	partial mRNA for immunoglobulin heavy chain variable region clone 39	*Gadus morhua*
AJ279366.1	partial mRNA for immunoglobulin heavy chain variable region clone 31	*Gadus morhua*
AJ279367.1	partial mRNA for immunoglobulin heavy chain variable region clone 25	*Gadus morhua*
AJ279368.1	partial mRNA for immunoglobulin heavy chain variable region clone 23	*Gadus morhua*
AJ279369.1	partial mRNA for immunoglobulin heavy chain variable region clone 263	*Gadus morhua*
AJ279370.1	partial mRNA for immunoglobulin heavy chain variable region clone 35	*Gadus morhua*
AJ279371.1	partial mRNA for immunoglobulin heavy chain variable region clone 33	*Gadus morhua*
AJ279372.1	partial mRNA for immunoglobulin heavy chain variable region clone 98	*Gadus morhua*
AJ279373.1	partial mRNA for immunoglobulin heavy chain variable region clone 9	*Gadus morhua*
AJ279374.1	partial mRNA for immunoglobulin heavy chain variable region clone 28	*Gadus morhua*
AJ279375.1	partial mRNA for immunoglobulin heavy chain variable region clone 127	*Gadus morhua*
AJ279376.1	partial mRNA for immunoglobulin heavy chain variable region clone 4	*Gadus morhua*
AJ279377.1	partial mRNA for immunoglobulin heavy chain variable region clone 11	*Gadus morhua*
AJ279378.1	partial mRNA for immunoglobulin heavy chain variable region clone 22	*Gadus morhua*
AJ279380.1	partial mRNA for immunoglobulin heavy chain variable region clone 264	*Gadus morhua*
AJ279381.1	partial mRNA for immunoglobulin heavy chain variable region clone 48	*Gadus morhua*
AJ279382.1	partial mRNA for immunoglobulin heavy chain variable region clone 109	*Gadus morhua*
AJ279383.1	partial mRNA for immunoglobulin heavy chain variable region clone 110	*Gadus morhua*
AJ279384.1	partial mRNA for immunoglobulin heavy chain variable region clone 45	*Gadus morhua*
AJ279385.1	partial mRNA for immunoglobulin heavy chain variable region clone 32	*Gadus morhua*

AJ279386.1	partial mRNA for immunoglobulin heavy chain variable region clone 244	*Gadus morhua*
AJ279387.1	partial mRNA for immunoglobulin heavy chain variable region clone 12	*Gadus morhua*
AJ279388.1	partial mRNA for immunoglobulin heavy chain variable region clone 20	*Gadus morhua*
AJ279389.1	partial mRNA for immunoglobulin heavy chain variable region clone 1	*Gadus morhua*
AJ279390.1	partial mRNA for immunoglobulin heavy chain variable region clone 30	*Gadus morhua*
AJ279391.1	partial mRNA for immunoglobulin heavy chain variable region clone 36	*Gadus morhua*
AJ279392.1	partial mRNA for immunoglobulin heavy chain variable region clone 82	*Gadus morhua*
AJ279393.1	partial mRNA for immunoglobulin heavy chain variable region clone 42	*Gadus morhua*
AJ279394.1	partial mRNA for immunoglobulin heavy chain variable region clone 8	*Gadus morhua*
AJ279395.1	partial mRNA for immunoglobulin heavy chain variable region clone 46	*Gadus morhua*
AJ279396.1	partial mRNA for immunoglobulin heavy chain variable region clone 17	*Gadus morhua*
AJ279397.1	partial mRNA for immunoglobulin heavy chain variable region clone 90	*Gadus morhua*
DQ230541.1	clone 1B07AVH1 CS3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230547.1	clone 3B11AVH1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230550.1	clone 3D04AVH1 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230551.1	clone 3D08AVH1 CS4 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230552.1	clone 3E09AVH1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230553.1	clone 3F02AVH1 CS3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230555.1	clone 3G07AVH1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230557.1	clone 3G12AVH1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230558.1	clone 6E04AVH1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*

DQ230560.1	clone 6G04AVH1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230562.1	clone 6H05AVH1 CS4 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
AY238358.1	immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492547.1	clone 15B02VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492548.1	clone 15B03VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492549.1	clone 15B04VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492550.1	clone 15B05VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492551.1	clone 15B06VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492558.1	clone 15C08VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492557.1	clone 15C07VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492554.1	clone 15B12VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492555.1	clone 15C02VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492553.1	clone 15B10VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492587.1	clone 19D12VH1PBL CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492591.1	clone 15D01VH1AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492596.1	clone 15D10VH1AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492595.1	clone 15D07VH1AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492597.1	clone 15D11VH1AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492590.1	clone 15C11VH1AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492598.1	clone 15D12VH1AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492594.1	clone 15D06VH1AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*

EU492592.1	clone 15D02VH1AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492637.1	clone 15E01VH1SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492642.1	clone 15E07VH1SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492643.1	clone 15E09VH1SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492641.1	clone 15E06VH1SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492695.1	clone 15A01VH1GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492696.1	clone 15A02VH1GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492697.1	clone 15A04VH1GL CS3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492698.1	clone 15A06VH1GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492699.1	clone 15A07VH1GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492700.1	clone 15A08VH1GL CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492701.1	clone 15A10VH1GL CS3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492702.1	clone 15A11VH1GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492733.1	clone 18D12VH1GL CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492763.1	clone 18C11VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	
EU492734.1	clone 18E01VH1GL CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492746.1	clone 18C12VH1SK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492747.1	clone 3F02AVH1 CS3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492750.1	clone 19B04VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492751.1	clone 19B09VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492835.1	clone 19G12VH1I3 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*

EU492837.1	clone 20B06VH1I3 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492838.1	clone 20B12VH1I3 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492840.1	clone 20F07VH1I3 CS6 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492845.1	clone 21D11VH1I3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492851.1	clone 21H11VH1I3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492850.1	clone 21H10VH1I3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492844.1	clone 21D08VH1I3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492842.1	clone 21D06VH1I3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492795.1	clone 19A07VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492796.1	clone 19A10VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492797.1	clone 19A11VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492802.1	clone 20A01VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492805.1	clone 21C10VH1I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492806.1	clone 21C11VH1I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492807.1	clone 21D01VH1I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492808.1	clone 21D03VH1I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492809.1	clone 21D04VH1I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492813.1	clone 19E02RevI2 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492816.1	clone 15G05VH1I3 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492818.1	clone 15H07VH1I3 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492819.1	clone 15H11VH1I3 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*

EU492820.1	clone 15H12VH1I3 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492843.1	clone 21D07VH1I3 CS6 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492752.1	clone 19B10VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492754.1	clone 20A06VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492755.1	clone 20A07VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230542.1	clone 1B10AVH1 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230544.1	clone 1F07AVH1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230548.1	clone 3C11AVH1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
GU296460.1	clone 5r21 immunoglobulin delta heavy chain membrane bound form mRNA, partial cds	*Ictalurus punctatus*
EU492836.1	clone 19H09VH1I3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492834.1	clone 19C10VH1I3 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492794.1	clone 19A02VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230556.1	clone 3G11AVH1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492552.1	clone 15B07VH1PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492640.1	clone 15E05VH1SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492761.1	clone 21C07VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492759.1	clone 21C05VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492762.1	clone 21H01VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492815.1	clone 15F08VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492756.1	clone 20A10VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492757.1	clone 20A11VH1SK CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*

EU492764.1	clone 15F02VH1I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492765.1	clone 15F06VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492766.1	clone 15F07VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492767.1	clone 15F11VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492768.1	clone 15F12VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492769.1	clone 15G02VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492770.1	clone 15G03VH1I2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230559.1	clone 6F04AVH1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
M58673.1	Ig heavy chain mRNA V-region clone NG64	*Ictalurus punctatus*
DQ230561.1	clone 6G05AVH1 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492772.1	clone 16F07VH2I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492773.1	clone 16F08VH2I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492774.1	clone 16F10VH2I2 CS7 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492776.1	clone 16F12VH2I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230565.1	clone 1D11AVH2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230567.1	clone 2C05AVH2 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230568.1	clone 2F06AVH2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230571.1	clone 3B10AVH2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230572.1	clone 3C07AVH2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230573.1	clone 3C12AVH2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230574.1	clone 3D09AVH2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*

DQ230576.1	clone 3D11AVH2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230577.1	clone 6A03AVH2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230579.1	clone 6D03AVH2 CS3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230580.1	clone 6D06AVH2 CS3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230581.1	clone 6E05AVH2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492749.1	clone 18E03VH2SK CS3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492777.1	clone 16G01VH2I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492704.1	clone 16A01VH2GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492648.1	clone 16E04VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492649.1	clone 16E05VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492650.1	clone 16E07VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492651.1	clone 16E08VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492652.1	clone 16E09VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492653.1	clone 16E10VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492655.1	clone 16E12VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492656.1	clone 16F01VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492657.1	clone 16F02VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492658.1	clone 16F03VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492585.1	clone 20C02RevPBL CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492589.1	clone 19E03VH2PBL CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492600.1	clone 16D01VH2AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*

EU492602.1	clone 16D06VH2AK CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492603.1	clone 16D07VH2AK CS5 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492604.1	clone 16D08VH2AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492605.1	clone 16D09VH2AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492606.1	clone 16D10VH2AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492607.1	clone 16D11RevAK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492608.1	clone 16D12VH2AK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492633.1	clone 19E04VH2AK CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492638.1	clone 15E03VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492639.1	clone 15E04VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492647.1	clone 16E03VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492659.1	clone 16F04VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492660.1	clone 16F05VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492691.1	clone 20C04RevSP CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492692.1	clone 20E01RevSP CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492705.1	clone 16A02VH2GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492706.1	clone 16A03VH2GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492708.1	clone 16A05VH2GL CS9 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492709.1	clone 16A06VH2GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492710.1	clone 16A07VH2GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492711.1	clone 16A09VH2GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*

EU492712.1	clone 16A11VH2GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492713.1	clone 16A12VH2GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492735.1	clone 20F02VH2GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230582.1	clone 6F06AVH2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230583.1	clone 6G06AVH2 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230584.1	clone 6H06AVH2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
AY238359.1	immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492559.1	clone 16B04VH2PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492560.1	clone 16B06VH2PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492561.1	clone 16B08VH2PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492562.1	clone 16B10VH2PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492563.1	clone 16B12VH2PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492564.1	clone 16C01VH2PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492565.1	clone 16C02VH2PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492566.1	clone 16C03VH2PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492821.1	clone 16H03VH2I3 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492568.1	clone 16C05VH2PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492822.1	clone 16H09VH2I3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492812.1	clone 20F03RevI2 CS8 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492736.1	clone 16A08RevI2 CS8 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492760.1	clone 21C06VH2SK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*

EU492778.1	clone 16G02VH2I2 CS9 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492853.1	clone 16H05VH2I3 CS5 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230570.1	clone 3B05AVH2 CS4 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230575.1	clone 3D10AVH2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
DQ230578.1	clone 6B05AVH2 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492599.1	clone 16C12VH2AK CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492654.1	clone 16E11VH2SP immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492694.1	clone 19E11VH2SP CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492758.1	clone 20C09VH2SK immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492810.1	clone 16G06RevI2 CS2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492823.1	clone 16H10VH2I3 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492846.1	clone 21H05VH2I3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492847.1	clone 21H06RevI3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492849.1	clone 21H09VH2I3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492852.1	clone 16H07RevI3 CS1 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
M58670.1	Ig heavy chain mRNA V-region clone NG22	*Ictalurus punctatus*
EU492779.1	clone 16G04VH2I2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
M58675.1	Ig heavy chain mRNA V-region, clone NG77	*Ictalurus punctatus*
EU492841.1	clone 20G04VH2I3 CS6 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492780.1	clone 16G08VH2I2 CS7 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492707.1	clone 16A04VH2GL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*

EU492569.1	clone 16C06VH2PBL immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
EU492601.1	clone 16D04VH2AK CS6 immunoglobulin heavy chain variable region mRNA, partial cds	*Ictalurus punctatus*
AF273412.1	clone 01-09 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273415.1	clone 05-12 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273410.1	clone 04-09 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273411.1	clone 01-06 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273413.1	clone 01-10 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273414.1	clone 05-11 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273416.1	clone 08-06 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273417.1	clone 08-07 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273418.1	clone 10-11 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF269076.1	clone d27 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF269078.1	clone 09-10 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273398.1	clone 08-01 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273399.1	clone 08-10 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273396.1	clone 04-04 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273397.1	clone 07-02 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF269079.1	clone 02-04 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF269080.1	clone 02-10 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF269081.1	clone 06-07 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF269082.1	clone 09-04 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*

AF269083.1	clone 09-06 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273429.1	clone 07-04 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF269084.1	clone 09-14 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273425.1	clone 03-08 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273426.1	clone 10-08 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273419.1	clone 01-03 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273421.1	clone 08-05 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273422.1	clone 09-15 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273423.1	clone 09-12 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273424.1	clone 10-02 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273420.1	clone 07-01 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273427.1	clone 02-05 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273428.1	clone 03-06 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273430.1	clone 08-03 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273431.1	clone 09-03 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF273432.1	clone 10-10 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AF269085.1	clone 10-06 immunoglobulin heavy chain variable region (IgH) mRNA, partial cds	*Salmo salar*
AY646275.1	isolate 4-8.2.1 immunoglobulin zeta heavy chain mRNA, partial cds	*Danio rerio*
AY646273.1	isolate 4-3.2.1 immunoglobulin zeta heavy chain mRNA, partial cds	*Danio rerio*
AY646274.1	isolate 4-6.0.2 immunoglobulin zeta heavy chain mRNA, partial cds	*Danio rerio*
AY646252.1	isolate 4-8.3.5 immunoglobulin mu heavy chain mRNA, partial cds	*Danio rerio*

AY646251.1	isolate 4-6.5.5 immunoglobulin mu heavy chain mRNA, partial cds	*Danio rerio*
AY646250.1	isolate 4-6.4.2 immunoglobulin mu heavy chain mRNA, partial cds	*Danio rerio*
AF273884.1	clone VH124 immunoglobulin heavy chain variable region mRNA, partial cds	*Danio rerio*
AF273876.1	clone VH101 immunoglobulin heavy chain variable region mRNA, partial cds	*Danio rerio*
AF273877.1	clone VH103 immunoglobulin heavy chain variable region mRNA, partial cds	*Danio rerio*
AF273878.1	clone VH119 immunoglobulin heavy chain variable region mRNA, partial cds	*Danio rerio*
AF273882.1	clone VH23 immunoglobulin heavy chain variable region mRNA, partial cds	*Danio rerio*
AF273885.1	clone VH88 immunoglobulin heavy chain variable region mRNA, partial cds	*Danio rerio*
AF273880.1	clone VH114 immunoglobulin heavy chain variable region mRNA, partial cds	*Danio rerio*
AF273886.1	clone VH350-6 immunoglobulin heavy chain variable region mRNA, partial cds	*Danio rerio*
AF273889.1	clone VH350-3 immunoglobulin heavy chain variable region mRNA, partial cds	*Danio rerio*
AY646245.1	isolate 1-2.1.1 immunoglobulin mu heavy chain mRNA, partial cds	*Danio rerio*
DQ106021.1	isolate A variant immunoglobulin heavy chain variable region gene, partial cds	*Danio rerio*
AF273897.1	clone VHE1 immunoglobulin heavy chain variable region gene, partial cds	*Danio rerio*
DQ106019.1	isolate A immunoglobulin heavy chain variable region gene, partial cds	*Danio rerio*
AY646263.1	isolate 1-2.2.1 immunoglobulin zeta heavy chain mRNA, partial cds	*Danio rerio*
AY646264.1	isolate 1-1.2.1 immunoglobulin zeta heavy chain mRNA, partial cds	*Danio rerio*
AY646267.1	isolate 1-2.1.1 immunoglobulin zeta heavy chain mRNA, partial cds	*Danio rerio*
AY608342.1	isolate MaryM7 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608355.1	isolate MaryM33 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608358.1	isolate JosefM2 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*

AY608362.1	isolate JosefM7 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608373.1	isolate JosefM21 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608376.1	isolate JosefM27 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608386.1	isolate M17 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608392.1	isolate M29 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608397.1	isolate M34 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY609272.1	clone 72S immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY609265.1	clone 47S immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY609266.1	clone 49S immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359839.1	clone G5G2-13 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359840.1	clone G5G2-16 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359841.1	clone G5G2-17 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359843.1	clone G5G2-33 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359844.1	clone G5G2-9 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359845.1	clone G5G2-13-2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359846.1	clone G5G2-B immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359832.1	clone G2G5-E11 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359857.1	clone G2G5-C2 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359856.1	clone G2G5-B9 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359858.1	clone G2G5-C6 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608337.1	isolate MaryM2 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*

AY608339.1	isolate MaryM4 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608340.1	isolate MaryM5 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608341.1	isolate MaryM6 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608346.1	isolate MaryM12 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608347.1	isolate MaryM13 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608349.1	isolate MaryM15 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608351.1	isolate MaryM17 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608353.1	isolate MaryM31 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608354.1	isolate MaryM32 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608356.1	isolate MaryM34 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608357.1	isolate JosefM1 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608361.1	isolate JosefM6 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608363.1	isolate JosefM11 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608364.1	isolate JosefM12 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608365.1	isolate JosefM13 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608366.1	isolate JosefM14 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608367.1	isolate JosefM15 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608368.1	isolate JosefM16 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608369.1	isolate JosefM17 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608370.1	isolate JosefM18 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608371.1	isolate JosefM19 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*

AY608372.1	isolate JosefM20 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608374.1	isolate JosefM22 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608375.1	isolate JosefM26 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608377.1	isolate M3 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608378.1	isolate M4 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608379.1	isolate M5 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608381.1	isolate M8 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608382.1	isolate M9 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608383.1	isolate M13 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608384.1	isolate M14 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608385.1	isolate M15 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608387.1	isolate M19 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608389.1	isolate M21 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608390.1	isolate M24 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608391.1	isolate M25 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608393.1	isolate M30 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608394.1	isolate M31 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608395.1	isolate M32 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608396.1	isolate M33 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608398.1	isolate M35 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608400.1	isolate M39 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*

AY608401.1	isolate M41 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608403.1	isolate M43 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY608404.1	isolate M47 immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY609249.1	clone 2S immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY609254.1	clone 21S immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359842.1	clone G5G2-31 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY609258.1	clone 27S immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ282627.1	clone G2G5-34 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY609264.1	clone 46S immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
AY609259.1	clone 29S immunoglobulin mu heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359827.1	clone G4G5-3 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359826.1	clone G4G5-17 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359833.1	clone G4G5-E30 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359828.1	clone G4G5-4 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359830.1	clone G4G5-39 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359829.1	clone G4G5-33 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359831.1	clone G4G5-66 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359835.1	clone G4G2-33 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359836.1	clone G4G2-41 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359834.1	clone G4G5-E35 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359837.1	clone G4G2-54 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*

AY609252.1	clone 15S	*Ginglymostoma cirratum*
GQ359848.1	clone G4G5-46 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359849.1	clone G4G5-76 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359850.1	clone G4G5-81 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359855.1	clone G4G5-A21 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359851.1	clone G4G5-88 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
JQ272797.1	clone 4-7 IgM G4 VDJ switch to G5 C-region mRNA sequence	*Ginglymostoma cirratum*
JQ272798.1	clone 4-21 IgM G4 VDJ switch to G5 C-region mRNA sequence	*Ginglymostoma cirratum*
JQ272799.1	clone 4-36 IgM G4 VDJ switch to G5 C-region mRNA sequence	*Ginglymostoma cirratum*
JQ272805.1	clone I7 IgM G4 VDJ switch to G5 C-region mRNA sequence	*Ginglymostoma cirratum*
JQ272806.1	clone I16 IgM G4 VDJ switch to G5 C-region mRNA sequence	*Ginglymostoma cirratum*
JQ272808.1	clone I29 IgM G4 VDJ switch to G5 C-region mRNA sequence	*Ginglymostoma cirratum*
JQ272809.1	clone I36 IgM G4 VDJ switch to G5 C-region mRNA sequence	*Ginglymostoma cirratum*
JQ272810.1	clone I53 IgM G4 VDJ switch to G5 C-region mRNA sequence	*Ginglymostoma cirratum*
JQ272812.1	clone I69 IgM G4 VDJ switch to G5 C-region mRNA sequence	*Ginglymostoma cirratum*
JQ272815.1	clone I-167 IgM G4 VDJ switch to G5 C-region mRNA sequence	*Ginglymostoma cirratum*
JQ272821.1	clone 61 IgM G5 VDJ switch to G4 C-region mRNA sequence	*Ginglymostoma cirratum*
GQ359852.1	clone G2G5-F27 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
GQ359853.1	clone G4G5-C33 immunoglobulin heavy chain variable region mRNA, partial cds	*Ginglymostoma cirratum*
JF507607.1	clone T0923W2J05 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507611.1	clone T1023W2J12 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*

JF507612.1	clone T1123W2J12 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507613.1	clone T0419W2J12 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507614.1	clone T0423W2J12 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507615.1	clone T1323W2J12 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507616.1	clone T0523W2J24 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507617.1	clone T0123W2J24 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507620.1	clone S1523W2J06 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507625.1	clone S0916W2J08 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507627.1	clone S2023W2J09 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507629.1	clone S1823W2J12 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507637.1	clone V1419W2J08 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507640.1	clone V1319W2J08 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507646.1	clone V1924W2J12 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507647.1	clone V1724W2J12 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507648.1	clone V2424W2J12 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507659.1	clone V1219W2J24 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
JF507660.1	clone V1424W2J25 IgWV TCR delta trans-rearrangement (TCRD) mRNA, partial cds	*Ginglymostoma cirratum*
KC920802.1	clone V5 secreted IgW heavy chain mRNA, partial cds	*Ginglymostoma cirratum*
KC920803.1	clone c1 secreted IgW heavy chain mRNA, partial cds	*Ginglymostoma cirratum*
AY524282.1	clone 7 1-2 immunoglobulin IgW short mRNA complete cds	*Ginglymostoma cirratum*
AY524295.1	clone L immunoglobulin IgW-like mRNA complete sequence	*Ginglymostoma cirratum*

LC000730.1	IGHV2S19 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000729.1	IGHV2S18 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
AB125608.1	IgVH mRNA for immunoglobulin heavy chain variable region partial cds clone: F-m161	*Takifugu rubripes*
LC000719.1	IGHV2S7 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
AB125607.1	IgVH mRNA for immunoglobulin heavy chain variable region partial cds clone: F-m146	*Takifugu rubripes*
AB125606.1	IgVH mRNA for immunoglobulin heavy chain variable region partial cds clone: F-m118	*Takifugu rubripes*
AB217624.1	IgM mRNA for immunoglobulin mu heavy chain partial cds clone: IgM_36	*Takifugu rubripes*
XM_011621003.1	Ig mu chain C region membrane-bound form (LOC445921) mRNA	*Takifugu rubripes*
LC000729.1	IGHV2S18 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000728.1	IGHV2S17 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000724.1	IGHV2S12 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000720.1	IGHV2S8 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000719.1	IGHV2S7 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000718.1	IGHV2S6 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000717.1	IGHV2S1 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000721.1	IGHV2S9 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000722.1	IGHV2S10 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000723.1	IGHV2S11 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000726.1	IGHV2S15 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000727.1	IGHV2S16 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000730.1	IGHV2S19 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*

LC000731.1	IGHV2S20 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
AB217616.1	IgH mRNA for Immunoglobulin heavy chain partial cds clone: IgH_4	*Takifugu rubripes*
AB159481.1	IgD mRNA for immunoglobulin D complete cds	*Takifugu rubripes*
AB217618.1	IgH mRNA for immunoglobulin heavy chain partial cds clone: IgH_6	*Takifugu rubripes*
AB125605.1	IgVH mRNA for immunoglobulin heavy chain variable region partial cds clone: F-m116	*Takifugu rubripes*
AB217620.1	IgH mRNA for immunoglobulin heavy chain partial cds clone: IgH_20	*Takifugu rubripes*
AB125604.1	IgVH mRNA for immunoglobulin heavy chain variable region partial cds clone: F-m106	*Takifugu rubripes*
LC000713.2	IGHV1S17 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000716.1	IGHV1S21 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000700.1	IGHV1S4 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000714.1	IGHV1S18 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000711.1	IGHV1S15 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000710.1	IGHV1S14 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000708.1	IGHV1S12 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000704.1	IGHV1S8 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000703.1	IGHV1S7 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000699.1	IGHV1S3 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000698.1	IGHV1S2 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000697.1	IGHV1S1 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000715.1	IGHV1S19 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000700.1	IGHV1S4 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*

LC000709.1	IGHV1S13 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*
LC000712.1	IGHV1S16 gene immunoglobulin heavy chain partial sequence	*Takifugu rubripes*

Appendix 2: Pairwise Comparisons of substrate specificity of Hs-AID using independent samples Kruskal-Wallis test

Sample 1 *vs.* Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
TGC *vs*. WRC	8.444	9.863	.856	.392	1.000
AGC *vs*. WRC	-5.222	9.863	-.529	.596	1.000
TAC *vs*. WRC	-3.222	9.863	-.327	.744	1.000
TGC *vs*. non-WRC	40.889	9.863	4.146	.000	.001
AGC *vs*. non-WRC	27.222	9.863	2.760	.006	.162
TAC *vs*. non-WRC	29.222	9.863	2.963	.003	.085
GGC *vs*. WRC	-40.222	9.863	-4.078	.000	.001
GTC *vs*. WRC	-31.889	9.863	-3.233	.001	.034
GAC *vs*. WRC	-25.222	9.863	-2.557	.011	.295
GGC *vs*. non-WRC	-7.778	9.863	-.789	.430	1.000
GTC *vs*. non-WRC	.556	9.863	.056	.955	1.000
GAC *vs*. non-WRC	7.222	9.863	.732	.464	1.000
WRC *vs*. non-WRC	32.444	6.974	4.652	.000	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

[a]. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Appendix 3: Pairwise Comparisons of substrate specificity of Dr-AID using independent samples Kruskal-Wallis test

Sample 1 *vs.* Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
TGC *vs*. WRC	18.444	9.860	1.871	.061	1.000
AGC *vs*. WRC	-.889	9.860	-.090	.928	1.000
TAC *vs*. WRC	-17.556	9.860	-1.780	.075	1.000
TGC *vs*. non-WRC	41.556	9.860	4.214	.000	.001
AGC *vs*. non-WRC	22.222	9.860	2.254	.024	.678
TAC *vs*. non-WRC	5.556	9.860	.563	.573	1.000
GGC *vs*. WRC	-37.889	9.860	-3.843	.000	.003
GTC *vs*. WRC	-33.222	9.860	-3.369	.001	.021
GAC *vs*. WRC	-.889	9.860	-.090	.928	1.000
GGC *vs*. non-WRC	-14.778	9.860	-1.499	.134	1.000
GTC *vs*. non-WRC	-10.111	9.860	-1.025	.305	1.000
GAC *vs*. non-WRC	24.889	9.860	2.524	.012	.325
WRC *vs*. non-WRC	23.111	6.972	3.315	.001	.026

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

[a]. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Appendix 4: Pairwise Comparisons of substrate specificity of Ip-AID using independent samples Kruskal-Wallis test

Sample 1 *vs.* Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
TGC *vs*. WRC	14.556	9.629	1.512	.131	1.000
AGC *vs*. WRC	2.556	9.629	.265	.791	1.000
TAC *vs*. WRC	-17.111	9.629	-1.777	.076	1.000
TGC *vs*. non-WRC	36.778	9.629	3.819	.000	.004
AGC *vs*. non-WRC	24.778	9.629	2.573	.010	.282
TAC *vs*. non-WRC	5.111	9.629	.531	.596	1.000
GGC *vs*. WRC	-34.111	9.629	-3.542	.000	.011
GTC *vs*. WRC	-25.778	9.629	-2.677	.007	.208
GAC *vs*. WRC	-6.778	9.629	-.704	.482	1.000
GGC *vs*. non-WRC	-11.889	9.629	-1.235	.217	1.000
GTC *vs*. non-WRC	-3.556	9.629	-.369	.712	1.000
GAC *vs*. non-WRC	15.444	9.629	1.604	.109	1.000
WRC *vs*. non-WRC	22.222	6.809	3.264	.001	.031

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

[a]. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Appendix 5: Pairwise Comparisons of substrate specificity of Gm-AID using independent samples Kruskal-Wallis test

Sample 1 *vs.* Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
TGC *vs*. WRC	13.222	9.863	1.341	.180	1.000
AGC *vs*. WRC	-2.444	9.863	-.248	.804	1.000
TAC *vs*. WRC	-10.778	9.863	-1.093	.274	1.000
TGC *vs*. non-WRC	42.778	9.863	4.337	.000	.000
AGC *vs*. non-WRC	27.111	9.863	2.749	.006	.167
TAC *vs*. non-WRC	18.778	9.863	1.904	.057	1.000
GGC *vs*. WRC	-42.444	9.863	-4.303	.000	.000
GTC *vs*. WRC	-27.111	9.863	-2.749	.006	.167
GAC *vs*. WRC	-19.111	9.863	-1.938	.053	1.000
GGC *vs*. non-WRC	-12.889	9.863	-1.307	.191	1.000
GTC *vs*. non-WRC	2.444	9.863	.248	.804	1.000
GAC *vs*. non-WRC	10.444	9.863	1.059	.290	1.000
WRC *vs*. non-WRC	29.556	6.974	4.238	.000	.001

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

[a]. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Appendix 6: List of bony fish species studied in this thesis. Basic habitat information was retrieved from FishBase database (www.fishbase.se).

Species	IUCN Red List Status	Habitat	Depth range (usual range) (m)	Temperature	Distribution	Comments
G. morhua	VU (1996)	Marine	0 – 600 (150 – 200)	0°C – 15°C	83°N – 35°N; 95°W – 86°E	
T. chalcogramma	NE	Marine	? – 1280 (30 – 400)	Polar	68°N – 34°N; 129°E – 120°W	
B. saida	NE	Marine	0 – 400	Polar -2°C – 8°C	87°N – 52°N; 180°W – 180°E	
A. glacialis	NE	Marine	0 – 1000	Deep-water	87°N – 69°N; 130°W – 150°E	
M. merlangus	LC (2013)	Marine	10 – 200 (30 – 100)	Temperate	72°N – 35°N; 27°W – 42°E	
M. aeglefinus	VU (1996)	Marine	10 – 450 (10 – 200)	Temperate	79°N – 35°N; 76°W – 52°E	
P. virens	NE	Marine	37 – 364	Temperate	77°N – 33°N; 76°W – 35°E	
G. argenteus	NE	Marine	100 – 1000	Temperate	74°N – 24°N; 18°W – 17°E	
T. minutus	NE	Marine	1 – 440 (15 – 200)	Temperate	66°N – 28°N; 13°W – 36°E	
B. brosme	NE	Marine	18 – 1000 (18 – 549)	Temperate	83°N – 37°N; 75°W – 57°E	
M. molva	NE	Marine	100 – 1000 (100 – 400)	Temperate	75°N – 35°N; 55°W – 44°E	
L. lota	LC (2012)	Freshwater	1 – 700	Temperate 4°C – 18°C	78°N – 40°N; 180°W – 180°E	The only member of Lotidae family which lives in freshwater.
P. phycis	LC (2015)	Marine	13 – 614 (100 – 200)	Subtropical	45°N – 13°N; 32°W – 36°E	
P. blennoides	NE	Marine	10 – 1200 (100 – 450)	Temperate	69°N – 20°N; 29°W – 36°E	
M. occidentalis	LC (2009)	Marine	140 – 1945 (300 – 500)	Deep-water	43°N – 37°S; 98°W – 13°E	In tropical and warm-temperate waters.
M. berglax	NE	Marine	100 – 1000 (300 – 500)	Temperate 0°C – 4°C	82°N – 37°N; 95°W – 61°E	
B. melanobranchus	LC (2012)	Marine	400 – 2600 (700 – 1400)	Deep-water	53°N – 34°S; 98°W – 20°E	

Species	IUCN Red List Status	Habitat	Depth range (usual range) (m)	Temperature	Distribution	Comments
L. laureysi	LC (2012)	Marine	200 – 618 (300 – ?)	Deep-water	8°N – 8°S; 13°W – 12°E	
M. mora	LC (2013)	Marine	450 – 2500	Deep-water	64°N – 51°S; 77°W – 174°W	
T. murrayi	NE	Marine	0 – 1630 (500 – 1630)	Temperate	65°N – 42°N; 71°W – 0°E	
T. scabrus	LC (2012)	Marine	395 – 1700	Deep-water	55°N – 27°S; 26°W – 36°E	
M. marmoratus	NE	Marine	30 – 1600	Polar	44°S – 56°S; 39°E – 76°E	
M. zugmayeri	LC (2012)	Marine	99 – 3000	Deep-water	60°N – 49°S; 81°W – 153°W	In tropical and temperate waters; rare in the temperate northeast Atlantic
M. merluccius	LC (2015)	Marine	30 – 1075 (70 – 400)	Temperate	76°N – 18°N; 30°W – 42°E	
B. cantori	LC (2013)	Marine	450 – 475	Deep-water	Western Atlantic: Cariaco Trench, Venezuela, Gulf of Mexico to southern Brazil	
S. chardatus	LC (2013)	Marine	300 – 800	Deep-water	Tropical to subtropical in all oceans.	
C. roseus	LC (2014)	Marine	150 – 730 (330 – 690)	Deep-water		
Z. faber	DD (2013)	Marine	5 – 400 (50 – 150)	Temperate	75°N – 49°S; 17°W – 177°E	
T. subterraneus	NT (2012)	Freshwater		Temperate	39°N – 34°N	
P. transmontana	LC (2012)	Freshwater		Temperate	44°N – 43°N	
P. japonica	NE	Marine	160 – 628	Deep-water	40°N – 6°N; 97°E – 154°W	
S. salar	LR/LC (1996)	Marine	0 – 210 (10 – 23)	Temperate 2°C - 9°C	72°N – 40°N; 80°W – 61°E	
D. rerio	LC (2009)	Freshwater		Tropical 18°C - 24°C	33°N – 8°N; 66°E – 98°E	
T. rubripes	NT	Marine		Temperate	46°N – 21°N; 119°E – 142°E	

Species	IUCN Red List Status	Habitat	Depth range (usual range) (m)	Temperature	Distribution	Comments
	(2011)					
O. latipes	NE	Freshwater		Subtropical 18°C - 24°C	55°N – 10°N; 85°E – 105°E	

Appendix 7: Nucleotide sequence of aicda homologs examined in this thesis

Species name	DNA sequence
Acanthochaenus luetkenii	ATGATTACAAAACTAGACCGTGTGCTTTTGGCCAAGGAAACGTTCATCTTCCATTATGAGAACATGCGCTGGGCAAAAGGTCGGCATGAGACATAC CTCTGCTTTGTAGTGAAGAGGCGGGTGGGGCCAGACTCCCTGTCCTTTGACTTTGGACACCTCCGCAACCGCACTGGCTGCCATGTAGAGCTGCTGT TCCTGCGCCACCTGGGAACCTTGTGCCCTGGACTGTGGGGGTACGGAGGCGCTGGAGAGAGGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG GTCCCCCTGCGCTGACTGCGCCTTCAGAGTGGCCCAGTTAATCGGCCGGACGCCCAACCTCCGCCTCAGGATCTTCGTCTCTCGCCTCTACTTCTGCG ACCTGGAGGACAGCCGCGAGAGAGGGGGCCTGAGGTTGCTGAAGAAAGCTGGCGTGCAGATCACTGTCATGAGCTACAAAGACTTTTTCTATTGCT GGCAGACCTTTGTGGCTAATGGAGGGAGCAGCTTCAAGGCCTGGGACGAGATGCACCAAAACTCTGTTCGCCTGGCCAGCCAACTCAACCACATCC TGCAGCCATGTGATACAGAGGACTTAAGAGATGCATTCAAGCTTCTTGGTCTGTGA
Anabas testudineus	ATGATTACAAAGCTAGACAGTGTGCTTTTGCCCCGAAAGAAGTTTATCTACCATTACAAGAATGTGCGCTGGGCGAGGGGTCGTCATGAAACATAC CTCTGTTTCGTAGTGAAGAGGCGGGTGGGCCCAGACTCCTTGACCTTTGACTTTGGACACCTCCGCAATCGCAATGGCTGCCATGTGGAGATGCTGT TCTTGCGCTATCTGGGAGCCTTATGTCCTGGTATTTGGGGGTACGGAGGTGCTGGAGAGAAAAGGCTCAGTTACTCAATTACCTGGTTCTGTTCCTG GTCTCCTTGTGCCAACTGCTCCCTTAGGCTGACCCAGTTCCTCAGTCAGACCCCCAACCTCCGCCTCAGGATCTTTGTGTCCCGCCTTTACTTCTGTG ACATGGAGGACAGCCGCGAGCGGGAGGGTCTGAGGATACTGAAAAATGCTGGCGTGCAGATCACAGTCATGACTTACAAAGACTTCTTCTATTGCT GGCAGACCTTTGTGGATCGTAAACAGAGCAGCTTCAAAGCGTGGGATGAGCTGCACCAAAACTCTGTTCGCCTCACCAGAAAACTCTACCGCATCC TTCAGCCCTGTGAAATAGAAGATTTAAGAGATGCCTTCAAGCTTCTTGGGCTGTGA
Antennarius striatus	ATGATTACGAAGCTTGACAGCGTGCTTTTGCCCCGAAAAAAGTTCATCTACCATTATAAGAACATGCGCTGGGCGAGAGGCCGGTGTGAGACGTAC CTCTGCTTTGTAGTGAAGAGACGAGAGGGGCCAGACACCTTAACTTTTGACTTTGGACACCTCCGTAATCGCAATGGCTGTCATGTGGAGCTACTTT TCTTACGCTATCTGGGGGCCTTGTGCCCTGGATTGTGGGGCAGTGGGGGTACTGGGGAGAAGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG GTCTCCCTGTGCCAACTGTTCCATCAGACAGTGTGAATTCCTGAGCCGAACGCCCAACCTTCGCCTCAGGATCTTTGTCTCTCGTTTGTACTTCTGCG ACCTGGAGGATAGCCGTGAAAGGGAAGGCCTAAGAATGCTGAAGAAAGCCGGCGTGCAGATCTCAGTCATGAGTTACAAAGACTTCTTCTACTGCT GGCAGACCTTTGTGGCTAGTAAACAAAGTAGTTTCAAGGCTTGGGAAGAGCTGCATCAAAATTCAGTACGCCTTGCCAGAAAACTGAACCGCATCC TCCAGCCGTGTGAAGCTGAAGATTTAAGAGATGCCTTTAAGCTTCTTGGACTGTGA
Arctogadus glacialis	ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAAAATAAATTCATCTACAATTACAAGAACATGCGATGGGCAAAAGGCCGCAACGAGACCTAT CTCTGCTTCGTAATGAAGAGAAGGCTTGGACCTGATTCCCTCTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGCAGAGCTGCTGT TCCTGAGCTACCTGGGGGCGCTGTGCCCGGGCCTCTGGGGCTGCGCAGACGACAGAAACCGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT GGTCGCCCTGTGCCAACTGTGCGACCACGCTGACCCGGTTCCTGAGGCAGACACCAAACCTGCGACTCAGGATCTTCGTGTCTCGCCTCTACTTTTG TGACCTGGAGGGCAGTCCGCATGTAGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAAGACTACTTCTACT GCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTATGTGCGTCTGTCAAGAAAAAANCNAAACC GCATCCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACCTGA
Astyanax mexicanus	ATGACGAGCAAGCTGGACAGCATTCTGCTCACCCAGAAGAAGTTTATCTATCACTACAAGAACGTGCGCTGGGCTCGTGGGAGGCATGAGACTTAC CTCTGCTTCGTGGTGAAGAGGCGAATCGGACCAAACTCGCTGTCCTTCGACTTCGGGCACCTGCGCAACCGCTCCGGCTGCCACGTGGAGCTCCTCT TCCTGCGCTACCTGGGGGCACTGTGCCCGGGCCTGGGGGGTCTGGGTGTGGACGGAGTGAAGGTGGGCTACGCTGTGACCTGGTTCTGCTCATGGT CGCCCTGCTCTAACTGCGCCCAGCGAATCGCCCACATCCTGTCCCAGACGCCCAGCCTGCGACTCCGCATCTTCGTCTCCCGCCTGTACTTCTGCGAC AACGAGGACAGCCTGGAGCGGGAGGGGCTGCGGCACCTGCTGAGGGCAGGGGTGCAGATTACAGTCATGACGTATAAAGATTTTTTCTACTGTTGG CAGACGTTTGTGGCTCGCAGGGAGAGTCGCTTTAAAGCCTGGGACGGTCTTCACCAAAACTCTGTCAGACTGTCCCGCAAACTCAAACGCATCCTCC AGCCCTGTCAGACTGAAGATCTGAGGGACGTCTTCGCTCTGCTGGGTCTCTGA
Bathygadus melanobranchus	ATGATTAGTAAGCTCGACAGTGTGCTTTTGGCCCAGAAAAAATTCATGTACAATTACAAGAACGTGCGCTGGGCAAAAGGCCGCCACGAGACCTAC CTCTGCTTCGTAGTGAGGAGAAGGCTCGGACCAAATTCCCTGTCTTTTGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTAGAGCTGCTGT TTCTGAGCCACCTGGGGGCGCTCTGCCCAGGCCTCTGGGGGTGCGTAGGTGATGACAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTG GTCTCCCTGCGCCAACTGTGCGGCCACACTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGT

	GACCTGGAGGACAGTCCGAATATAGAGGGCTTGAGAGAGCTGAGGAGGGCAGGAGTCCAGGTCATCGTTATGAGCTACAAAGACTACTTCTACTG CTGGCAGACATTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAAAACTAAACCGCAT CCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGAGTTATTGGGCTGTTAAGCTGA
Benthosema glaciale	ATGATTACTAAACTAGACAGTGTGCTTTTGGGTCAGAAGAAGTTCCTCTTCCACTATAAGAACGTGCGCTGGGCGTGGGGTCGAAATGAAACGTAC CTCTGCTTTGTGGTGAAGAGGAGAGTAGGACCAAACTCCCTCTCCTTTGACTTTGGACATCTCCGCAACCGCTCCAGCTGCCACGCGGAGCTGCTGT TCCTTCGCCACCTGGGGGGCGCCCTGTGCCCTGGTCTGTGGGGCTACGGAGGTGACGGGGGAGAGGGGAGGTTCAACTACTCGGTCACCTGGTTCT GCTCGTGGTCTCCGTGCGCCGACTGTTCTCTGAGACTGGCCCAGTTCCTCAGCCGGACCCCCAACCTGCGCCTCCGCATCTTCGTCTCTCGCCTCTAC TTCTGTGACGCGGAGGACAGCCGGGAGAGGGAGGGTCTGAGGACGCTGAAAAGGGCAGGTGTACAGATCACCGTCATGAACTACAAAGACTACTA CTATTGTTGGCAGACCTTTGTGGCTCACAGACAGAGCAGCTTCAAGGCCTGGGCTGATCTGCACCAGAACTCTGTCCGTCTGGCCAGGAAACTCCAC CGCATCCTCCAGCCTTGTGAGACAGAGGATTTTAGAGACGCATTCAAGCTTCTTGGGTTGTGA
Beryx splendens	ATGATTACAAAACTAGACAGTGTGCTTTTGGCCAAGAAAAAGTTCATCTACCATTACAAGAACATGCGCTGGGCAAAGGGCCGGCATGAGACATAC CTCTGCTTTGTGGTGAAGAGGCGAGTGGGGCCAGACTCCCTGTCCTTCGACTTCGGACACCTCCGCAACCGCGCTGGCTGCCATGTAGAGCTGCTGT TCCTGCGCCACCTGGGAGCCCTGTGCCCTGGACTGTGGGGGCATGGAGGCAGCGGAGAGAGGAAGCTGAGTTACTCCATCACCTGGTTCTGCTCCT GGTCTCCCTGCGCTGACTGCTCCTTCAGACTGGCCCAGTTCCTCAACCGGACGCCCAACCTCCGCCTCAGGATCTTCGTCTCCCGCCTCTACTTCTGC GACCAGGAGGACAGCCGCGAGAGAGACGGCCTGAGGCTGCTGAAAAAGGCCGGCGTGAACATCACTGTCATGAGCTACAAAGACTTCTTCTACTG CTGGCAGACCTTTGTGGCTAACAGAACGAGCAGATTCAAGGCCTGGGATTTGCTGCACCAAAACTCTGTTCGCCTGGCCAGGAAACTCAACCGCAT CCTCCAGCCTTATGAGATAGAAGATTTAAGAGATGCCTTCAGACTTCTTGGTTTTTGA
Boreogadus saida	ATGATTAGGAAGCTAGACAGTGTGCTCTTGGCCCAGAATAAATTCATCTACAATTACAAGAACATGCGATGGGCAAAAGGCCGCAACGAGACCTAT CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCTGATTCCCTCTCTTTCGACTTCGGACACCTACACAATCGCACTGGCTGCCACGCAGAGCTGCTGT TCCTGAGCTACCTGGGGGCGCTGTGCCCGGGCCTCTGGGGCTGCGCAGACGACAGAAACCGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT GGTCGCCCTGTGCCAACTGTGCGACCACGCTGACCCGGTTCCTGAGGCAGACACCAAACCTGCGACTCAGGATCTTCGTGTCTCGCCTCTACTTTTG TGACCTGGAGGGCAGTCCGCATGTAGAGGGCTTGAGGGACCCGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAAGACTACTTCTACT GCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTATGTGCNTCTGTCAAGAAAACTAAACCGCA TCCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACCTGA
Borostomias antarcticus	ATGATCAGTAAACTAGACAGTGTTCTCCTGGCCCAGAAGAAGTTCCTCTTCCACTACAAGAACGTGCGCTGGGCCCGAGGACGACACGAGACGTAC CTGTGCTTCGTGGTGAAGAGGAGGGTGGGACCCGACTCGCTTACCTTCGACTTCGGACACCTGCGCAATCGCACCGGCTGCCACGTTGAGCTGCTGT TCCTGCGCCATCTAGGGGTGCTGTGTCCGGGCCTGTCGGCGTCTGGAGGTGCTGGAGGGGGCAGGGGGCTGAACTACTCCATCACCTGGTTCTGCTC ATGGTCCCCCTGCTTCGACTGCTCGGCCCGGCTGGCCCAGTTCCTGAGACGGACCCCCAACCTCAGGCTCCGCCTCTTCGTCTCCCGCCTCTACTTCT GTGACCCGGAGGACCGCCACGAGAGAGAGGGGCTCCGGGCGCTGAAGAGAGCCGGAGTCCACATCACCGTCATGAGCTATAAAGATTATTTTTACT GCTGGCAGACGTTTGTAGCTCACAGACAGAGGGCCTTCAAAGCCTGGGAAGATCTTCAGCAGAACTCCGTCCGCCTGGCCAGGAAGCTCAACAGCA TCCTGCTGCCCTGTGAGACGGAGGATCTGAGAGACCCGTTCAGGCTGCTTGGACTGTGA
Brosme brosme	ATGATGAGTAAGCTAGACAGTGTGCTCTTAGCCCAGAAGAAATTCATATACAATTACAAGAACCTGCGATGGGCAAAAGGCCGCAACGAGACCTA CCTCTGCTTCGTAGTGAAGAGAAGGCTCGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGTACTGGCTGCCACGTAGAGCTGCTG TTTCTGAGCTACCTGGGGGCGCTGTGCCCAGGCCTCTGGGGGTGCGGTGGCGACAGAAACCAAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCT GGTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCACACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGT GACCTGGAGGGCAGTCCGCATATAGAGGGCTTGAGGGACCTGAGGAGAGCAGGGGTCCAGGTCAAAGTTATGAGCTACAAAGACTACTTCTACTG CTGGCAGACCTTCGTAGCTCACAGACTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCAAGAGCTCTAAACCGCATC CTCCAGCCATGTGAAACAGAAGATTTAAGAGATCCTTTCAGACTTTTTGGACTGTTAACCTGA
Brotula barbata	ATGATTGCGAAACTAGACAGTGTACTTTTACCACGGAAAAAGTTCATCTACCATTTCAAGAACATGCGCTGGGCTAAGGGTCGGCATGAGACGTAC CTGTGCTTTGTGGTGAAGAGGCGAGTAGGGCCGGACTCGCTGTCCTTTGACTTTGGACACCTCCGCAATCGCAATGGCTGCCACGTAGAGCTACTGT TCTTACGCTACCTAGGAGCTTTATGCCCTGGACTGTGGGGCTGTGGGAATTCTGGACAGAGGTTGTGTTACTCCATCACTTTGTTCTGCTCTTGGTCC CCCTGTGCCAACTGTTCCGAGAGACTGGCCAAGTTCCTCGGCCGGACACCCAACCTTCGCCTCAGGATCTTTGTCTCTCGCCTCTACTTCTGCGACAT GGAAGACAGCCGTGAAAGAGAGGGTCTGAGGATGCTGAAAAATGCTGGCGTAAACATCACAGTCATGAGCTACAAAGACTATTTCTATTGCTGGC

	AAACCTTTGTGGCTCGTGGCGCAAGCAACTTCAAAGCCTGGGATGGGCTGCAAGAGAATTCAATTCGCCTTGCCAGGAAACTCACCCACATCCTAC AGCCAGGTGAGACGGAAGATTTAAGGGACGCATTCAAACTTCTGGGTATGTGA
Carapus acus	ATGACTGCCAAGCTAGACAGGGTCCTTTTGCCACGGAAAAAGTTCCTCTTCCATTACAAGAACGTGCGCTGGGCGAAGGGCCGCCACGAGACGTAC CTCTGCTTCGTGGTGAAGAGGCGAGTGGGTCCAGACTCCATGTCCTTTGACTTTGGACACCTCCGCAATCGCAGTGGCTGCCACGTAGAGCTCTTGT TCCTGCGCTACCTGGGAGCTCTGTGTCCTGGACTGTGGGGGTATGAAGGTTCTGGACAGAGGAGACTCAGCTACTCCATCACCTGGTTCTGCTCTTG GTCCCCGTGCGCCAACTGCTCGGAGCGACTCGCCCAGTTCCTCAATCGGACCCCCAACCTCCGCCTCAGGATCTTCGTCTCTCGCCTCTACTTCTGCG ACCTGGAGGACAGCCGTGAGAGGGAGGGCCTGAGGACGCTGGAGAAAGCTGGCGTGCACATCACCATCATGAGCTACAAAGACTATTTCTACTGC TGGCAAACCTTTGTGGCTTGTGGAACTTCAAAATTCAAAGCCTGGGATGAGCTCCACCAAAACACCACTCGTCTCAAGAGAAAACTGAATCGGATC CTCCAGCCATGTGAGACAGAAGATTTAAGGGACGCATTCAAACTTCTAGGGTTGCTGTGA
Chaenocephalus aceratus	ATGATCACAAAGCTTGACAGCATGCTTTTGCCTCGAAAAAAGTTCATCTACCATTACAAGAACATGCGCTGGGCAAGGGGCCGGTGTGAGACATAC CTCTGCTTTGTAGTGAAGAGGCGGGTGGGACCAGACTCCTTAACCTTTGACTTCGGACACCTTCGCAATCGCAATGGCTGCCATGTAGAGATGCTGT TCCTGCGCTACCTGGACGCCCTGTGCCCTGGTCTGTTGGGATGTGAAGGTACTGGAGAGAAGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG GTCCCCCTGTGCAAACTGCTCCATCAGGCTGTCCCAGTTCCTCAGCCAGACGCCCAATCTTCGCCTCAGGATCTTCGTCTCTCGTCTTTACTTCTGTG ACATGGAGAATAGCCCTGCAAGAGACGGCCTAATAATGCTGAAAAAAGCTGGCGTGCAGACTTCAGTCATGAGTTACAAAGACTTTTTCTATTGCT GGCATAACTTTGTGGATTGTAAACAGAGTAAATTCAAGCCATGGGAAGATCTGCACCAAAACTCTGTTCGCCTTGCCAGAAAACTCAAACGCATCC TTCAGCTGTGTGAAACTGAAGATTTGAGAGATGCCTTCAAGCTTCTTGGACTGTAA
Chatrabus melanurus	ATGATTACAAAACTAGACAGTGTGCTTTTGCCACGGAAGAAGTTCATCTACCATTACAAGAACATGCGCTGGGCAAAGGGCCGGCACGAGACATAC CTCTGCTTTGTGGTGAAGAGACGAATGGGGCCAGACTCCCTGTCCTTTGATTTCGGACACCTCCGCAATCGCAACGGCTGCCATGTAGAGCTGCTGT TCCTGCGTTACCTGGGAGCCTTGTGCCCTGGTCTGTGGGGGTATGGAATTGCTGGAGAGAGGAAGCTTAGTTACTCCGTCACCTGGTTCTGCTCCTG GTCCCCCTGTGTCAACTGCTCCCTCAGACTGACACAGTTCCTCATGCAGACGCCTAATCTTCGCCTCAGGATCTTCGTCTCTCGCCTTTACTTCTGTG ATATGGAAGACAGCCGTGAGAGAGAAGGTCTGAGGATGCTGAAAAAAGCCGGCGTGCACATCACAGTGATGAGTTACAAAGACTTCTTCTACTGCT GGCAGACCTTTGTGGCTTGTAAAGAGAGCAAATTCAAGGCATGGGAGGCGCTGCACCAAAACTCTGTTCGTCTGGCTAGAAAGCTCAACCGCATCC TCCAGCCCTGTGAGACAGAAGACTTCAGAGATGCCTTCAAGCTTCTTGGACTGTGA
Chromis chromis	ATGATCACAAAACTCGACAGTGTGCTTTTGCCCCAGAAGAAGTTCATCTACCATTATAAGAACATGCGCTGGGCGAGAGGCCGCTGTGAGACGTAC CTCTGCTTCGTGATTAAGAAAAGAGCCGGTCCAGATTCTATATCCTTCGACTTCGGACATCTACGGAACCGCAACGGCTGCCATGTAGAGCTGCTGT TCCTGCGCTACCTGGGCGCCTTGTGTCCTGGTCTCTGGGGTTATGGACAGAACCGGATCAGCTACTCCATCACCTGGTTCTGCTCCTGGTCTCCCTGC GCTAACTGCTCCCTCAGACTGGCCCAGTTCCTGAACCAGACGCCCAACCTTCGTCTCCGGATCTTCGTCTCTCGGCTCTACTTCTGCGACATGGAGGA CAGCCGGGAGAGGGAAGGTCTGAGGATCCTGAAGAAGGCCGGCGTTAACATCACCGTCATGAGCTACAAAGACTACTTCTACTGCTGGCAGACCTT CGTGGCTCGGAGGCTGAGTAAGTTCAAACCGTGGGACGGGCTGCAACAGAACTACGTCCGTCTGTCCAGAAAACTGAACCGCATCCTGCAGCCCTG TGAGACTGAAGACTTTCGAGACGCCTTCAGGCTCCTTGGACTCTGA
Cyttopsis roseus	ATGATTACTAAACTAGACAGTGTGCTTCTGGCTCGGAAGACATTCATTTACCACTATAAGAACATGCGCTGGGCAAAAGGCCGGCATGAGACATAC CTCTGCTTCGTCGTCAAGAGAAGAGTTGGACCCGATTCCTTGTCCTTTGACTTTGGACACCTTCGCAATCGGACTGGCTGCCATGTAGAGCTCCTGTT TCTACGTCACCTGGGGGCCCTGTGCCCTGGACTGTGGGGACAAGGAGGCGCTGATGAAAGAAGGCTCAGTTACTCGGTCACCTGGTTCTGCTCCTG GTCTCCCTGCGCCAACTGCTCCCTCAGACTGGTCCAATTCCTCGGGCAGACGCCCAACCTCCGTCTCAGGATCTTCGTCTCCCGTCTCTACTACTGTG ACCTTGAAGACAGCCGCGAGAGAGAGGGCTTAAGAACCCTGAAAAGAGCCGGAGTCCAAATCACAGTCATGAGCTACAAAGACTATTTCTATTGCT GGCAGACGTTCGTGGCTCGCCGACAGACCCGCTTCAAGGCGTGGGATGAGCTGCACCAAAACTCAGTTCGTCTGGCCAGGAAACTAAACCGCATCC TCCAGCCTTGTGAAACGGAAGATTTAAGAGATGCTTTCAAACTTCTCGGGTTCTTGTAA
Danio rerio	GenBank: BC162573.1
Gadiculus argenteus	ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATATACAATTACAATAACATGCGATGGGCAAAAGGCCGCAACGAGACCTAC CTCTGCTTCGTCGTGAAGAGAAGGCTTGGACCTGACTCCCTCTCCTTCGACTTCGGACACCTACGCAATCGCACCGGCTGCCACGCAGAGGTGCTGT TCCTGAGCTACCTCGGGGCACTGTGTCCGGGCCTCTGGGGCTGCGCAGGCGACAGAAGCCTAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCTG GTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTCGCTCGCCTCTACTTCTGTG ACCTGGAGGGCAGTCCGCATGTGGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTTATGAGCTACAAAGACTACTTCTACTGCT

	GGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTGCGTCTGTCAAGGAAACTAAACCGCATCC TCCAGCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACCTGA
Gadus morhua	NCBI Reference Sequence: XM_030370988.1
Gasterosteus aculeatus	ATGATTGCAAAGCTTGACAGTGTGCTTCTGCCCCGAAAAAAGTTCATCTACCACTACACGAACATGCGCTGGGCGAGGGGCCGACACGAGACTTAC CTCTGCTTTGTTGTGAAAAGGCGAGTGGGGCCGGATTCCTTGTCCTTCGACTTTGGACACCTGCGCAATCGCAGTGGCTGCCATGTCGAGTTGTTGTT CCTGCGCCACCTCGGAGCCTTGTGCCCTGGTTTCTTGGGTTGTGGAGACACCGGAGGGAGGAGGCTGAGTTACTCCATCACCTGGTTCTGCTCGTGG TCTCCCTGCGTAAACTGCTCCATCAGTCTGTCCCAGTTCCTCAGCCGGACGCCCAACCTCCGCCTCAGGATCTTCGTCTCTCGCCTTTACTTTTGTGAC ATGGAGAACAGTCGGGAAAGAGACGGCCTGAGAATGCTGAAAAAAGCTGGCGTGCAGGTCACAGTCATGAGTTACAAAGATTTCTTCTATTGCTGG CAGACTTTTGTAGATCGCAAACAAAGCCAGTTCAAGGCCTGGAAAGAGCTTCACCAAAACTCTGTTCGCCTTTCCAGAAAGCTCAAGCGCATCCTCC AGCCTTGTGAAACAGAAGATTTAAGGGATGCCTTCAAGCTGCTTGGACTGTGA
Guentherus altivela	ATGATTACTAAACTAGACAGCATACTTATGGCCCAGAAGAAGTTCATCTTCCACTATAAGAACATGCGATGGGCCAAGGGTCGAAATGAGACACAC CTCTGCTTTGTGGTGAAGAGAAGGCTGGGACCAAACTCCCTGTCCTTTGACTTTGGACACCTGCGTAATCGCACTGGCTGCCATGTAGAGCTACTCT TCTTGCGCCACCTGGGATTCCTGTGCCCTGGCTTGTGGGGGTACGGAGAGCCAGGTGAAGGGAGGCTGAATTACTCTGTCACCTGGTTCTGCTCCTG GTCCCCCTGTGCAGATTGTTCCTTCACGCTGACCCACTTCCTCAGAGAGACTCCCAACCTCCGTCTTAGAATCTTTGTGTCTCGCCTCTACTTCTGTGA CGAGGAGGACAGCAGTGCAAGGGAAGGCCTGCGAATGTTGAAGAAAGCCGGTGTGAACATCACTGTCATGAGCTACAAAGACTACTTCTATTGCT GGAAGACCTTTGTGGCTCACAGACAAAGGAACTTCAAGGCCTGGGATGGGCTAGACCAGAACTCTGTTCACCTAGCCAGGAAACTCAGCCACATCC TCCAGCCCTGGGAAACAGCAGATTTAAGAGATGCCTTTAAACTTCTTGGACTGTGA
Helostoma temminckii	ATGATTACAAAGCTAGACAGTGTGCTTTTGCCCCGAAAAAAGTTCATCTACCATTACAAAAATGTGCGCTGGGCAAGGGGTCGGCATGAGACATAC CTCTGTTTTGTAGTGAAGAGGCGGGTGGGCCCAGACTCCTTGACCTTTGACTTTGGGCATCTCCGCAATCGCAATGGTTGCCATGTAGAGATGCTGT TCTTGCGATATCTGGGAGCTTTGTGCCCTGGACTTTGGGGGTGTGGAGGTACTGGAGAGAGAAGGCTCAGTTACTCTATTACCTGGTTCTGCTCCTG GTCTCCTTGTTCTAACTGCTCCCTTAGACTGGCCCAGTTCCTCAGTCAGACCCCAAACCTCCGCCTCAGGATCTTTGTGTCTCGCCTATACTTCTGTGA CATGGAGGACAGTCGCGAGAGGGAGGGTCTCAGGATCCTGAAAAACGCTGGAGTGCAGATCACAGTCATGAGTTACAAAGACTTCTTCTACTGCTG GCAGACATTTGTGGCACGTAAGCAGAGCAACTTCAAAGCATGGGAGGAGCTGCACCAAAACTCTGTTCGCCTTACCAGAAAACTCCATCGCATCCT TCAGCCTTGTGAAACAGAAGATTTAAGAGATGCTTTCAAGCTCCTTGGACTGTGA
Holocentrus rufus	ATGATTACAAAACTAGACAGTGTGCTTTTGGCCAAGAAAAAGTTCATCTACCATTATAAGAACTTGCGCTGGGCAAAAGGCCGGCATGAGACATAC CTCTGCTTTGTCGTGAAGAGGCGGGCGGGGCCGGACTCCATCGCCTTCGACTTTGGACACCTCCGCAACCGTGCTGGCTGCCATGTAGAGCTGCTAT TCCTTCGCTACCTGGGAGCCTTGTGCCCTGGACTGTGGGGCTACGGAGGAACTGGTGAGAGGAAGATGAGCTACTCCATCACATGGTTCTGCTCCTG GTCTCCTTGTGCCAACTGCTCCTACAGACTCGCCCAGTTCCTCAACCGGACGCCCAACCTCCGCCTCAGGCTCTTCGTCGCTCGCCTCTATTTCTGTG ACATCGAGGACAGCCGTGAGAGAGAGGGCCTGAGAATGCTGAAGAATGCCGGTGTGCACATCACTGTCATGAGCTACAAAGACTACTTCTACTGCT GGCAGACATTTGTGGCTCGTAAAACGAGCAACTTCAAGGCCTGGGATGGGCTGCACCAAAACTATGTTCGTCTGGCCAGGAAACTCAACCGCATCC TCCAGCCTTGTGAGACAGAAGATTTAAGAGATGCATTCAGGCTTCTTGGCTTGTGA
Laemonema laureysi	ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTTCAATTACAAGAACATGCGCTGGGCAAGAGGCCGCAACGAGACCTAC CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCCAATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACAGGCTGCCATGTAGAGCTGCTGT TTTTGAGCTATCTGGGGGGCACTGTGCCCAGGCCTGTGGGGGTGCAGAGGCGACGAAAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCT GGTCTCCATGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTG TGACCTGGAGGACAGTCCCCATATAGAGGGCTTGAGGGACCTGAGGAGAGCAGGGGTGCGGGTCACCGTTATGAGCTACAAAGACTACTTCTACTG CTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAAAACTAAACCGCAT CCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGGCTGTTAACCTGA
Lampris guttauts	ATGATCAGCAAACTAGACAGTGTGCTTCTGACCCAGAAGAAGTTCCTCTACCATTATAAGAACGTGCGTTGGGCAAAAGGTCGGCATGAGACATAT CTCTGCTTTGTGGTGAAGAGGAGGGTGGGACCGGACTCCATGTCCTTCGATTTTGGACACCTCCGCAATCGAGCTGGCTGCCATGTAGAGCTGCTGT TCCTGCGCTACCTGGGGGCCCTGTGTCCTGGACTGTGGGGCTACGGGGACACCGGAGACAGGAGGCTCAGTTACTCGGTCACCTGGTTCTGCTCCTG GTCTCCCTGCGCCAACTGCTCCTTCAGACTGGCCCAGTTCCTCCAAAGGACGCCCAACTTCCGCCTCAGGCTCTTTGTCTCCCGTCTGTACTTCTGTG ACATGGAGGACAGCAGTGAGAGGGACGGCCTGAGGTTGCTGAAAAACGCAGGGGTGCAGATCACCGTCATGAGCTACAAAGACTACTTCTATTGC

	TGGCAGACTTTTGTGGCTCACAGAAAGAGCAGTTTCAAGGCCTGGGATGGGCTGCACCAAAACACTGTTCGCTTGGCCCGGTTACTCAACCGCATCC TCCAGCCTTGTGAGGCAGAGGATTTGCGGGATGCGTTCAAACTTCTCGGGTTTTGA
Lamprogrammus exutus	ATGATTGCAAAACTAGACAGTGTGCTTTTGCCCCGCAAAAAGTTCATCTTCCATTACAAGAACATGCGCTGGGCTAAGGGTCGGCACGAGACATAC CTCTGCTTTGTAGTGAAGAGACGAGTGGGTCCAGACTCCCTGTCCTTTGACTTTGGACACCTCCGCAATCGCAATGGCTGCCATGTAGAGCTACTGT TCCTGCGCTACCTGGGAGCTCTATGCCCTGGACTGTGGGGGTGTGGAGGTTCTGGTGAGAGGAGACTCAGTTACTCCATCACCTGGTTCTGCTCTTG GTCCCCCTGTGCCAACTGCTCCCAGAGACTATCCCAATTCCTCAGCCAGACACCCAACCTTCGCCTCAGGATCTTTGTCTCTCGCCTCTACTTCTGTG ACATGGAGAACAGCCGTGAGAGAGAGGGCCTGAGGATGCTGAAAAATGCTGGTGTGCAAATCACAGTCATGAGCTACAAAGACTTTTTCTATTGCT GGCAAACCTTTGTGGCTTGTGGGAAAAGCAAATTCAAGGCCTGGGATGAGCTGCACCGAAACTCTGTTCGCCTCACCAGGAAACTGAACCGCATCC TCCAGCCATGGGAGACAGAAGATTTAAGAGATGCATTCAGACTTCTTGGATTTTGA
Lesueurigobius cf sanzoi	ATGATTACCAAGCTAGACAGTGTACTTTTACCAAAGAAGAAGTTTATCTTCCATTACAAGAACGTGCGCTGGGCGAAGGGTCGGCATGAGACGTAC CTCTGCTTTGTGGTCAAGAGGCGCGTGGGGCCAAATTCTATGTCCTTTGACTTTGGACATCTTCGCAATCGCAGCGGCTGCCATGTGGAGATTCTGTT CCTGCGTTACCTTGGTGCTCTGTGCCCTGGACTCTGGGGGGCTGGAGGCTCGGAGGAGAGGCGACTGAGTTACTCCATCACTTGGTTCTGCTCCTGG TCTCCATGCGCCAACTGCTCCACGAAACTGTCGCAGTTCCTCGCCAAAACCCCAAACTTGCGTCTGCGGATATTTGTCTCACGCCTTTACTTCTGCGA CCTGGAGGACAGCATAGAACGAGAGGGTCTGAGGATGCTAAAGAGAGCAGGCGTGCAGTTAACGGTCATGAAATACAAAGACTACTTTTACTGCT GGCACACGTTTGTGGCTCGAAACCAAAGCAACTTCAAGGCCTGGGAAGAGCTTCACCAAAACTCAGTGCGACTGACCAGGAAACTCAGTCGCATCC TTCAGCCATGTGAGACAGAGGATTTAAGAGATGCCTTCAGACTTCTTGGTTTGTGA
Lota lota	ATGATAAGTAAGCTAGACAGTGTGCTCTTAGCCCAGAAGAAATTCATATACAATTACAAGAACATAAGATGGGCAAAAGGCCGCAACGAGACCTA CCTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTAGAGCTGCTG TTTCTGAGCTACCTGGGGGCGCTGTGCCCGGGCCTCTGGGGGTGCGGAGGCGACAGAAACCGAAGACTCAGCTACTCGGTCACCTGGTTTTGCTCCT GGTCTCCCTGTGCCAACTGTGCGGCTACACTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGT GACCTGGAGGGCAGTCCGCATATAGAGGGCTTGAGGGACCTGAGGAGAGCCGGGGTCCAGGTCAAAGTTATGAGCTACAAAGACTACTTCTACTG CTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCATGGGAAGGGCTGCATACCAATTCGGTCCGTCTGTCAAGAAAACTAAACCGCAT CCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTTTTGGACTGCTAACCTGA
Macrourus berglax	ATGATTAGTAAGCTTGACAGCATACTCTTGGCCCAGAAGAAATTCAAGTACAATTACAATAACATGCGATGGGCAAAGGGCCGCAACGAGACCTAC CTCTGCTTCGTAGTGAAGAGAAGGCTCGGACCCAATTCACTGTCCTTTGACTTCGGACACCTACGCAATCGTGCTGGCTGCCACGTAGAGCTGCTGT TTCTGAGCCACCTGGGGGCGCTGTGCCCGGGCCTGTGGGGCTTTGGAGGGGCAGAAAACATAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTG GTCTCCCTGCGCCAACTGTGCGGCCACACTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGT GAACTGGCGGACAGTCCGCACTCAGAGGGCTTGAGGGAGCTGAGGAGAGCAGGGGTCCAGGTCAACGTTATGACCTACAAAGACTACTTCTACTG CTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAAAACTAAACCTCATC CTCCAGCCATGTGAAACAGAAGATTTAAGAGACGCTTTCAGACTTATTGGCCTGTTAACCTGA
Malacocephalus occidentalis	ATGATTAGTAAGCTCGACAGCGTGCTCTTGGCCCAGAAGAAATTCATATACAATTACAAGAACATACGCTGGGCAAAGGGCCGCAACGAGACCTAC CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCCAATTCACTGTCCTTCGACTTCGGACACCTACGCAACCGCACTGGCTGCCATGTAGAGCTGCTGT TTCTGAGCTACTTGGGGGGCGCTGTGCCCGGGCCTGTGGGGCTGTGGAGGTGCAGATAACAGAAGACTCAACTACTCGGTCACCTGGTTCTGCTCCTG GTCTCCCTGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTTGTGGCTCGCCTCTACTTCTGC GACCTGGACGACAGTCCACACACAGAGGGCTTAAGGGAGCTGAGGAGAGCAGGGGTCCAGTTCACCGTAATGAGCTACAAAGACTACTTCTACTG CTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAAAACTAAACCGCAT CCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGGCTGTTATCCTGA
Melanogrammus aeglefinus	ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATCTACAATTACAAGAACATGCGATGGGCAAAGGGCCGCAACGAGACCTAT CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCTGATTCCCTCTCTTTCGACTTTGGACACCTACGCAATCGCACTGGCTGCCACGCAGAGCTGCTGT TCCTGAGCTACCTGGGGGCACTGTGCCCAGGCCTCTGGGGCTGTGCAGGCGACAGAAACCGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCTG GTCGCCCTGTGCCAACTGTGCGACCACGCTGACCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGTCTCGCCTCTACTTCTGT GACCTGGAGGGCAGTCCGCATGTAGAGGGCTTGAGGGATCTAAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAAGACTACTTCTACTG

	CTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTATGTGCGTCTGTCAAGAAAACTAAACCGCAT CCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACCTGA
Melanonus zugmayeri	ATGATTAGTAACCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTACAAGAACATGCATTGGGCAAAAGGCCGCAACGCGACCTAC CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACACAATCGCACTGGCTGCCACGCAGAGCTGCTGT TTCTCAGCCACCTGGGGGCACTGTGCCCAGGCCTGTGGGGNTGCGGAGGCGACAAAAACAGAAGACTCAGCTATTCGGTTACCTGGTTCTGCTCCT GGTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGCTTTCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCGTCTACTTCTGT GAACAGGAGGACAGTCCGCATATAGAGGGCTTGAGGGATCTGAGGAGAGCAGGGGTCCAGGTCACCGTTATGAGCTACAAAGACTACTTCTACTG CTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGACCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAAAACTAAACCGCAT CCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGACTGTTAACCTGA
Merlangius merlangus	ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATCTACAATTACAAGAACATGCGGTGGGCAAAAGGCCGCAACGAGACCTAT CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCTGATTCCCTTTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGCAGAGCTGCTGT TCCTGAGCTACCTGGGGGCACTGTGCCCAGGCCTCTGGGGCTGCGCAGGCGACAGAAACCGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT GGTCGCCCTGTGCCAACTGTGCGACCACGCTGAGCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGTCTCGCCTCTACTTCTG TGACCTGGAGGGCAGTCCGCATGTAGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAAGACTACTTCTACT GCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTATGTGCGTCTGTCAAGAAAACTAAACCGCA TCCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACCTGA
Merluccius merluccius	ATGATTAGTAAGCTCGACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTACAAGAACATGCGCTGGGCAAAAGGCCGCAACCAGACCTAC CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACACAATCGCACTGGCTGCCACGCAGAGCTGCTGT TCCTGAGCCACCTAGGGGCGCTGTGCCCGGGTCTGTGGGGGTGCGGAGGTGACGAAAACCGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCT GGTCTCCCTGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGACTCACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGT GACGTGGAGGACAGTCCGCACAGGGAGGGCTTGAGGAACCTGAGGAGAGCAGGGGTCCTGGTCAACGTTATGAGCTACAAAGACTACTTCTACTG CTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAACACTAAACCGCATC CTCCAGCCATGTGAAACAGAAGATTTAAGAGACGCTTTCAGACTTATTGGGCTGTTAACCTGA
Merluccius polli	ATGATTAGTAAGCTCGACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTACAAGAACATGCGCTGGGCAAAAGGCCGCAACCAGACCTAC CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACACAATCGCACAGGCTGCCACGCAGAGCTGCTGT TCCTGAGCCACCTAGGGGCGCTGTGCCCGGGTCTGTGGGGGTGCGGAGGTGACGAAAACCGAAGACTCAGCTACTCTGTCACCTGGTTCTGCTCCT GGTCTCCCTGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGACTCACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGT GACGTGGAGGACAGTCCGCACAGGGAGGGCTTGAGGAACCTGAGGAGAGCAGGGGTCCTGGTCAACGTTATGAGCTACAAAGACTACTTCTACTG CTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAACACTAAACCGCATC CTCCAGCCATGTGAAACAGAAGATTTAAGAGACGCTTTCAGACTTATTGGGCTGTTAACCTGA
Molva molva	ATGATTAGTAAGCTAGACAGTGTGCTCTTAGCCCAGAAGAAATTCATATACAACTACAAGAACATGCGATGGGCAAAAGGCCGCAATGAGACCTAC CTCTGCTTCGTAGTGAAGAGAAGGCTCGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTAGAGCTGCTGT TTCTGAGCTACCTGGGGGGCGCTGTGCCCGGGCCTCTGGGGGGTGCGGAGGCGACACTAACCGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTG GTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCACACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTG ACCTGGAGGGCAGTCCGCATATAGAGGGCTTAAGGGACCTGAGGAGAGCAGGGGTCCAGGTCAAAGTTATGAGCTACAAAGACTACTTCTACTGCT GGCAGACCTTCGTAGCTCACAAGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTATGTCCGTCTGTCAAGAAAACTAAACCGCATCC TCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTTTTGGACTGTTAACCTGA
Monocentris japonica	ATGATTACAAAACTAGACAGTGTGCTTTTGGCGCAGAAAAAGTTCATCTACCATTATAAGAACATGCGCTGGGCAAGGGGTCGGCATGAGACATAC CTCTGCTTTGTAGTGAAGAGGAGAGTGGGACCAGACTCCCTGTCCTTTGACTTTGGACACCTCCGCAATCGCTCTGGCTGCCATGTAGAGCTGCTGT TCCTGCGCCACCTGGGAGCCTTGTGCCCTGGACTGTGGGGGTATGGAGGCACTGGTGAGAGGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG GTCTCCCTGCGCTGACTGCTCCTTTAGATTGGTCCAGTTCCTCGGCCGGACGCCCAACCTCCGCCTCAGGATCTTCGTCTCTCGCCTCTACTTCTGTG ACGTGGAGGACAGCCGCGAGAGACAGGGCCTGAGAATGCTGAAAAAAGCCGGCGTGCAAATCACTGTCATGAGCTACAAAGACTACTTCTATTGC

	TGGCAGACCTTCGTGGCTCACAGACAGAGCAGTTTCAAGGCCTGGGATGAGCTGCACCAAAACTCTGTTCGCCTGGCCAGGAAACTCAACCGCATC CTCCAGCCTTGTGAGACAGAAGATTTAAGAGATGCGTTCAAGCTTCTTGGGTTGTGA
Mora moro	ATGATTAGTACACTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTACAAGAACATGCGTTGGGCAAAAGGCCGCAACGAGACCTAC CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTAGAGCTGCTGT TTCTGAGCCACCTGGGGGCACTGTGCCCAGGCCTGTGGGGGTGCGGAGGCGATGAAAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCT GGTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTATTTCTGT GACCTGGAGGACAGTCCGCATATAGAGGGCTTGAGGGACCTGAGGAGAGCAGGGGTGCAGGTCACTGTTATGAGCTACAAAGACTACTTCTACTGC TGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAAAACTAAACCGCATC CTCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGGCTGTTAACCTGA
Muraenolepis marmoratus	ATGATTAGCAAACTAGACAGTGTGCTCTTGGGCCAGAAGAAATTCATATACAATTACAAGAACATGCGTTGGGCAAAAGGCCGCAACGAGACCTAC CTCTGCTTCGTGGTGAAGAGAAGGCTCGGACCCGATTCCATGTCTTTCGACTTCGGGCACCTACGCAATCGCGCAGGCTGCCACGTGGAGCTGCTGT TTCTCAGCCACCTGGGGGGCGCTGTGCCCGGGTCTGTGGGGTTGCGGAGGCGACGAGAACAGACGGCTCAGCTACTCGGTCACCTGGTTCTGCTCCTG GTCCCCCTGTGCCAACTGTGCCGCCACGCTGGCCCGGCTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGGCCCGCCTGTACTTCTGT GACCTGGAGGGCAGTCCGCACTCAGAGGGCCTGAGGGACCTGAGGAGGGCCGGGGTCCAGGTCAACGTTATGAGCTACAAAGACTACTTCTACTG CTGGCAGACCTTTGTAGCGCACAGGGTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAAAACTAAACCGCAT CCTCCAGCCACGCGAAACAGACGATTTAAGAGATGCCTTCAGACTTATTGGTCTGTTAACCTAA
Myoxocephalus scorpius	ATGATTACAAAGCTAGACAGTGTGCTATTGCAGCAAAAAAAGTTCATCTACCATTACAAGAACATGCGCTGGGCAAGGGGCCGACATGAGACTTAC CTCTGCTTTGTAGTGAAGAGGCGAGTGGGGCCAGACTCCTTATCCTTTGACTTTGGACACCTCCGCAATCGCACTGGCTGCCATGTAGAGCTGTTGT TCCTACGCTACCTGGGAGCCTTGTGCCCTGGTTTGTGGGGTTACGGAGGCACTGGAGAGAAGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG GTCTCCCTGCATAAACTGCTCCATCAGTTTGTCCCAGTTCCTCAACCGGACGCCCAACCTTCGCCTCAGGATCTTTGTCTCTCGTCTTTACTTCTGTGA CAAGGAGAACAGCCGGGAAAGAGATGGCCTGAGAATGCTGAAAAATGCTGGCGTGCAGATCACAGTCATGAGTTACAAAGACTTCTTCTATTGCTG GCAGACATTTGTGGATCGCAAGAAAAGCAACTTCAAGGCCTGGGAAGAGCTGCACCAGAACTCTGTTCGCCTTGCCAGAAAACTCAACCGCATCCT CCAGCCTTGTGAAGCAGAAGATTTAAGGGATGCCTTCAAGCTTCTTGGACTGTGA
Myripristis jacobus	ATGATTACAAAGCTAGACAGCATGCTTTTGGCCAAGAAAAAGTTCATTTACCATTATAAGAACATGCGCTGGGCTAAAGGTCGGCATGAGACATAC CTGTGCTTTGTAGTGAAGAGACGAGTGGGGCCAGACTCCATGTCCTTTGACTTTGGACATCTCCGCAATCGTGCTGGCTGCCATGTAGAGCTGCTGT TCCTGCGCTACCTGGGAGCGCTTTGCCCTGGACTGTGGGGGTGTGGAGGCAACACTGAGAAGAAGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG GTCTCCCTGCGCCGACTGCTCTTTCAGACTGGCCCAGTTCCTCAACCGGACGCCCAACCTCCGCCTCAGGATCTTTGTCTCTCGCCTCTATTTCTGCG ACCTGGAGGACAGCCGTGAGAGAGAGGGCCTGAGGATGCTGAAAAAAGCCGGCGTGCAAATCACTGTTATGAGTTACAAAGATTACTTCTATTGCT GGCAGACATTTGTGGCACATAGAATGAGCAGCTTCAAGGCTTGGGATGGGCTGCACCAAAACTATGTTCGCCTGGCCAGGAAACTCAACCGCATCC TCCAGGCTAGTGAGACAGAAGATTTAAGAGATGCATTCAAGCTTCTTGGATTGTGA
Neoniphon sammara	ATGATTACAAAGCTAGACAGTGTGCTTTTGGCCAAGAAAAAGTTCATCTACCATTATAAGAACTTGCGCTGGGCAAAAGGCCGGCATGAGACATAC CTCTGCTTTGTCGTGAAGAGGCGGGTGGGGCCAGACTCCATTGCCTTCGACTTTGGACACCTCCGCAATCGTGCTGGCTGCCATGTAGAGCTGCTAT TCCTTCGCTACCTGGGAGCCTTGTGCCCTGGACTGTGGGGGTATGGAGGAACTGGGGAGAGGAAGCTGAGTTACTCCATCACGTGGTTCTGCTCCTG GTCTCCCTGTGCCAACTGCTCCTTCAGACTCGCCCAGTTCCTCAACCGGACGCCCAACCTCCGCCTCAGGATCTTTGTCTCTCGCCTCTATTTCTGTG ACGTGGAGGACAGCCGTGAGAGAGAGGGCCTGAGAATGCTGAAAAATGCCGGCGTGCACATCACTGTTATGAGCTACAAAGACTACTTCTACTGCT GGCAGACATTTGTGGCTCGTAAAACGAGCAGCTTCAAGGCTTGGGATGGGCTGCACCAAAACTATGTTCGCCTGGCCAGGAAACTCAACCGCATCC TCCAGCCTTGTGACACAGAAGATTTAAGAGATGCATTCAGGCTTCTTGGATTGTGA
Oreochromis niloticus	ATGATTGCAAAGCTAGACAGTATGCTTTTGCCCAGAAAAAAGTTCCTCTATCATTACAAGAATGTGCGCTGGGCGAGGGGCCGGAATGAAACATAC CTCTGTTTTGTAGTAAAAAGACGAGTAGGGCCTGACTCCTTGTCCTTTGACTTTGGACACCTCCGCAATCGCAATGGTTGCCACGTTGAGCTGCTGTT CCTGCGCCAACTTGGTACATTATGCCCTGGCCTGTCTGGGTATGGATTTCATGGGGAGAGGAGGGTCAGCTACTCCATCACCTGGTTCTGCTCCTGG TCTCCCTGTGCAAACTGCTCTTCCAGACTGGCCCAGTTCCTCAAACAGACACCCAACCTTCGCCTCAGGATCTTTGTCTCTCGTCTTTACTTCTGTGA CATGGAGGACAGTCGTGAAAGAGAGGGTCTCAGGCTGCTTAAAAAAGTCGGCGTGCACATCACAGTCATGAGTTACAAAGACTTCTTCTACTGCTG

	GGAGAATTTTGTGGCCCAGCAAAGCAAATTTAAGGCCTGGGAAGGTCTGCATCAAAACACAGTACGCCTGGCCAGAAAACTCAACCGCATCCTCCA GCCCTGTGACACAGAAGATTTAAGAGATGCCTTCAAGCTTCTTGGACTGTAA
Oryzias latipes	NCBI Reference Sequence: XM_020710629.2
Osmerus eperlanus	ATGATCAGTACGCTAGACGGCGTGCTTCTGGCCCAGAAGAAGTTCATCTACCACTACAAGAACATGCGCTGGGCCAGAGGTCGACACGAGACCTAC CTGTGCTTTGTGATCAAGAGGAGGGTTGGGCCCGACTCGCTCTCCTTTGACTTCGGACACCTGCGCAATCGCACCGGCTGCCATGTAGAGCTGCTGT TTCTACGCCACCTGGGGGCTCTGTGTCCCGGCCTGTGGGGTACGGGTGGTGCCGGTGGTGGGGTGAGGTTGAGCTACTCCATCACCTGGTTCTGCTC CTGGTCTCCCTGCTCCGCCTGCTCCCACAGGCTGTCTGACTTCCTCAGCCGGACCCCCAACCTCCGCCTCCGGATCTTCGTGTCTCGTCTCTACTTCTG CGACCCGGAAGACAGCCTGGAGAGGGAGGGGCTCCGTATGCTGAAGAGAGCCGGAGTAAACATCACTGTCATGAGTTATAAAGACTATTTCTACTG CTGGGAGACTTTTGTAGCTCGCAGAAAGACAGGCTTCAAGGCCTGGGACGGGCTTCACCACAACTCGGTTCGCCTGGCCAGGAAGCTCTACCGTAT CCTACAGCCTTGTGAGACAGAAGATCTGAGAGATGCTTTCACGCTGCTGGGACTGTGA
Parablennius parvicornis	ATGATTGCCAAGCTCGACAGTATGCTCCTGCCCAGAAAAAAGTTCATCTATCATTACAAGAACATGCGCTGGGCGAAGGGTCGGCATGAGACTTAC CTCTGCTTCGTGGTGAAGCGGCGACTGGGCCCAGACTCTTTGTCCTTTGACTTCGGGCATCTCCGAAATCGCAATGGTTGCCATGTAGAGTTGCTGTT CCTGCGCCACCTGGGGACTTTGTGCCCTGGTCTGTCGGGGTACGGAGTACATGGAGAAAAAAGGCTTAGCTACTCCATCACCTGGTTCTGCTCCTGG TCTCCCTGTTCCAACTGTTCTCACCGACTAGCCCAGTTCCTGAGCCGAACGCCCAACATTCGACTCAGAATCTTTGTCTCCCGCCTGTACTTCTGCGA CTTGGAGGACAGCCGCGAGAGAGAGGGTCTCCGGCTGCTGAAAAAAACTGGCGTGCATATCACGGTCATGAGCTACAAAGATTATTTCTATTGCTG GCAAACTTTTGTGGCAAGTAATCAGAGCAGGTTTAAGCCTTGGGATGAGCTGCAGCGAAACTCCATCCGCCTCACCAGAAAACTCAACCGCATCCT CCAGCCCTGCGAAACAGAAGATTTAAGAGATGCCTTCAAGCTTCTTGGACTCTGA
Parasudis fraserbrunneri	ATGATTACTAATCTAGACAGTGTGCTTCTGGCCCAGAAGAAGTTCATCTACCATTACAAGAACATGCGGTGGGCAAGGGGCCGGCATGAGACTTAT CTCTGCTTTGTAGTGAAGAGAAGGTTGGGTCCAGACTCCTTGTCCTTCGACTTTGGACACCTTCGCAATCGCTCTGGCTGCCATGTAGAGCTGCTTTT CCTGCGTCACCTGGGCGCCCTTTGCCCTGGCCTGTGGGGATATGGAGGAGAGAAGAGGCTGAGCTACTCTGTCACCTGGTTCTGCTCCTGGTCGCCC TGCGCCGACTGCTCCACCAGACTGTCCCAGTTCCTCAGCAGGACGCCCAACCTCCGCCTGAGGATCTTCGTCTCGCGCCTCTACTTCTGCGACCTGG AGGACAGCCTCGCAAGAGAGGGCCTGAGGACACTGAAGAGAGTCGGCGTGCAGGTCACTGTCATGAGCTACAAAGACTACTTCTACTGCTGGCAG ACCTTCGTGGCTCGCAGACAGAGCAGCTTCAAGGCTTGGGATGGGCTGCAGCAGAACTCTGTCCGCCTGGCCAGGAAACTCAACCGCATCCTCCAG CCTTGTGAGACAGAAGACTTACGAGATGCATTCAAGCTTCTTGGACTGTGA
Perca fluviatilis	ATGATTACAAAGCTAGACAGTGTGCTTTTGCCCCGAAAAAAGTTCATCTACCATTACAAGAACATGCGCTGGGCAAGGGGTCGCCATGAGACATAT CTCTGCTTTGTAGTGAAGAGGCGAGTGGGGCCAGACTCCTTATCCTTTGACTTTGGACACCTCCGCAATCGCAATGGCTGCCATGTAGAGCTGCTGT TCCTGCGCTACATTGGAGCCTTGTGCCCTGGTTTGTGGGGATGCAGCGGTACTGGAGAGAGGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG GTCTCCTTGTGCCAACTGCTCCATCAGACTGTCCCAGTTCCTCAGCCAGACGCCCAACCTTCGCCTAAGGATTTTCGTCTCTCGCCTTTACTTCTGTG ACACGGAGAACAGCCCTGAAAGAGACGGCCTAAGAATGCTGAAAAAAGCTGGCGTGCAGATCACAGTCATGAGTTACAAAGACTTCTTTTATTGCT GGCAGACCTTTGTGGATCGTAAGCAAAGCAACTTCAAGGCCTGGGAAGAGCTGCACTCAAACTCTGTTCGCCTTTCCAGAAAACTCAACCGCATCC TCCAGCCTTTTGAAACAGAAGATTTAAGAGATGCCTTCAAGCTTCTTGGACTGTGA
Percopsis transmontana	ATGATTACCAAGCTAGACAGTGTGCTTCTGGCGCAGAAGAAATTCATCTTCCACTACAAGAACATGCGCTGGGCAAGGGGTCGCCATGAGACATAT CTCTGCTTTGTCATTAAGAGGAGAGTGGGGCCAAACTCCCTGTCCTTTGACTTTGGACACCTCCGCAATCGCTCCGGTTGCCATGTAGAGATCCTGTT CCTGCGCCACTTGGGAGCGCTGTGCCCTGGACTGTGGGGAGAGGGGGGTACTGGTGAGAGAAGATTAAGTTACTCCATCACCTGGTTCTGCTCCTG GTCTCCCTGTGCCAACTGCTCCCTCAGACTGGCCCAGATCCTCAGACAGCTGCCCAACCTCCGCCTGAGGATCTTTGTGTCCCGCCTCTACTTCTGTG ACCTGGAGGACAGCAAAGAGAGAGATGGCCTCAGAATGCTGAAGAACGTGGGTGTGCAGATCACCGTCATGAGCTACAAAGACTATTTCTATTGCT GGCAGACCTTTGTAGCTCACAGAAAGAGTAACTTCAAAGCCTGGGACGGGCTGCACCAAAACTCTGTTCGCCTGGCTCGGAAACTCAACCGCATCC TCCAGCCTTGTGAGATAGAAGATTTAAGAGATGCCTTCAAACTTCTTGGGTTTTGA
Phycis blennoides	ATGATTAGTAAGCTAGACAGTGTGCTCTTAGCCCAGAAGAAATTCATATACAATTACAAGAACATACGATGGGCAAAAGGCCGCAACGAGACCTAC CTCTGCTTTGTAGTGAAGAGAAGGCTCGGACCCAATTCCCTGTCCTTCGACTTCGGTCACCTACGCAATCGCGCTGGCTGCCACGTAGAGCTGCTGT TTCTGAGCCACCTGGGGGCGCTGTGCCCGGGCCTCTGGGGGTGCGTGGATGACAGCAACAGGAGACTGAGCTACTCGGTCACCTGGTTCTGCTCCT GGTCTCCCTGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTACGGATGACACCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGT

	GACCTGGAGGACAGTCCGCATATTGAGGGCTTGAGGCACCTGAGGAGAGCAGGGGTTGAGGTCAAAGTTATGAGCTACAAAGACTACTTCTACTGT TGGCAGACCTTCGTAGCTCACAGGCTGAGTCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCAAGAAAACTAAACCGCATC CTCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTTTTGGACTGTTAACCTGA
Phycis phycis	ATGATTAGTAAGCTAGACAGTGTGCTCTTAGCCCAGAAGAAATTCCTATACAATTACAAGAACATACGATGGGCAAAAGGCCGCAACGAGACCTTC CTCTGCTTTGTAGTGAAGAGAAGGCTCGGACCCAATTCCTTGTCCTTCGACTTCGGTCACCTACGCAATCGCGCTGGCTGCCACGTAGAGCTGCTGT TTCTGAGCCACCTGGGGGCGCTGTGCCCGGGCCTCTGGGGGTGCGTAGATGACAGCAACAGGAGACTGAGCTACTCGGTCACCTGGTTCTGCTCCT GGTCTCCATGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTCAGGATGACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTG TGACCTGGAAGACAGTCCGCATATTGAGGGCTTGAGGCACTTGAGGAGAGCGGGGGTCGAGGTCAAAGTTATGAGCTATAAAGACTACTTCTACTG CTGGCAGACCTTCGTAGCTCACAGGCTGAGTCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCAAGAAAACTAAACCGCAT CCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTTTTGGACTGTTAACCTGA
Poecilia formosa	ATGATTACAAAGCTAGACAGGGCACTATTACCCAGAAAAAAATTCATCTATCATTACAAGAACTTGCGCTGGGCAAGAGGTCGATGTGAGACGTAC CTCTGTTTTGTGGTGAAGAAGCGAGTGGGACCAGACTCCCTGTCCTTTGACTTTGGGCATCTCCGCAACCGCAACAACTGCCATGTGGAGCTGCTGT TCCTGCGCCACCTGGGAGCGTTGTGCCCTGGCCTGTGGGGTTATGGAGTCACTGGTGAAAGAAAAGTCAGCTACTCTGTCACCTGGTTTTGCTCCTG GTCTCCCTGTGCAAACTGCTCCATCCGACTGGCTCAGTTCCTCCACCAGACCCCCAACCTCCGCCTCAGGATCTTTGTATCCCGGCTTTATTTCTGCG ACTTGGAGGACAGCCGTGAAAGAGAGGGACTTAGAATACTGAAAAAAGCTGGCGTGCACATCACAGTCATGAGTTACAAAGATTACTTTTACTGCT GGCAGACCTTTGTGGCAAAAAGCCAAAGCAAGTTCAAGCCGTGGGATGGGCTGCACCAAAACTATATCCGGCTGTCAAGGAAACTCAACCGCATTC TTCAGCCATGTGAGACAGAAGATTTAAGAGATGCCTTCAGGCTTCTTGGACTGTGA
Pollachius virens	ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATCTACAATTACAAGAACATGCGATGGGCAAAAGGCCGCAACGAGACCTAT CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCTGATTCCCTCTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGCAGAGCTGCTGT TCCTGAGCTACCTGGGGGCGCTGTGCCCAGGCCTCTGGGGCTGCGCAGACGACAGAAACCGAAGACTAATTTACTCCGTCACCTGGTTCTGCTCCTG GTCGCCCTGTGCCAACTGTGCGACCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGTCTCGCCTCTACTTCTGT GACCTGGAGGGCAGTCCGCATGTAGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAAGACTACTTCTACTG CTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTATGTGCGTCTGTCAAGAAAACTAAACCGCAT CCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGTTTTCGGACTTTTTGGACTGTTAACTTGA
Polymixia japonica	ATGATTACTAAACTAGACAGTGTGCTTTTGGCCCAGAAGAAATTCATCTACCATTATAAGAACATGCGCTGGGCGAAGGGTCGACACGAGACGTAT CTCTGCTTTGTAGTCAAGAGGAGGGTGGGACCGGACTCCATGTCCTTTGATTTTGGACACCTACGCAATCGCTCTGGCTGCCATGTAGAGCTGCTGT TCCTGCGCCACCTGGGGGCCTTGTGCCCTGGACTGTGGGGATACGGAGGTACTGGTGAGAAGAGGCTCAGTTACTCCGTCACCTGGTTCTGCTCCTG GTCGCCCTGCTCCAACTGCTCCTACAGACTGGCCCAGTTCCTCAGCCAGACGCCCAACCTCCGCCTCAGGATCTTCGTCTCTCGACTTTACTTCTGCG ACCTGGAGGACAGCCGGGAGCGAGACGGCCTCAGAATGCTCAAAAGGGCTGGAGTGCAAATCACAGTCATGACCTACAAAGACTACTTCTATTGCT GGCAGACCTTTGTGGCTCACAGAACAAGCAAGTTCAAGGCCTGGGATGAGCTGCACCGGAACTCTGTCCGCCTGTCCAGGATACTCAACCGCATCC TCCAGCCTTGTGAGACAGAAGATTTAAGAGATGCCTTCAGACTTCTTGGGTTGTGA
Pseudochromis fuscus	ATGATTGCAAAGCTTGACAGTGTGCTTTTGCCAAAAAAGAAATTCATCTTTCATTACAAGAACATGCGCTGGGCAAGGGGCCGACATGAGACATAC CTCTGCTTTGTGGTGAAAAGGCGAAGGGGCCCAGACTCTCTGTCCTTTGACTTTGGACATCTCCGCAATCGCAACGGCTGCCATGTAGAGCTGCTAT TCCTACGGTACCTGGGAGCCTTGTGCCCTGGTCTGTGGGGGTATGGGGCTACTGGGGCGAGCAGGCTCAGCTACTCCATCACGTGGTTCTGCTCCTG GTCTCCTTGTGCCAACTGCTCTTTCAGACTGGCCCAGTTCCTCAGCCAGACGCCCAATCTTCGCCTCAGGATCTTCGTCTCTCGCCTTTACTTTTGTGA CATGGAGGACAGCCGTGAAAGAGAGGGTCTAAGGCAGCTGAAAAAAGCCGGAGTGCACATCACAGTCATGAGTTACAAAGACTACTTCTACTGCT GGCAGACCTTTGTGGCTCGTAATCAAAGCAAATTCAAGCCCTGGGATGAATTGCACCAAAACTCTGTCCGCCTGTCCAGAAAACTCAACCGCATCCT CCAGCCTTGTGAGACAGAAGATTTAAGAGATGCCTTCAAGCTTCTTGGACTGTGA
Rondeletia loricata	ATGATTACAAAACTAGACAGTGTGCTTTTGGCCAAGAAAAAGTTCATCTACCATTATAAGAACATGCGCTGGGCAAGGGGTCGGCATGAGACATAC CTCTGCTTTGTAGTGAAGAGGCGAGTGGGGCCAGACTCCCTGTCCTTCGACTTTGGACACCTCCGCAACCGCACTGGCTGCCATGTAGAGCTGCTGT TCCTGCGCCACCTGGGAGCCTTGTGCCCTGGACTGTGGGGGCATGGAGGCACTGGAGAGAGGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG GTCTCCCTGCGCTGACTGCTCCTTCAGACTGGCCCAGTTCCTCGGCCGGATGCCCAACCTCCGCCTCAGGATCTTCGTCTCTCGCCTCTACTTCTGCG ACCTGGAGGACAGCCGCGAGAGAGAGGGCCTGAGGTTGCTGAAAAAAGCCGGCGTGCAGATCACTGTCATGAGCTACAAAGACTTCTTCTATTGCT

	GGCAGACCTTTGTGGCTCATAGAAATTGCAGCTTCAAGGCCTGGGATGAGATGCATCAAAACTCTGTTCGCCTGGCCAGGAAACTCAACCGCATCC TGCAGCCTTGTGAGACAGAAGATTTAAGAGATGCGTTCAAGCTTCTTGGGTTGTGA
Salmo salar 1	NCBI Reference Sequence: XM_014151382.1
Salmo salar 2	NCBI Reference Sequence: XM_014154598.1
Sebastes norvegicus	ATGATTACAAAGCTAGACAGTGTGCTTTTGCCTCGAAAAAAGTTCATCTTCCATTACAAGAACATGCGCTGGGCAAGAGGCCGGCATGAGACATAC CTCTGCTTCGTAGTGAAGAGGCGAGTGGGGCCAGACTCCTTAACCTTTGACTTTGGACACCTCCGCAATCGCAATGGCTGCCATGTAGAGCTGCTGT TCATGCGCTACCTGGGAGCCTTGTGCCCTGGTTTGTGGGGGCAGGGAGTCCCCGGAGAGAAGAGGCTCAGTTACTCCATCACCTGGTTTTGCTCCTG GTCTCCCTGCGTCAACTGCTCCGTCACACTGTCCCAGTTCCTCAGCAAAACGCCCAACCTTCGCCTCAGGATCTTCGTCTCTCGCCTTTACTTCTGTG ACATGGAGAACAGCCGTGAAAGAGATGGACTAAGAATGCTGAAAAAAGCTGGCGTGCAGATCTCAGTCATGAGTTACAAAGACTACTTCTATTGCT GGCAGACCTTTGTGGATCGGAAGCAGAGCAAGTTCAAGGCCTGGGATGAGATGCACCAAAACTCTGTTCGCCTTACCAGAAAACTCAGCCGCATCC TCCAGCCTAGTGAAACAGAAGATTTAAGGGATGCCTTCAAGCTTCTTGGACTGTGA
Selene dorsalis	ATGATTACTAAGCTAGACAGTGTGCTTTTGCCCCGCAAAAAGTTTATCTTCCATTACAAGAATGTGCGCTGGGCGAAGGGCCGGCATGAGACATAC CTCTGCTTTGTTGTGAAGAGACGAGTGGGCCCAGACTCCATGACTTTTGACTTTGGACACCTCCGCAATCGTAATGGCTGCCATGTAGAGATTCTGT TCCTGCGTTACCTTGGTGCCTTGTGTCCTGGTCTATGGGGTTATGGGGTTGGTGGAGAGAAGAGACTCAGTTACTCCATCACCTGGTTCTGCTCCTGG TCTCCCTGTGCCAACTGCTCCAGCAGGCTGGCCCAGTTCTTAAAGCAGACGCCCAACCTTCGCCTAAGGATCTTCGTTTCACGCCTTTATTTCTGTGA CTTGGAGGACAGCCAAGAGAGAGAGGGCCTGAGGATATTGAAAAAAGCTGGAGTGCACATAACAGTCATGACTTACAAAGACTTCTTCTATTGCTG GCAGACCTTTGTGGCTCGTAAACAGAGTAGCTTCAAAGCCTGGGATGAGCTGCACCAAAATTCTGTTCGTCTTGCTAGAAAACTTCAGCGTATCCTC CAGCCATGTGAAACAGAAGATTTGAGGGATGCCTTCAAACTTCTTGGACTGTGA
Spondyliosoma cantharus	ATGATTACAAAGCTAGACAGTGTGCTTTTGCCTAAAAAAAAATTCATCTACCACTACAAGAATGTGCGCTGGGCAAGGGGCCGACATGAGACTTAC CTGTGCTTTGTAGTGAAGAGGCGAGTGGGGCCAGACACCTTAACCTTTGACTTCGGACACCTCCGCAATCGCAACGGCATCCATGTCGAGTTGCTGT TCCTGCGCTATCTGGGAGCCTTGTGCCCTGGTTTGTGGGGGTATGGAGGCACTGGAGAGAAGAGGCTGAGTTACTCTATCACCTGGTTCTGCTCCTG GTCTCCCTGTGCCAACTGCTCACTCAGACTGTGCCAGTTCCTCAGCCAGACTCCCAACCTTCGCCTTAGGATCTTCGTCTCTCGCCTCTACTTCTGTG ACATGGAGGACAGCCGTGAAAGAGAGGGCCTAAGAATGCTGAAAAAAGCCGGCGTGCAGATCACAGTCATGAGTTACAAAGACTTCTTCTATTGC TGGCAGACATTTGTGGCTCGTAGGGCGAGCCAATTCAAGGCCTGGGAAGAGCTGCAACGTAACTCTGTTCGCCTTACCAGAAAACTGAACCGCATC CTCCAGCCCTGTGAAACAGAAGATTTAAGAGATGCCTTTAAGCTTCTTGGACTGTGA
Stylephorus chordatus	ATGATTGCAAAACTAGACAGTGTGCTTCTGGCCCGGAATAAATTCATCTACCATTATAAGAACATGCGCTGGGCGAAAGGGCGCAACGAGACCTAC CTCTGCTTTGTAGTGAAGAGAAGGGTTGGACCTGATTCCCTGGCTTTCGACTTTGGACACCTCCGCAATCGTACCGGTTGCCACGTAGAGCTCCTGTT CCTGCGCCACCTGGGGGCCCTGTGCCCTGGACTGTGGGGAGGTGCTGCTGGTGATAAAAGGCTCAGCTATTCAGTCACCTGGTTCTGCTCTTGGTCT CCCTGTGCCAACTGTGCTTCCACGCTGGCCCAATTCCTGAGACAGACGCCAAACCTCCGTCTCAGGCTCTTTGTGGCTCGTCTCTACTTCTGTGACCT GGAGGATAGTCCTGACAGAGAGGGCCTACGGATTTTGAGAAGAGCTGGGGTGCATATCACAGTTATGAGATACAAAGACTACTTCTACTGCTGGCA GACCTTTGTGGCTCACAACCAGAGCCGCTTCAAAGCCTGGGAAGGACTGCACCCAAACTCTGTCCGTTTGTCCAGAACATTAAACCGCATCCTCCAG CCTTGTGAAACAGAAGATTTAAGAGATGCTTTCAAACTCCTTGGATTGTAA
Symphodus melops	ATGAATACAAAACTCGACAGCGTGCTTTTGCCACGAAAGAAGTTCATTTACCATTACAAGAACGTGCGCTGGGCAAGGGGCCGGCATGAGACATAC CTTTGCTTTGTAATCAAGAGACGGGTGGGGCCGGACACCTTAACCTTTGATTTTGGACACCTCCGCAATCGCAATGGCTGCCATGTAGAGCTGCTGT TCCTGCGCTACCTGGGGGGCCTTGTGTCCTGGTTTATTGGGGTATGGAGGCGCCGGAGAGAAGAGGCTCAGCTACTCTATCACCTGGTTCTGCTCCTG GTCTCCATGCTCCAACTGCTCCACAATACTTTGCCAGTTCCTCAGTAAGATGCCCAACCTTCGCCTCCGGCTCTTCGTCTCTCGCCTTTACTTCTGTGA CATGGAGGATAGTCGTGAAAGAGAGGGCTTAAGAATGCTGAAAAAAGTCGGGGTGCAGATCACAATCATGAGTTACAAAGATTTCTTCTATTGTTG GCAGAAATTTGTGGCACGTAGGCAAAGCAACTTCAAGGCATGGGAAGAGCTGCACCAGAACTCTGTTCGTCTTTCCAGGAAACTCAACCGCATCCT ACAGCCCTGTGAAACAGAAGACTTGAGAGATGCGTTCAAGCTTCTTGGACTTTGA
Takifugu rubripes	NCBI Reference Sequence: XM_003966246.3
Tetraodon nigroviridis	ATGATTACCAAGCTAGACAGTATGCTTTTGCCAAGAAAAAAGTTCCTCTACCATTACAAGAACGTGCGATGGGCGCGGGGCCGACACGAGACCTAC CTCTGCTTTGTTGTGAAGCGGAGAGTGGGCCCAGACACGCTAACCTTTGACTTCGGGCACCTCCGCAATCGCAACGGTTGCCACGTAGAGCTGCTCT

	TCCTGCGCTACCTGGGGGCCCTGTGCCCGGGTTTGTGGGGTTATGGCGCTGCCGGGGAGAAGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG GTCCCCCTGCGCCAACTGCTCCATCCAACTTTCCCAGTTTCTGAGGAACACGCCCAACCTTCGCCTCAGAATCTTTGTCTCCCGCCTTTACTTCTGTG ACATGGAGGACAGCCTTGAACGGGAAGGCCTGAGGATGCTGTCCAGGGCCGGCGTGAGGATTTCAGTGATGAGCTACAAAGACTTTTTCTATTGCT GGCAGAAATTTGTGGATAGCAAAACGAGCAGCTTTAAAGCCTGGGAAGAGCTGCACCAGAACTCTGTACGCCTCACTCGAAAACTCAACCGCATTC TCCAGAGCTGGGATTTAGAAGATTTACGAGACGCCCTTAAGCTTCTTGGACTCTAA
Theragra chalcogramma	ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAAAAATTCATCTACAATTACAAGAACATGCGATGGGCAAAAGGCCGCAACGAGACCTAT CTCTGCTTCGTAGTGAAGAGAAGGCTTGGACCTGATTCCCTCTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGCAGAGCTGCTGT TCCTGAGCTACCTGGGGGCGCTGTGCCCGGGCCTCTGGGGCTGCGCAGACGACAGAAACCGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT GGTCGCCCTGTGCCAACTGTGCGACCACGCTGACCCGGTTCCTGAGGCAGACACCCAACCTGCGACTCAGGATCTTCGTGTCTCGCCTCTACTTCTG TGACCTGGAGGGCAGTCCGCATGTAGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAAGACTACTTCTACT GCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTATGTGCGTCTGTCAAGAAAACTAAACCGCA TCCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACCTGA
Thunnus albacares	ATGATTACAAAACTAGACAGTGTGCTTTTGCCCCGGAAAAAGTTCATCTACCATTACAAGAACGTGCGCTGGGCAAGAGGACGGCATGAAACATAC CTCTGCTTTGTAGTGAAGAGGCGAGTGGGGCCAGACTCTTTATCCTTTGACTTTGGACACCTGCGCAATCGCAATGGCTGCCATGTAGAGCTGCTGT TCCTGCGATATCTGGGAGCCTTGTGCCCTGGTGTGTGGGGGTATGGAAATACTGGACAGAGGATCAGTTACTCCATCACCTGGTTCTGCTCTTGGTC TCCCTGTGCCAACTGCTCTCGCAGACTGGCCCAGTTCCTCAGCCAGGTACCCAACGTTCGCCTTAGGATCTTCGTATCACGCCTCTACTTTTGTGACT TGGAGGACAGCCGTGAGAGAGACGGCCTGAGGTTGCTAAAAAACGCCGGCGTGCAGATCACAGTCATGAGTTACAAAGACTTCTTCTACTGCTGGC AGACTTTTGTAGCTCGTAATCAGAGCAAATTCAAGGCCTGGGAAGAGCTGCACCGAAACTCTGTTCGCCTAACAAGAACACTCAACCGCATACTCC AGCCCTGTGACATTGATGATTTAAGAGATGCCTTCAAGCTTCTTGGGCTGTGA
Trachyrincus murrayi	ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATCTACAATTACAAGAACATGCGTTGGGCAAAAGGCCGCAACGAGACCTAC CTATGCTTTGTGGTGAAGAGAAGGCTTGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCCTTGGCTGCCACGTAGAGCTGCTGTT TCTGAGCCACCTGGGGGCGCTGTGCCCGGGCCTGTGGGGGTGTGGAGGCGACGTAAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTG GTCTCCCTGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGT GACCTGGAGGACAGTCCGCATATAGAGGGCTTGAGGGATCTGAGGAGAGCAGGGGTCCAGGTCACCGTTATGAGCTACAAAGACTACTTCTACTGC TGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAAAACTAAACCGCATC CTCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGACTGTTAACCTGA
Trachyrincus scabrus	ATGATAAGTAAGCTAGACAGTGTGCTCTTGGCTCAGAAGAAATTCATCTACAATTACAAGAACATGCGTTGGGCAAAAGGCCGGAATGAGACCTAC CTATGCTTTGTGGTGAAGAGAAGGCTTGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCCTTGGCTGCCACGTAGAGCTGCTGTT TCTGAGCCACCTGGGGGCACTGTGCCCGGGCCTGTGGGGGTGCGGAGGCGACGAAAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTG GTCTCCCTGCGCCAACTGTGCGGCCACACTGGCCCGGTTCCTGAGGCACACGCCCAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGT GACCTGGAGGACAGTCCGCATATAGAGGGCTTGAGGGATCTGAGGAGAGCAGGGGTCCAGGTCACTGTTATGAGCTACAAAGACTACTTCTACTGC TGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGAAAACTAAACCGCATC CTCCAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGACTGTTAACCTGA
Trisopterus minutus	ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTTATATACAATTACAAGAACCTACGATGGGCAAAAGGACGCAACGAGACCTAC CTCTGCTACGTAGTGAAGAGGAGGCTCGGACCTGATTCCCTCTCCTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTAGAGCTGCTGT TCCTCAGCTACCTTGGGGCACTATGCCCGGGCCTCTGGGGCTGCACCGATGACAGAAACCGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCTG GTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGTCGCCCGCCTCTACTTCTGC GACCTGGAGGGCAGTCCGCACATAGAGGGCTTGAGGCACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTCATGAGCTACAAAGACTACTTCTACTG CTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGTGCGTCTGTCAAGAAAACTAAACCGCAT CCTCCAGCCATGTGAAACAGAAGATTTAAGAGATGTTTTCGGACTTTTTGGACTGTTAACCTGA
Typhlichthys subterraneus	ATGATTAGCAAGCTAGACAGTGTCCTTCTGGCGCAGAAGAAATTCATCTTCCACTATAAGAATATGCGCTGGGCAAGGGGGCGCAATGAGACATAT CTCTGCTTTGTCATTAAGAGGAGAGTGGGGCCGGACTCCCTGTCCTTTGACTTTGGACACCTCCGCAATCGCTCCGGCTGCCATGTAGAGCTGCTGTT CCTGCGCCACTTGGGGGCGCTGTGCCCTGGCCTGTGGGGACAGGGGGGTACAGGTGACAACAGACTCAGTTACTCCATCACCTGGTTCTGCTCCTGG

	TCCCCCTGTTCCAACTGCTCTCACAGACTGGCCCAGTTCCTCAGCCAGCTGCCCAACCTCCGCCTGAGGATCTTTGTGTCCCGTCTGTACTTCTGTGA CCTGGAGGACAGCAGGGAGAGAGAGGGCCTGAGAATGCTGAAGAATGCGGGCGTGCACATAACCGTCATGAGCTACAAAGACTATTACTATTGCT GGCAGACCTTTGTAGCTCGCAGAAAGAGTAAATTCAAAGCATGGGAAGGGCTGCACCAAAACTCTGTTCGCCTGGCCAGGAAACTCAACCGCATCC TCCAGCCGTGCGAGATAGAAGATTTAAGAGATGCCTTCAAACTTCTTGGGTTTTGA
Xiphophorus maculatus	ATGATTACAAAGCTAGACAGGGTACTATTACCCAAAAAAAAATTCATCTATCATTACAAGAACATGCGCTGGGCAAGAGGTCGATGTGAGACATAC CTCTGCTTTGTGGTGAAGAAGCGAGTGGGACCAGACTCCCTGTCCTTTGACTTTGGACATCTCCGCAACCGCAACAACTGTCATGTGGAGCTGCTGT TCCTGCGCCACCTGGGAGCGTTGTGCCCTGGCCTGTGGGGTTATGGAGTCACTGGTGAGAGAAAAGTCAGCTACTCCATCACCTGGTTTTGCTCCTG GTCTCCCTGTGCAAACTGCTCCTTCCGACTGGCTCAGTTCCTCCACCAGACCCCCAACCTCCGCCTCAGGATCTTTGTATCCCGGCTTTATTTCTGTG ACTTGGAGGACAGCCGTGAAAGAGAGGGACTTAGAATGCTGAAAAAAGCTGGCGTGCACATCACAGTCATGAGTTACAAAGATTACTTTTACTGCT GGCAGACCTTTGTGGCAAAAAGTCAAAGCAAGTTCAAGCCGTGGGATGGGCTGCACCAAAACTGTATCCGGCTGACAAGGAAACTCAACCGCATA CTTCAGCCATGTGAGACAGAAGATTTAAGAGATGCCTTCAGGCTTCTTGGACTGTGA
Zeus faber	ATGATAACTAAACTAGACAGTGTGCTTCTGGCTCGGAAGAAATTCATTTACCACTATAAGAACATGCGCTGGGCAAAAGGCCGCTGTGAGACGTAC CTCTGCTTTGTCGTCAAGAGGAGAGTTGGACCCAATTCCCTGTCCTTTGACTTTGGACACCTTCGCAATCGGACCGGCTGCCATGTAGAGCTCCTGTT TCTACGTCACCTGGGAGCCTTGTGCCCTGGACTGTGGGGACACGGAGGCCCCTATGGAGGGCGGCTCAGTTACTCAGTCACCTGGTTCTGCTCGTGG TCTCCCTGCGCCAACTGCTCCTTCAGACTGGCCCAATTCCTCGGGCAGACGCCCAACCTCCGCCTCAGGATCTTTGTCTCCCGCCTCTACTACTGCGA CCTTGAAGATAGCCGCGAGAGAGAGGGCTTACGGATCCTGAAAAGAGCCGGAGTCCAAATCACAGTCATGAGCTACAAAGACTACTTCTATTGCTG GCAAACCTTCGTGGCTCACAGACAGACCAGCTTCAAGGCGTGGGATGAGCTGCACCAAAACTCAGTTCGCCTGGCCAGGAAACTAAACCGCATCCT CCAGCCTTGTGAAACAGAAGATTTAAGAGATGCCTTCAAACTTCTTGGGTTCTTGTGA
Lampetra tridentata	ATGGCCAACGATGAGTACGTGAGAGTCGGCGATAAGTTGGACAGCTGCACGTTTAGGACGCAGTTTTTTAACTTTAAAAGATCCACGTCGCATATA TGCTGCGTTCTCTTTGAATTAAAACAGCAGGATAGCGTCGCTTTTTGGGGCTATGCTGTGAATAAACCACGGAGCAATGCAGACCTAGGAATTCACG CCGAAATTTTTTGCATTAAAAAAATCAGAGAGTACCTGCACGAAAACCCTGGAATATACACGATAAATTGGTACTCATCCTGGAGTTCGTGTGCAG ATTGCGCTGAAGAGATCTTAACATGGTATAAGAAGGAGGTGATGAAGATGGGCCACACTTTGAATATCTGGGCTTGCAAACTCTATTTCGAGAACA TTACGCGGAATCAAATTGGGTTGTGGAACCTCAGAAAAATCGGGGTTGGGTTGGAAATAATGCTTGGTGAACACTACCAATGGTGCTGGAACAACT ACATCCAAACGTTGGACAGCAATTTGAATGAAAATAGATGGCTTCAGAAGACTTCGAATCGAGCTCTTACACGACAGAACGAGTTGTCCATTATGA TTCAGGTAAAAAGACTCCACACCGCTAAGACTCCTGCTGTTTAG

Appendix 8: Our computationally predicted 3D structure of Gm-AID used to guide amino acid alignment and as the structure template in ProtASR analyses

```
ATOM      1  N   MET A   1      45.786  76.689  64.991  1.00  7.26
ATOM      2  CA  MET A   1      45.759  77.128  63.593  1.00  7.26
ATOM      3  HA  MET A   1      45.603  76.258  62.956  1.00  7.26
ATOM      4  CB  MET A   1      44.601  78.122  63.380  1.00  7.26
ATOM      5  HB1 MET A   1      44.602  78.442  62.338  1.00  7.26
ATOM      6  HB2 MET A   1      44.771  79.001  64.005  1.00  7.26
ATOM      7  CG  MET A   1      43.213  77.554  63.713  1.00  7.26
ATOM      8  HG1 MET A   1      43.228  77.183  64.738  1.00  7.26
ATOM      9  HG2 MET A   1      43.002  76.708  63.060  1.00  7.26
ATOM     10  SD  MET A   1      41.853  78.755  63.592  1.00  7.26
ATOM     11  CE  MET A   1      41.751  79.020  61.801  1.00  7.26
ATOM     12  HE1 MET A   1      40.959  79.739  61.590  1.00  7.26
ATOM     13  HE2 MET A   1      42.695  79.409  61.420  1.00  7.26
ATOM     14  HE3 MET A   1      41.517  78.082  61.304  1.00  7.26
ATOM     15  C   MET A   1      47.064  77.797  63.159  1.00  7.26
ATOM     16  O   MET A   1      47.348  77.866  61.964  1.00  7.26
ATOM     17  N   ILE A   2      47.834  78.273  64.141  1.00  6.71
ATOM     18  H   ILE A   2      47.484  78.131  65.078  1.00  6.71
ATOM     19  CA  ILE A   2      49.173  78.835  63.960  1.00  6.71
ATOM     20  HA  ILE A   2      49.275  79.188  62.932  1.00  6.71
ATOM     21  CB  ILE A   2      49.423  80.038  64.901  1.00  6.71
ATOM     22  HB  ILE A   2      49.291  79.710  65.934  1.00  6.71
ATOM     23  CG2 ILE A   2      50.868  80.550  64.735  1.00  6.71
ATOM     24 1HG2 ILE A   2      51.064  81.374  65.419  1.00  6.71
ATOM     25 2HG2 ILE A   2      51.592  79.770  64.975  1.00  6.71
ATOM     26 3HG2 ILE A   2      51.034  80.888  63.711  1.00  6.71
ATOM     27  CG1 ILE A   2      48.403  81.166  64.612  1.00  6.71
ATOM     28 1HG1 ILE A   2      47.392  80.766  64.692  1.00  6.71
ATOM     29 2HG1 ILE A   2      48.540  81.526  63.592  1.00  6.71
ATOM     30  CD1 ILE A   2      48.486  82.359  65.574  1.00  6.71
ATOM     31  HD1 ILE A   2      47.652  83.034  65.381  1.00  6.71
ATOM     32  HD2 ILE A   2      48.428  82.010  66.606  1.00  6.71
ATOM     33  HD3 ILE A   2      49.414  82.909  65.424  1.00  6.71
ATOM     34  C   ILE A   2      50.192  77.721  64.149  1.00  6.71
ATOM     35  O   ILE A   2      50.756  77.295  63.158  1.00  6.71
ATOM     36  N   SER A   3      50.388  77.184  65.359  1.00  6.18
ATOM     37  H   SER A   3      49.926  77.600  66.153  1.00  6.18
ATOM     38  CA  SER A   3      51.432  76.176  65.639  1.00  6.18
ATOM     39  HA  SER A   3      52.312  76.456  65.059  1.00  6.18
ATOM     40  CB  SER A   3      51.889  76.231  67.098  1.00  6.18
ATOM     41  HB1 SER A   3      52.640  75.460  67.280  1.00  6.18
ATOM     42  HB2 SER A   3      51.039  76.075  67.763  1.00  6.18
ATOM     43  OG  SER A   3      52.472  77.496  67.338  1.00  6.18
ATOM     44  HG  SER A   3      53.299  77.542  66.812  1.00  6.18
ATOM     45  C   SER A   3      51.123  74.737  65.166  1.00  6.18
ATOM     46  O   SER A   3      51.425  73.736  65.821  1.00  6.18
ATOM     47  N   LYS A   4      50.522  74.656  63.978  1.00  5.46
ATOM     48  H   LYS A   4      50.303  75.551  63.556  1.00  5.46
ATOM     49  CA  LYS A   4      50.555  73.526  63.049  1.00  5.46
ATOM     50  HA  LYS A   4      51.438  72.914  63.241  1.00  5.46
```

```
ATOM     51  CB  LYS A   4      49.276  72.650  63.140  1.00  5.46
ATOM     52  HB1 LYS A   4      49.380  71.884  62.369  1.00  5.46
ATOM     53  HB2 LYS A   4      48.412  73.262  62.872  1.00  5.46
ATOM     54  CG  LYS A   4      48.927  71.908  64.441  1.00  5.46
ATOM     55  HG1 LYS A   4      48.499  72.608  65.159  1.00  5.46
ATOM     56  HG2 LYS A   4      49.832  71.481  64.862  1.00  5.46
ATOM     57  CD  LYS A   4      47.910  70.789  64.108  1.00  5.46
ATOM     58  HD1 LYS A   4      48.360  70.120  63.370  1.00  5.46
ATOM     59  HD2 LYS A   4      47.024  71.242  63.656  1.00  5.46
ATOM     60  CE  LYS A   4      47.472  69.945  65.315  1.00  5.46
ATOM     61  HE1 LYS A   4      47.000  70.603  66.049  1.00  5.46
ATOM     62  HE2 LYS A   4      48.364  69.509  65.774  1.00  5.46
ATOM     63  NZ  LYS A   4      46.530  68.861  64.917  1.00  5.46
ATOM     64  HZ1 LYS A   4      46.269  68.262  65.689  1.00  5.46
ATOM     65  HZ2 LYS A   4      45.652  69.233  64.556  1.00  5.46
ATOM     66  HZ3 LYS A   4      46.913  68.267  64.192  1.00  5.46
ATOM     67  C   LYS A   4      50.665  74.053  61.596  1.00  5.46
ATOM     68  O   LYS A   4      49.871  73.684  60.725  1.00  5.46
ATOM     69  N   LEU A   5      51.502  75.062  61.380  1.00  5.61
ATOM     70  H   LEU A   5      52.194  75.249  62.100  1.00  5.61
ATOM     71  CA  LEU A   5      51.579  75.860  60.151  1.00  5.61
ATOM     72  HA  LEU A   5      51.653  75.182  59.298  1.00  5.61
ATOM     73  CB  LEU A   5      50.298  76.732  60.013  1.00  5.61
ATOM     74  HB1 LEU A   5      50.436  77.675  60.541  1.00  5.61
ATOM     75  HB2 LEU A   5      49.466  76.225  60.501  1.00  5.61
ATOM     76  CG  LEU A   5      49.853  77.023  58.570  1.00  5.61
ATOM     77  HG  LEU A   5      49.699  76.080  58.044  1.00  5.61
ATOM     78  CD1 LEU A   5      48.521  77.785  58.587  1.00  5.61
ATOM     79 1HD1 LEU A   5      48.231  78.049  57.575  1.00  5.61
ATOM     80 2HD1 LEU A   5      47.753  77.162  59.043  1.00  5.61
ATOM     81 3HD1 LEU A   5      48.626  78.695  59.178  1.00  5.61
ATOM     82  CD2 LEU A   5      50.862  77.861  57.784  1.00  5.61
ATOM     83 1HD2 LEU A   5      50.466  78.094  56.797  1.00  5.61
ATOM     84 2HD2 LEU A   5      51.081  78.784  58.320  1.00  5.61
ATOM     85 3HD2 LEU A   5      51.784  77.296  57.652  1.00  5.61
ATOM     86  C   LEU A   5      52.852  76.724  60.207  1.00  5.61
ATOM     87  O   LEU A   5      53.723  76.624  59.348  1.00  5.61
ATOM     88  N   ASP A   6      53.018  77.447  61.319  1.00  5.08
ATOM     89  H   ASP A   6      52.194  77.509  61.904  1.00  5.08
ATOM     90  CA  ASP A   6      54.285  77.539  62.051  1.00  5.08
ATOM     91  HA  ASP A   6      54.996  78.114  61.457  1.00  5.08
ATOM     92  CB  ASP A   6      54.054  78.294  63.372  1.00  5.08
ATOM     93  HB1 ASP A   6      53.074  78.045  63.768  1.00  5.08
ATOM     94  HB2 ASP A   6      54.062  79.367  63.174  1.00  5.08
ATOM     95  CG  ASP A   6      55.099  77.965  64.438  1.00  5.08
ATOM     96  OD1 ASP A   6      56.298  78.151  64.136  1.00  5.08
ATOM     97  OD2 ASP A   6      54.680  77.501  65.525  1.00  5.08
ATOM     98  C   ASP A   6      54.864  76.133  62.256  1.00  5.08
ATOM     99  O   ASP A   6      54.154  75.181  62.592  1.00  5.08
ATOM    100  N   SER A   7      56.154  76.004  61.965  1.00  4.11
ATOM    101  H   SER A   7      56.701  76.848  61.874  1.00  4.11
ATOM    102  CA  SER A   7      56.771  74.746  61.570  1.00  4.11
ATOM    103  HA  SER A   7      56.189  73.915  61.970  1.00  4.11
```

```
ATOM    104  CB  SER A   7      56.761  74.630  60.045  1.00  4.11
ATOM    105  HB1 SER A   7      57.414  75.387  59.608  1.00  4.11
ATOM    106  HB2 SER A   7      55.747  74.778  59.671  1.00  4.11
ATOM    107  OG  SER A   7      57.208  73.344  59.679  1.00  4.11
ATOM    108  HG  SER A   7      56.455  72.741  59.781  1.00  4.11
ATOM    109  C   SER A   7      58.186  74.660  62.122  1.00  4.11
ATOM    110  O   SER A   7      59.122  75.262  61.596  1.00  4.11
ATOM    111  N   VAL A   8      58.325  73.985  63.264  1.00  3.62
ATOM    112  H   VAL A   8      57.506  73.520  63.622  1.00  3.62
ATOM    113  CA  VAL A   8      59.403  74.280  64.218  1.00  3.62
ATOM    114  HA  VAL A   8      59.609  75.348  64.122  1.00  3.62
ATOM    115  CB  VAL A   8      58.948  74.094  65.686  1.00  3.62
ATOM    116  HB  VAL A   8      58.899  73.030  65.921  1.00  3.62
ATOM    117  CG1 VAL A   8      59.920  74.781  66.659  1.00  3.62
ATOM    118 1HG1 VAL A   8      60.929  74.388  66.541  1.00  3.62
ATOM    119 2HG1 VAL A   8      59.930  75.857  66.475  1.00  3.62
ATOM    120 3HG1 VAL A   8      59.599  74.601  67.684  1.00  3.62
ATOM    121  CG2 VAL A   8      57.559  74.701  65.946  1.00  3.62
ATOM    122 1HG2 VAL A   8      57.314  74.642  67.006  1.00  3.62
ATOM    123 2HG2 VAL A   8      57.536  75.749  65.639  1.00  3.62
ATOM    124 3HG2 VAL A   8      56.788  74.159  65.400  1.00  3.62
ATOM    125  C   VAL A   8      60.727  73.586  63.905  1.00  3.62
ATOM    126  O   VAL A   8      61.138  72.684  64.628  1.00  3.62
ATOM    127  N   LEU A   9      61.358  73.981  62.800  1.00  2.69
ATOM    128  H   LEU A   9      60.879  74.673  62.233  1.00  2.69
ATOM    129  CA  LEU A   9      62.575  73.386  62.237  1.00  2.69
ATOM    130  HA  LEU A   9      62.264  72.506  61.675  1.00  2.69
ATOM    131  CB  LEU A   9      63.224  74.370  61.245  1.00  2.69
ATOM    132  HB1 LEU A   9      64.143  73.918  60.878  1.00  2.69
ATOM    133  HB2 LEU A   9      63.496  75.273  61.793  1.00  2.69
ATOM    134  CG  LEU A   9      62.375  74.775  60.023  1.00  2.69
ATOM    135  HG  LEU A   9      61.445  75.225  60.363  1.00  2.69
ATOM    136  CD1 LEU A   9      63.134  75.814  59.199  1.00  2.69
ATOM    137 1HD1 LEU A   9      62.517  76.132  58.358  1.00  2.69
ATOM    138 2HD1 LEU A   9      63.356  76.684  59.816  1.00  2.69
ATOM    139 3HD1 LEU A   9      64.062  75.391  58.817  1.00  2.69
ATOM    140  CD2 LEU A   9      62.051  73.607  59.093  1.00  2.69
ATOM    141 1HD2 LEU A   9      61.422  73.958  58.274  1.00  2.69
ATOM    142 2HD2 LEU A   9      62.973  73.197  58.682  1.00  2.69
ATOM    143 3HD2 LEU A   9      61.504  72.835  59.631  1.00  2.69
ATOM    144  C   LEU A   9      63.597  72.900  63.284  1.00  2.69
ATOM    145  O   LEU A   9      64.041  73.647  64.158  1.00  2.69
ATOM    146  N   LEU A  10      63.961  71.621  63.173  1.00  2.50
ATOM    147  H   LEU A  10      63.623  71.113  62.362  1.00  2.50
ATOM    148  CA  LEU A  10      64.843  70.917  64.092  1.00  2.50
ATOM    149  HA  LEU A  10      64.462  71.078  65.100  1.00  2.50
ATOM    150  CB  LEU A  10      64.800  69.404  63.776  1.00  2.50
ATOM    151  HB1 LEU A  10      65.125  69.263  62.744  1.00  2.50
ATOM    152  HB2 LEU A  10      63.769  69.057  63.853  1.00  2.50
ATOM    153  CG  LEU A  10      65.686  68.534  64.696  1.00  2.50
ATOM    154  HG  LEU A  10      66.685  68.949  64.725  1.00  2.50
ATOM    155  CD1 LEU A  10      65.146  68.473  66.129  1.00  2.50
ATOM    156 1HD1 LEU A  10      65.755  67.807  66.733  1.00  2.50
```

```
ATOM    157 2HD1 LEU A  10      65.151  69.463  66.578  1.00  2.50
ATOM    158 3HD1 LEU A  10      64.121  68.097  66.109  1.00  2.50
ATOM    159  CD2 LEU A  10      65.875  67.106  64.205  1.00  2.50
ATOM    160 1HD2 LEU A  10      66.764  66.681  64.656  1.00  2.50
ATOM    161 2HD2 LEU A  10      65.041  66.497  64.520  1.00  2.50
ATOM    162 3HD2 LEU A  10      65.970  67.082  63.121  1.00  2.50
ATOM    163  C   LEU A  10     66.291  71.436  64.025  1.00  2.50
ATOM    164  O   LEU A  10     66.841  71.736  62.966  1.00  2.50
ATOM    165  N   ALA A  11     66.953  71.422  65.182  1.00  2.57
ATOM    166  H   ALA A  11     66.448  71.175  66.015  1.00  2.57
ATOM    167  CA  ALA A  11     68.383  71.640  65.285  1.00  2.57
ATOM    168  HA  ALA A  11     68.590  72.633  64.880  1.00  2.57
ATOM    169  CB  ALA A  11     68.754  71.658  66.773  1.00  2.57
ATOM    170  HB1 ALA A  11     69.816  71.881  66.884  1.00  2.57
ATOM    171  HB2 ALA A  11     68.181  72.431  67.288  1.00  2.57
ATOM    172  HB3 ALA A  11     68.541  70.689  67.225  1.00  2.57
ATOM    173  C   ALA A  11     69.245  70.626  64.505  1.00  2.57
ATOM    174  O   ALA A  11     69.168  69.412  64.714  1.00  2.57
ATOM    175  N   GLN A  12     70.166  71.148  63.694  1.00  2.56
ATOM    176  H   GLN A  12     70.144  72.144  63.536  1.00  2.56
ATOM    177  CA  GLN A  12     71.071  70.366  62.855  1.00  2.56
ATOM    178  HA  GLN A  12     70.467  69.901  62.073  1.00  2.56
ATOM    179  CB  GLN A  12     72.015  71.345  62.155  1.00  2.56
ATOM    180  HB1 GLN A  12     72.517  71.970  62.896  1.00  2.56
ATOM    181  HB2 GLN A  12     71.400  72.000  61.539  1.00  2.56
ATOM    182  CG  GLN A  12     73.057  70.672  61.252  1.00  2.56
ATOM    183  HG1 GLN A  12     73.073  71.190  60.293  1.00  2.56
ATOM    184  HG2 GLN A  12     72.800  69.631  61.054  1.00  2.56
ATOM    185  CD  GLN A  12     74.438  70.752  61.883  1.00  2.56
ATOM    186  OE1 GLN A  12     74.766  70.055  62.833  1.00  2.56
ATOM    187  NE2 GLN A  12     75.267  71.659  61.420  1.00  2.56
ATOM    188 1HE2 GLN A  12     74.927  72.330  60.732  1.00  2.56
ATOM    189 2HE2 GLN A  12     76.155  71.751  61.868  1.00  2.56
ATOM    190  C   GLN A  12     71.780  69.212  63.582  1.00  2.56
ATOM    191  O   GLN A  12     71.677  68.065  63.150  1.00  2.56
ATOM    192  N   LYS A  13     72.399  69.472  64.744  1.00  2.63
ATOM    193  H   LYS A  13     72.506  70.437  65.013  1.00  2.63
ATOM    194  CA  LYS A  13     73.018  68.408  65.555  1.00  2.63
ATOM    195  HA  LYS A  13     73.835  67.974  64.971  1.00  2.63
ATOM    196  CB  LYS A  13     73.587  68.963  66.868  1.00  2.63
ATOM    197  HB1 LYS A  13     73.697  68.133  67.567  1.00  2.63
ATOM    198  HB2 LYS A  13     72.896  69.686  67.304  1.00  2.63
ATOM    199  CG  LYS A  13     74.968  69.601  66.668  1.00  2.63
ATOM    200  HG1 LYS A  13     74.878  70.501  66.057  1.00  2.63
ATOM    201  HG2 LYS A  13     75.615  68.888  66.154  1.00  2.63
ATOM    202  CD  LYS A  13     75.593  69.950  68.025  1.00  2.63
ATOM    203  HD1 LYS A  13     75.547  69.072  68.672  1.00  2.63
ATOM    204  HD2 LYS A  13     75.028  70.760  68.489  1.00  2.63
ATOM    205  CE  LYS A  13     77.058  70.364  67.848  1.00  2.63
ATOM    206  HE1 LYS A  13     77.095  71.309  67.299  1.00  2.63
ATOM    207  HE2 LYS A  13     77.561  69.604  67.241  1.00  2.63
ATOM    208  NZ  LYS A  13     77.741  70.488  69.159  1.00  2.63
ATOM    209  HZ1 LYS A  13     78.709  70.759  69.036  1.00  2.63
```

```
ATOM    210  HZ2 LYS A  13      77.283  71.184  69.735  1.00  2.63
ATOM    211  HZ3 LYS A  13      77.716  69.601  69.647  1.00  2.63
ATOM    212  C   LYS A  13      72.067  67.240  65.822  1.00  2.63
ATOM    213  O   LYS A  13      72.396  66.096  65.514  1.00  2.63
ATOM    214  N   LYS A  14      70.858  67.534  66.320  1.00  2.36
ATOM    215  H   LYS A  14      70.585  68.506  66.337  1.00  2.36
ATOM    216  CA  LYS A  14      69.831  66.514  66.566  1.00  2.36
ATOM    217  HA  LYS A  14      70.238  65.781  67.264  1.00  2.36
ATOM    218  CB  LYS A  14      68.550  67.128  67.161  1.00  2.36
ATOM    219  HB1 LYS A  14      67.691  66.524  66.865  1.00  2.36
ATOM    220  HB2 LYS A  14      68.399  68.130  66.763  1.00  2.36
ATOM    221  CG  LYS A  14      68.581  67.197  68.694  1.00  2.36
ATOM    222  HG1 LYS A  14      67.843  67.931  69.023  1.00  2.36
ATOM    223  HG2 LYS A  14      69.566  67.525  69.027  1.00  2.36
ATOM    224  CD  LYS A  14      68.230  65.840  69.326  1.00  2.36
ATOM    225  HD1 LYS A  14      68.898  65.064  68.946  1.00  2.36
ATOM    226  HD2 LYS A  14      67.204  65.578  69.061  1.00  2.36
ATOM    227  CE  LYS A  14      68.364  65.926  70.847  1.00  2.36
ATOM    228  HE1 LYS A  14      67.773  66.771  71.214  1.00  2.36
ATOM    229  HE2 LYS A  14      69.414  66.114  71.092  1.00  2.36
ATOM    230  NZ  LYS A  14      67.909  64.676  71.498  1.00  2.36
ATOM    231  HZ1 LYS A  14      68.175  64.628  72.469  1.00  2.36
ATOM    232  HZ2 LYS A  14      66.898  64.544  71.442  1.00  2.36
ATOM    233  HZ3 LYS A  14      68.238  63.840  71.018  1.00  2.36
ATOM    234  C   LYS A  14      69.530  65.710  65.307  1.00  2.36
ATOM    235  O   LYS A  14      69.554  64.476  65.374  1.00  2.36
ATOM    236  N   PHE A  15      69.337  66.412  64.181  1.00  2.18
ATOM    237  H   PHE A  15      69.419  67.421  64.211  1.00  2.18
ATOM    238  CA  PHE A  15      69.106  65.760  62.892  1.00  2.18
ATOM    239  HA  PHE A  15      68.164  65.212  62.941  1.00  2.18
ATOM    240  CB  PHE A  15      68.975  66.805  61.761  1.00  2.18
ATOM    241  HB1 PHE A  15      69.725  67.577  61.880  1.00  2.18
ATOM    242  HB2 PHE A  15      68.011  67.302  61.870  1.00  2.18
ATOM    243  CG  PHE A  15      69.107  66.279  60.337  1.00  2.18
ATOM    244  CD1 PHE A  15      67.974  66.151  59.514  1.00  2.18
ATOM    245  HD1 PHE A  15      66.995  66.403  59.897  1.00  2.18
ATOM    246  CE1 PHE A  15      68.111  65.703  58.188  1.00  2.18
ATOM    247  HE1 PHE A  15      67.237  65.609  57.569  1.00  2.18
ATOM    248  CZ  PHE A  15      69.369  65.340  57.681  1.00  2.18
ATOM    249  HZ  PHE A  15      69.465  64.987  56.663  1.00  2.18
ATOM    250  CE2 PHE A  15      70.502  65.457  58.501  1.00  2.18
ATOM    251  HE2 PHE A  15      71.478  65.192  58.121  1.00  2.18
ATOM    252  CD2 PHE A  15      70.374  65.965  59.804  1.00  2.18
ATOM    253  HD2 PHE A  15      71.259  66.109  60.400  1.00  2.18
ATOM    254  C   PHE A  15      70.186  64.724  62.614  1.00  2.18
ATOM    255  O   PHE A  15      69.848  63.577  62.344  1.00  2.18
ATOM    256  N   ILE A  16      71.465  65.088  62.769  1.00  2.39
ATOM    257  H   ILE A  16      71.682  66.035  63.067  1.00  2.39
ATOM    258  CA  ILE A  16      72.553  64.156  62.458  1.00  2.39
ATOM    259  HA  ILE A  16      72.399  63.766  61.452  1.00  2.39
ATOM    260  CB  ILE A  16      73.933  64.865  62.496  1.00  2.39
ATOM    261  HB  ILE A  16      74.082  65.289  63.491  1.00  2.39
ATOM    262  CG2 ILE A  16      75.069  63.855  62.221  1.00  2.39
```

```
ATOM    263 1HG2 ILE A  16      74.944  63.409  61.233  1.00  2.39
ATOM    264 2HG2 ILE A  16      76.041  64.341  62.280  1.00  2.39
ATOM    265 3HG2 ILE A  16      75.067  63.060  62.965  1.00  2.39
ATOM    266  CG1 ILE A  16      73.998  66.010  61.455  1.00  2.39
ATOM    267 1HG1 ILE A  16      73.154  66.678  61.601  1.00  2.39
ATOM    268 2HG1 ILE A  16      73.921  65.593  60.450  1.00  2.39
ATOM    269  CD1 ILE A  16      75.259  66.879  61.539  1.00  2.39
ATOM    270  HD1 ILE A  16      75.150  67.733  60.869  1.00  2.39
ATOM    271  HD2 ILE A  16      75.388  67.246  62.558  1.00  2.39
ATOM    272  HD3 ILE A  16      76.140  66.315  61.237  1.00  2.39
ATOM    273  C   ILE A  16     72.493  62.961  63.410  1.00  2.39
ATOM    274  O   ILE A  16     72.579  61.813  62.966  1.00  2.39
ATOM    275  N   TYR A  17      72.315  63.210  64.712  1.00  2.35
ATOM    276  H   TYR A  17      72.190  64.169  65.016  1.00  2.35
ATOM    277  CA  TYR A  17      72.514  62.165  65.715  1.00  2.35
ATOM    278  HA  TYR A  17      73.456  61.655  65.513  1.00  2.35
ATOM    279  CB  TYR A  17      72.600  62.804  67.113  1.00  2.35
ATOM    280  HB1 TYR A  17      72.741  62.006  67.842  1.00  2.35
ATOM    281  HB2 TYR A  17      71.640  63.278  67.329  1.00  2.35
ATOM    282  CG  TYR A  17      73.705  63.839  67.338  1.00  2.35
ATOM    283  CD1 TYR A  17      73.600  64.728  68.428  1.00  2.35
ATOM    284  HD1 TYR A  17      72.751  64.659  69.091  1.00  2.35
ATOM    285  CE1 TYR A  17      74.600  65.693  68.667  1.00  2.35
ATOM    286  HE1 TYR A  17      74.528  66.361  69.511  1.00  2.35
ATOM    287  CZ  TYR A  17      75.719  65.772  67.813  1.00  2.35
ATOM    288  OH  TYR A  17      76.696  66.694  68.032  1.00  2.35
ATOM    289  HH  TYR A  17      77.425  66.503  67.439  1.00  2.35
ATOM    290  CE2 TYR A  17      75.837  64.882  66.727  1.00  2.35
ATOM    291  HE2 TYR A  17      76.691  64.931  66.068  1.00  2.35
ATOM    292  CD2 TYR A  17      74.837  63.920  66.494  1.00  2.35
ATOM    293  HD2 TYR A  17      74.949  63.255  65.652  1.00  2.35
ATOM    294  C   TYR A  17     71.421  61.099  65.623  1.00  2.35
ATOM    295  O   TYR A  17     71.697  59.900  65.695  1.00  2.35
ATOM    296  N   ASN A  18     70.187  61.538  65.382  1.00  2.29
ATOM    297  H   ASN A  18     70.048  62.542  65.308  1.00  2.29
ATOM    298  CA  ASN A  18      69.026  60.670  65.338  1.00  2.29
ATOM    299  HA  ASN A  18      69.180  59.835  66.017  1.00  2.29
ATOM    300  CB  ASN A  18      67.806  61.472  65.819  1.00  2.29
ATOM    301  HB1 ASN A  18      66.915  60.863  65.668  1.00  2.29
ATOM    302  HB2 ASN A  18      67.695  62.382  65.227  1.00  2.29
ATOM    303  CG  ASN A  18      67.880  61.851  67.291  1.00  2.29
ATOM    304  OD1 ASN A  18      68.919  61.915  67.932  1.00  2.29
ATOM    305  ND2 ASN A  18      66.751  62.062  67.917  1.00  2.29
ATOM    306 1HD2 ASN A  18      65.882  62.039  67.419  1.00  2.29
ATOM    307 2HD2 ASN A  18      66.825  62.244  68.913  1.00  2.29
ATOM    308  C   ASN A  18     68.793  60.074  63.944  1.00  2.29
ATOM    309  O   ASN A  18     68.256  58.968  63.869  1.00  2.29
ATOM    310  N   TYR A  19     69.188  60.745  62.856  1.00  2.22
ATOM    311  H   TYR A  19     69.565  61.685  62.942  1.00  2.22
ATOM    312  CA  TYR A  19      69.103  60.172  61.515  1.00  2.22
ATOM    313  HA  TYR A  19      68.293  59.445  61.511  1.00  2.22
ATOM    314  CB  TYR A  19      68.788  61.200  60.414  1.00  2.22
ATOM    315  HB1 TYR A  19      69.034  60.775  59.441  1.00  2.22
```

```
ATOM    316  HB2 TYR A  19      69.454  62.051  60.521  1.00  2.22
ATOM    317  CG  TYR A  19      67.356  61.657  60.249  1.00  2.22
ATOM    318  CD1 TYR A  19      67.155  62.984  59.847  1.00  2.22
ATOM    319  HD1 TYR A  19      68.014  63.629  59.757  1.00  2.22
ATOM    320  CE1 TYR A  19      65.870  63.448  59.528  1.00  2.22
ATOM    321  HE1 TYR A  19      65.707  64.470  59.228  1.00  2.22
ATOM    322  CZ  TYR A  19      64.770  62.579  59.637  1.00  2.22
ATOM    323  OH  TYR A  19      63.526  63.068  59.429  1.00  2.22
ATOM    324  HH  TYR A  19      62.849  62.448  59.737  1.00  2.22
ATOM    325  CE2 TYR A  19      64.962  61.236  60.028  1.00  2.22
ATOM    326  HE2 TYR A  19      64.112  60.572  60.099  1.00  2.22
ATOM    327  CD2 TYR A  19      66.258  60.771  60.316  1.00  2.22
ATOM    328  HD2 TYR A  19      66.393  59.732  60.571  1.00  2.22
ATOM    329  C   TYR A  19     70.333  59.370  61.069  1.00  2.22
ATOM    330  O   TYR A  19     70.286  58.798  59.979  1.00  2.22
ATOM    331  N   LYS A  20     71.425  59.261  61.837  1.00  2.42
ATOM    332  H   LYS A  20     71.532  59.847  62.659  1.00  2.42
ATOM    333  CA  LYS A  20      72.520  58.384  61.404  1.00  2.42
ATOM    334  HA  LYS A  20      72.809  58.723  60.407  1.00  2.42
ATOM    335  CB  LYS A  20      73.774  58.531  62.293  1.00  2.42
ATOM    336  HB1 LYS A  20      73.612  58.030  63.249  1.00  2.42
ATOM    337  HB2 LYS A  20      73.976  59.586  62.483  1.00  2.42
ATOM    338  CG  LYS A  20      74.990  57.917  61.563  1.00  2.42
ATOM    339  HG1 LYS A  20      75.281  58.586  60.751  1.00  2.42
ATOM    340  HG2 LYS A  20      74.705  56.962  61.122  1.00  2.42
ATOM    341  CD  LYS A  20      76.211  57.651  62.455  1.00  2.42
ATOM    342  HD1 LYS A  20      75.918  56.973  63.257  1.00  2.42
ATOM    343  HD2 LYS A  20      76.576  58.585  62.886  1.00  2.42
ATOM    344  CE  LYS A  20      77.301  57.006  61.583  1.00  2.42
ATOM    345  HE1 LYS A  20      77.782  57.778  60.974  1.00  2.42
ATOM    346  HE2 LYS A  20      76.833  56.302  60.890  1.00  2.42
ATOM    347  NZ  LYS A  20      78.332  56.276  62.359  1.00  2.42
ATOM    348  HZ1 LYS A  20      78.971  55.828  61.690  1.00  2.42
ATOM    349  HZ2 LYS A  20      78.931  56.896  62.883  1.00  2.42
ATOM    350  HZ3 LYS A  20      77.951  55.544  62.941  1.00  2.42
ATOM    351  C   LYS A  20     72.101  56.917  61.297  1.00  2.42
ATOM    352  O   LYS A  20     71.912  56.228  62.301  1.00  2.42
ATOM    353  N   ASN A  21     72.032  56.438  60.059  1.00  2.36
ATOM    354  H   ASN A  21     72.167  57.109  59.318  1.00  2.36
ATOM    355  CA  ASN A  21      71.537  55.117  59.665  1.00  2.36
ATOM    356  HA  ASN A  21      70.579  54.977  60.147  1.00  2.36
ATOM    357  CB  ASN A  21      71.244  55.170  58.163  1.00  2.36
ATOM    358  HB1 ASN A  21      70.602  56.028  57.986  1.00  2.36
ATOM    359  HB2 ASN A  21      70.682  54.281  57.874  1.00  2.36
ATOM    360  CG  ASN A  21      72.455  55.279  57.254  1.00  2.36
ATOM    361  OD1 ASN A  21      73.607  55.094  57.625  1.00  2.36
ATOM    362  ND2 ASN A  21      72.213  55.612  56.013  1.00  2.36
ATOM    363 1HD2 ASN A  21      71.254  55.792  55.727  1.00  2.36
ATOM    364 2HD2 ASN A  21      72.960  55.558  55.337  1.00  2.36
ATOM    365  C   ASN A  21     72.409  53.903  60.041  1.00  2.36
ATOM    366  O   ASN A  21     72.215  52.811  59.507  1.00  2.36
ATOM    367  N   MET A  22      73.411  54.082  60.905  1.00  2.51
ATOM    368  H   MET A  22      73.483  54.973  61.369  1.00  2.51
```

```
ATOM    369  CA  MET A  22      74.385  53.036  61.199  1.00  2.51
ATOM    370  HA  MET A  22      74.841  52.768  60.245  1.00  2.51
ATOM    371  CB  MET A  22      75.517  53.561  62.096  1.00  2.51
ATOM    372  HB1 MET A  22      75.915  54.473  61.654  1.00  2.51
ATOM    373  HB2 MET A  22      76.318  52.820  62.099  1.00  2.51
ATOM    374  CG  MET A  22      75.124  53.840  63.551  1.00  2.51
ATOM    375  HG1 MET A  22      74.756  52.920  64.004  1.00  2.51
ATOM    376  HG2 MET A  22      74.328  54.584  63.581  1.00  2.51
ATOM    377  SD  MET A  22      76.524  54.422  64.541  1.00  2.51
ATOM    378  CE  MET A  22      75.843  54.235  66.212  1.00  2.51
ATOM    379  HE1 MET A  22      76.584  54.564  66.941  1.00  2.51
ATOM    380  HE2 MET A  22      74.942  54.838  66.319  1.00  2.51
ATOM    381  HE3 MET A  22      75.610  53.186  66.401  1.00  2.51
ATOM    382  C   MET A  22     73.729  51.772  61.770  1.00  2.51
ATOM    383  O   MET A  22     72.882  51.823  62.660  1.00  2.51
ATOM    384  N   ARG A  23     74.164  50.610  61.275  1.00  3.46
ATOM    385  H   ARG A  23     74.863  50.676  60.551  1.00  3.46
ATOM    386  CA  ARG A  23      73.583  49.280  61.547  1.00  3.46
ATOM    387  HA  ARG A  23      72.504  49.361  61.392  1.00  3.46
ATOM    388  CB  ARG A  23      74.151  48.294  60.500  1.00  3.46
ATOM    389  HB1 ARG A  23      74.025  48.719  59.503  1.00  3.46
ATOM    390  HB2 ARG A  23      73.552  47.383  60.528  1.00  3.46
ATOM    391  CG  ARG A  23      75.646  47.956  60.724  1.00  3.46
ATOM    392  HG1 ARG A  23      76.002  48.433  61.638  1.00  3.46
ATOM    393  HG2 ARG A  23      76.239  48.360  59.902  1.00  3.46
ATOM    394  CD  ARG A  23      75.897  46.442  60.826  1.00  3.46
ATOM    395  HD1 ARG A  23      75.991  46.030  59.819  1.00  3.46
ATOM    396  HD2 ARG A  23      75.041  45.956  61.299  1.00  3.46
ATOM    397  NE  ARG A  23      77.116  46.137  61.604  1.00  3.46
ATOM    398  HE  ARG A  23      77.980  46.036  61.099  1.00  3.46
ATOM    399  CZ  ARG A  23      77.177  45.914  62.908  1.00  3.46
ATOM    400  NH1 ARG A  23      78.318  45.669  63.485  1.00  3.46
ATOM    401 1HH1 ARG A  23      79.173  45.653  62.959  1.00  3.46
ATOM    402 2HH1 ARG A  23      78.343  45.509  64.478  1.00  3.46
ATOM    403  NH2 ARG A  23      76.127  45.939  63.676  1.00  3.46
ATOM    404 1HH2 ARG A  23      75.224  46.213  63.307  1.00  3.46
ATOM    405 2HH2 ARG A  23      76.210  45.743  64.659  1.00  3.46
ATOM    406  C   ARG A  23     73.758  48.725  62.972  1.00  3.46
ATOM    407  O   ARG A  23     73.746  47.511  63.186  1.00  3.46
ATOM    408  N   TRP A  24     74.042  49.600  63.925  1.00  3.49
ATOM    409  H   TRP A  24     74.017  50.573  63.648  1.00  3.49
ATOM    410  CA  TRP A  24      74.395  49.297  65.315  1.00  3.49
ATOM    411  HA  TRP A  24      73.867  48.397  65.627  1.00  3.49
ATOM    412  CB  TRP A  24      75.912  49.059  65.386  1.00  3.49
ATOM    413  HB1 TRP A  24      76.421  50.025  65.412  1.00  3.49
ATOM    414  HB2 TRP A  24      76.237  48.565  64.471  1.00  3.49
ATOM    415  CG  TRP A  24      76.405  48.210  66.520  1.00  3.49
ATOM    416  CD1 TRP A  24      77.419  48.543  67.350  1.00  3.49
ATOM    417  HD1 TRP A  24      77.965  49.481  67.321  1.00  3.49
ATOM    418  NE1 TRP A  24      77.671  47.503  68.220  1.00  3.49
ATOM    419  HE1 TRP A  24      78.385  47.542  68.938  1.00  3.49
ATOM    420  CE2 TRP A  24      76.834  46.434  67.991  1.00  3.49
ATOM    421  CZ2 TRP A  24      76.731  45.161  68.567  1.00  3.49
```

```
ATOM    422  HZ2 TRP A  24      77.405  44.860  69.357  1.00  3.49
ATOM    423  CH2 TRP A  24      75.731  44.284  68.114  1.00  3.49
ATOM    424  HH2 TRP A  24      75.630  43.301  68.558  1.00  3.49
ATOM    425  CZ3 TRP A  24      74.849  44.692  67.098  1.00  3.49
ATOM    426  HZ3 TRP A  24      74.066  44.018  66.774  1.00  3.49
ATOM    427  CE3 TRP A  24      74.969  45.969  66.517  1.00  3.49
ATOM    428  HE3 TRP A  24      74.267  46.263  65.749  1.00  3.49
ATOM    429  CD2 TRP A  24      75.975  46.872  66.936  1.00  3.49
ATOM    430  C   TRP A  24     73.936  50.427  66.244  1.00  3.49
ATOM    431  O   TRP A  24     74.536  50.713  67.274  1.00  3.49
ATOM    432  N   ALA A  25     72.866  51.110  65.831  1.00  3.55
ATOM    433  H   ALA A  25     72.447  50.813  64.962  1.00  3.55
ATOM    434  CA  ALA A  25      72.273  52.303  66.430  1.00  3.55
ATOM    435  HA  ALA A  25      73.061  53.051  66.526  1.00  3.55
ATOM    436  CB  ALA A  25      71.243  52.830  65.419  1.00  3.55
ATOM    437  HB1 ALA A  25      70.715  53.680  65.841  1.00  3.55
ATOM    438  HB2 ALA A  25      71.745  53.152  64.508  1.00  3.55
ATOM    439  HB3 ALA A  25      70.516  52.053  65.179  1.00  3.55
ATOM    440  C   ALA A  25     71.651  52.160  67.838  1.00  3.55
ATOM    441  O   ALA A  25     70.737  52.916  68.194  1.00  3.55
ATOM    442  N   LYS A  26     72.120  51.199  68.635  1.00  3.82
ATOM    443  H   LYS A  26     72.962  50.725  68.332  1.00  3.82
ATOM    444  CA  LYS A  26      71.705  51.016  70.023  1.00  3.82
ATOM    445  HA  LYS A  26      70.651  50.746  70.014  1.00  3.82
ATOM    446  CB  LYS A  26      72.547  49.883  70.648  1.00  3.82
ATOM    447  HB1 LYS A  26      73.548  50.268  70.855  1.00  3.82
ATOM    448  HB2 LYS A  26      72.660  49.073  69.926  1.00  3.82
ATOM    449  CG  LYS A  26      71.961  49.313  71.953  1.00  3.82
ATOM    450  HG1 LYS A  26      71.758  50.132  72.643  1.00  3.82
ATOM    451  HG2 LYS A  26      72.708  48.668  72.419  1.00  3.82
ATOM    452  CD  LYS A  26      70.671  48.499  71.752  1.00  3.82
ATOM    453  HD1 LYS A  26      69.945  49.084  71.190  1.00  3.82
ATOM    454  HD2 LYS A  26      70.236  48.284  72.730  1.00  3.82
ATOM    455  CE  LYS A  26      70.934  47.164  71.042  1.00  3.82
ATOM    456  HE1 LYS A  26      71.381  46.475  71.766  1.00  3.82
ATOM    457  HE2 LYS A  26      71.641  47.299  70.221  1.00  3.82
ATOM    458  NZ  LYS A  26      69.677  46.587  70.529  1.00  3.82
ATOM    459  HZ1 LYS A  26      69.719  45.570  70.436  1.00  3.82
ATOM    460  HZ2 LYS A  26      68.924  46.651  71.219  1.00  3.82
ATOM    461  HZ3 LYS A  26      69.337  46.988  69.674  1.00  3.82
ATOM    462  C   LYS A  26     71.857  52.331  70.795  1.00  3.82
ATOM    463  O   LYS A  26     72.847  53.042  70.649  1.00  3.82
ATOM    464  N   GLY A  27     70.860  52.660  71.610  1.00  3.51
ATOM    465  H   GLY A  27     70.078  52.031  71.705  1.00  3.51
ATOM    466  CA  GLY A  27      70.932  53.824  72.490  1.00  3.51
ATOM    467  HA1 GLY A  27      71.964  53.978  72.812  1.00  3.51
ATOM    468  HA2 GLY A  27      70.345  53.610  73.384  1.00  3.51
ATOM    469  C   GLY A  27     70.428  55.141  71.928  1.00  3.51
ATOM    470  O   GLY A  27     70.558  56.165  72.598  1.00  3.51
ATOM    471  N   ARG A  28     69.848  55.162  70.718  1.00  2.53
ATOM    472  H   ARG A  28     69.858  54.323  70.151  1.00  2.53
ATOM    473  CA  ARG A  28      69.236  56.405  70.227  1.00  2.53
ATOM    474  HA  ARG A  28      69.937  57.175  70.515  1.00  2.53
```

```
ATOM    475  CB  ARG A  28      69.140  56.448  68.699  1.00  2.53
ATOM    476  HB1 ARG A  28      68.228  55.940  68.383  1.00  2.53
ATOM    477  HB2 ARG A  28      69.996  55.918  68.276  1.00  2.53
ATOM    478  CG  ARG A  28      69.156  57.887  68.140  1.00  2.53
ATOM    479  HG1 ARG A  28      68.222  58.388  68.401  1.00  2.53
ATOM    480  HG2 ARG A  28      69.189  57.818  67.054  1.00  2.53
ATOM    481  CD  ARG A  28      70.358  58.749  68.592  1.00  2.53
ATOM    482  HD1 ARG A  28      70.764  59.246  67.719  1.00  2.53
ATOM    483  HD2 ARG A  28      71.153  58.121  68.997  1.00  2.53
ATOM    484  NE  ARG A  28      69.951  59.803  69.539  1.00  2.53
ATOM    485  HE  ARG A  28      69.397  60.557  69.144  1.00  2.53
ATOM    486  CZ  ARG A  28      70.182  59.893  70.835  1.00  2.53
ATOM    487  NH1 ARG A  28      69.483  60.722  71.544  1.00  2.53
ATOM    488 1HH1 ARG A  28      68.679  61.160  71.104  1.00  2.53
ATOM    489 2HH1 ARG A  28      69.555  60.698  72.540  1.00  2.53
ATOM    490  NH2 ARG A  28      71.069  59.174  71.466  1.00  2.53
ATOM    491 1HH2 ARG A  28      71.640  58.532  70.950  1.00  2.53
ATOM    492 2HH2 ARG A  28      71.111  59.178  72.467  1.00  2.53
ATOM    493  C   ARG A  28     67.964  56.804  70.969  1.00  2.53
ATOM    494  O   ARG A  28     67.780  58.000  71.167  1.00  2.53
ATOM    495  N   ASN A  29     67.202  55.809  71.445  1.00  2.51
ATOM    496  H   ASN A  29     67.597  54.892  71.311  1.00  2.51
ATOM    497  CA  ASN A  29      65.986  55.798  72.290  1.00  2.51
ATOM    498  HA  ASN A  29      65.545  54.804  72.186  1.00  2.51
ATOM    499  CB  ASN A  29      66.409  55.930  73.763  1.00  2.51
ATOM    500  HB1 ASN A  29      65.509  56.022  74.370  1.00  2.51
ATOM    501  HB2 ASN A  29      67.004  56.833  73.900  1.00  2.51
ATOM    502  CG  ASN A  29      67.196  54.738  74.288  1.00  2.51
ATOM    503  OD1 ASN A  29      67.656  53.863  73.568  1.00  2.51
ATOM    504  ND2 ASN A  29      67.379  54.663  75.585  1.00  2.51
ATOM    505 1HD2 ASN A  29      67.029  55.374  76.199  1.00  2.51
ATOM    506 2HD2 ASN A  29      67.866  53.852  75.928  1.00  2.51
ATOM    507  C   ASN A  29     64.825  56.754  71.922  1.00  2.51
ATOM    508  O   ASN A  29     63.655  56.447  72.139  1.00  2.51
ATOM    509  N   GLU A  30     65.131  57.902  71.344  1.00  2.23
ATOM    510  H   GLU A  30     66.112  58.120  71.262  1.00  2.23
ATOM    511  CA  GLU A  30      64.227  58.738  70.574  1.00  2.23
ATOM    512  HA  GLU A  30      63.311  58.880  71.138  1.00  2.23
ATOM    513  CB  GLU A  30      64.902  60.111  70.341  1.00  2.23
ATOM    514  HB1 GLU A  30      64.211  60.759  69.800  1.00  2.23
ATOM    515  HB2 GLU A  30      65.783  59.960  69.715  1.00  2.23
ATOM    516  CG  GLU A  30      65.335  60.839  71.628  1.00  2.23
ATOM    517  HG1 GLU A  30      66.019  60.208  72.197  1.00  2.23
ATOM    518  HG2 GLU A  30      64.456  61.016  72.245  1.00  2.23
ATOM    519  CD  GLU A  30      66.054  62.165  71.333  1.00  2.23
ATOM    520  OE1 GLU A  30      67.182  62.149  70.788  1.00  2.23
ATOM    521  OE2 GLU A  30      65.556  63.260  71.683  1.00  2.23
ATOM    522  C   GLU A  30     63.883  58.105  69.218  1.00  2.23
ATOM    523  O   GLU A  30     64.569  57.214  68.722  1.00  2.23
ATOM    524  N   THR A  31     62.868  58.664  68.570  1.00  1.83
ATOM    525  H   THR A  31     62.369  59.411  69.029  1.00  1.83
ATOM    526  CA  THR A  31      62.503  58.441  67.166  1.00  1.83
ATOM    527  HA  THR A  31      63.362  58.031  66.649  1.00  1.83
```

```
ATOM    528  CB  THR A  31      61.357  57.418  67.044  1.00  1.83
ATOM    529  HB  THR A  31      61.749  56.429  67.284  1.00  1.83
ATOM    530  CG2 THR A  31      60.207  57.708  68.004  1.00  1.83
ATOM    531 1HG2 THR A  31      59.383  57.022  67.819  1.00  1.83
ATOM    532 2HG2 THR A  31      60.536  57.554  69.029  1.00  1.83
ATOM    533 3HG2 THR A  31      59.875  58.733  67.882  1.00  1.83
ATOM    534  OG1 THR A  31      60.875  57.391  65.719  1.00  1.83
ATOM    535  HG1 THR A  31      59.965  57.015  65.742  1.00  1.83
ATOM    536  C   THR A  31     62.157  59.786  66.566  1.00  1.83
ATOM    537  O   THR A  31     61.800  60.686  67.337  1.00  1.83
ATOM    538  N   TYR A  32     62.342  59.977  65.248  1.00  1.68
ATOM    539  H   TYR A  32     62.495  59.200  64.620  1.00  1.68
ATOM    540  CA  TYR A  32      62.298  61.339  64.728  1.00  1.68
ATOM    541  HA  TYR A  32      61.787  61.985  65.443  1.00  1.68
ATOM    542  CB  TYR A  32      63.707  61.907  64.555  1.00  1.68
ATOM    543  HB1 TYR A  32      64.261  61.297  63.842  1.00  1.68
ATOM    544  HB2 TYR A  32      64.229  61.893  65.513  1.00  1.68
ATOM    545  CG  TYR A  32      63.636  63.324  64.050  1.00  1.68
ATOM    546  CD1 TYR A  32      62.989  64.314  64.809  1.00  1.68
ATOM    547  HD1 TYR A  32      62.632  64.101  65.807  1.00  1.68
ATOM    548  CE1 TYR A  32      62.748  65.575  64.239  1.00  1.68
ATOM    549  HE1 TYR A  32      62.255  66.338  64.814  1.00  1.68
ATOM    550  CZ  TYR A  32     63.150  65.849  62.913  1.00  1.68
ATOM    551  OH  TYR A  32      62.929  67.061  62.345  1.00  1.68
ATOM    552  HH  TYR A  32      63.143  67.055  61.410  1.00  1.68
ATOM    553  CE2 TYR A  32      63.830  64.861  62.177  1.00  1.68
ATOM    554  HE2 TYR A  32      64.128  65.054  61.161  1.00  1.68
ATOM    555  CD2 TYR A  32      64.076  63.605  62.750  1.00  1.68
ATOM    556  HD2 TYR A  32      64.555  62.834  62.171  1.00  1.68
ATOM    557  C   TYR A  32     61.460  61.391  63.471  1.00  1.68
ATOM    558  O   TYR A  32     61.828  60.820  62.448  1.00  1.68
ATOM    559  N   LEU A  33     60.264  61.952  63.606  1.00  1.67
ATOM    560  H   LEU A  33     60.031  62.453  64.458  1.00  1.67
ATOM    561  CA  LEU A  33      59.208  61.724  62.640  1.00  1.67
ATOM    562  HA  LEU A  33      59.495  60.931  61.954  1.00  1.67
ATOM    563  CB  LEU A  33      57.924  61.325  63.411  1.00  1.67
ATOM    564  HB1 LEU A  33      57.087  61.926  63.054  1.00  1.67
ATOM    565  HB2 LEU A  33      58.029  61.559  64.472  1.00  1.67
ATOM    566  CG  LEU A  33      57.514  59.853  63.275  1.00  1.67
ATOM    567  HG  LEU A  33      56.569  59.715  63.801  1.00  1.67
ATOM    568  CD1 LEU A  33      57.298  59.492  61.811  1.00  1.67
ATOM    569 1HD1 LEU A  33      56.714  58.575  61.760  1.00  1.67
ATOM    570 2HD1 LEU A  33      56.766  60.302  61.326  1.00  1.67
ATOM    571 3HD1 LEU A  33      58.246  59.349  61.299  1.00  1.67
ATOM    572  CD2 LEU A  33      58.543  58.905  63.872  1.00  1.67
ATOM    573 1HD2 LEU A  33      58.262  57.874  63.652  1.00  1.67
ATOM    574 2HD2 LEU A  33      59.534  59.093  63.480  1.00  1.67
ATOM    575 3HD2 LEU A  33      58.573  59.036  64.948  1.00  1.67
ATOM    576  C   LEU A  33     58.973  62.976  61.817  1.00  1.67
ATOM    577  O   LEU A  33     58.660  64.025  62.403  1.00  1.67
ATOM    578  N   CYS A  34     59.046  62.837  60.481  1.00  1.81
ATOM    579  H   CYS A  34     59.402  61.969  60.103  1.00  1.81
ATOM    580  CA  CYS A  34      59.052  64.038  59.636  1.00  1.81
```

```
ATOM    581  HA  CYS A  34      58.830  64.905  60.260  1.00  1.81
ATOM    582  CB  CYS A  34      60.455  64.281  59.089  1.00  1.81
ATOM    583  HB1 CYS A  34      60.415  64.554  58.036  1.00  1.81
ATOM    584  HB2 CYS A  34      61.055  63.390  59.203  1.00  1.81
ATOM    585  SG  CYS A  34      61.231  65.608  60.037  1.00  1.81
ATOM    586  HG  CYS A  34      60.632  66.617  59.391  1.00  1.81
ATOM    587  C   CYS A  34      57.963  64.068  58.563  1.00  1.81
ATOM    588  O   CYS A  34      58.023  63.409  57.517  1.00  1.81
ATOM    589  N   PHE A  35      56.915  64.811  58.888  1.00  2.04
ATOM    590  H   PHE A  35      56.986  65.462  59.665  1.00  2.04
ATOM    591  CA  PHE A  35      55.591  64.533  58.373  1.00  2.04
ATOM    592  HA  PHE A  35      55.569  63.527  57.977  1.00  2.04
ATOM    593  CB  PHE A  35      54.588  64.594  59.546  1.00  2.04
ATOM    594  HB1 PHE A  35      53.751  65.241  59.275  1.00  2.04
ATOM    595  HB2 PHE A  35      55.061  65.062  60.409  1.00  2.04
ATOM    596  CG  PHE A  35      54.024  63.250  59.966  1.00  2.04
ATOM    597  CD1 PHE A  35      54.729  62.425  60.862  1.00  2.04
ATOM    598  HD1 PHE A  35      55.673  62.758  61.265  1.00  2.04
ATOM    599  CE1 PHE A  35      54.188  61.182  61.251  1.00  2.04
ATOM    600  HE1 PHE A  35      54.703  60.559  61.967  1.00  2.04
ATOM    601  CZ  PHE A  35      52.960  60.749  60.726  1.00  2.04
ATOM    602  HZ  PHE A  35      52.556  59.788  61.017  1.00  2.04
ATOM    603  CE2 PHE A  35      52.237  61.589  59.865  1.00  2.04
ATOM    604  HE2 PHE A  35      51.271  61.274  59.503  1.00  2.04
ATOM    605  CD2 PHE A  35      52.768  62.836  59.487  1.00  2.04
ATOM    606  HD2 PHE A  35      52.205  63.488  58.836  1.00  2.04
ATOM    607  C   PHE A  35      55.172  65.516  57.286  1.00  2.04
ATOM    608  O   PHE A  35      55.364  66.740  57.448  1.00  2.04
ATOM    609  N   VAL A  36      54.587  64.948  56.207  1.00  2.06
ATOM    610  H   VAL A  36      54.482  63.936  56.201  1.00  2.06
ATOM    611  CA  VAL A  36      54.331  65.694  54.972  1.00  2.06
ATOM    612  HA  VAL A  36      54.380  66.755  55.227  1.00  2.06
ATOM    613  CB  VAL A  36      55.392  65.500  53.874  1.00  2.06
ATOM    614  HB  VAL A  36      55.292  64.508  53.446  1.00  2.06
ATOM    615  CG1 VAL A  36      55.235  66.529  52.748  1.00  2.06
ATOM    616 1HG1 VAL A  36      56.036  66.417  52.017  1.00  2.06
ATOM    617 2HG1 VAL A  36      54.287  66.384  52.239  1.00  2.06
ATOM    618 3HG1 VAL A  36      55.263  67.533  53.165  1.00  2.06
ATOM    619  CG2 VAL A  36      56.810  65.663  54.407  1.00  2.06
ATOM    620 1HG2 VAL A  36      57.525  65.572  53.590  1.00  2.06
ATOM    621 2HG2 VAL A  36      56.911  66.641  54.881  1.00  2.06
ATOM    622 3HG2 VAL A  36      57.023  64.882  55.136  1.00  2.06
ATOM    623  C   VAL A  36      52.933  65.490  54.405  1.00  2.06
ATOM    624  O   VAL A  36      52.709  64.662  53.507  1.00  2.06
ATOM    625  N   VAL A  37      52.012  66.277  54.963  1.00  2.12
ATOM    626  H   VAL A  37      52.316  66.912  55.695  1.00  2.12
ATOM    627  CA  VAL A  37      50.595  66.352  54.598  1.00  2.12
ATOM    628  HA  VAL A  37      50.252  65.336  54.411  1.00  2.12
ATOM    629  CB  VAL A  37      49.770  66.900  55.771  1.00  2.12
ATOM    630  HB  VAL A  37      50.103  67.916  55.950  1.00  2.12
ATOM    631  CG1 VAL A  37      48.266  66.891  55.456  1.00  2.12
ATOM    632 1HG1 VAL A  37      47.705  67.267  56.309  1.00  2.12
ATOM    633 2HG1 VAL A  37      48.044  67.539  54.611  1.00  2.12
```

```
ATOM    634 3HG1 VAL A  37      47.937  65.877  55.229  1.00  2.12
ATOM    635  CG2 VAL A  37      49.977  66.093  57.061  1.00  2.12
ATOM    636 1HG2 VAL A  37      51.033  66.044  57.321  1.00  2.12
ATOM    637 2HG2 VAL A  37      49.490  66.612  57.882  1.00  2.12
ATOM    638 3HG2 VAL A  37      49.564  65.092  56.962  1.00  2.12
ATOM    639  C   VAL A  37     50.345  67.155  53.321  1.00  2.12
ATOM    640  O   VAL A  37     50.816  68.287  53.166  1.00  2.12
ATOM    641  N   LYS A  38     49.586  66.564  52.385  1.00  2.41
ATOM    642  H   LYS A  38     49.304  65.598  52.545  1.00  2.41
ATOM    643  CA  LYS A  38     49.461  67.083  51.020  1.00  2.41
ATOM    644  HA  LYS A  38     49.635  68.158  51.066  1.00  2.41
ATOM    645  CB  LYS A  38     50.545  66.513  50.085  1.00  2.41
ATOM    646  HB1 LYS A  38     50.315  66.842  49.072  1.00  2.41
ATOM    647  HB2 LYS A  38     50.521  65.423  50.094  1.00  2.41
ATOM    648  CG  LYS A  38     51.950  67.014  50.458  1.00  2.41
ATOM    649  HG1 LYS A  38     52.212  66.615  51.436  1.00  2.41
ATOM    650  HG2 LYS A  38     51.935  68.102  50.530  1.00  2.41
ATOM    651  CD  LYS A  38     53.054  66.599  49.474  1.00  2.41
ATOM    652  HD1 LYS A  38     53.160  65.514  49.512  1.00  2.41
ATOM    653  HD2 LYS A  38     53.997  67.036  49.807  1.00  2.41
ATOM    654  CE  LYS A  38     52.798  67.016  48.019  1.00  2.41
ATOM    655  HE1 LYS A  38     51.926  66.468  47.649  1.00  2.41
ATOM    656  HE2 LYS A  38     53.653  66.712  47.409  1.00  2.41
ATOM    657  NZ  LYS A  38     52.568  68.474  47.868  1.00  2.41
ATOM    658  HZ1 LYS A  38     53.364  69.030  48.143  1.00  2.41
ATOM    659  HZ2 LYS A  38     52.310  68.667  46.901  1.00  2.41
ATOM    660  HZ3 LYS A  38     51.751  68.775  48.395  1.00  2.41
ATOM    661  C   LYS A  38     48.095  66.944  50.361  1.00  2.41
ATOM    662  O   LYS A  38     47.688  65.845  49.977  1.00  2.41
ATOM    663  N   ARG A  39     47.399  68.070  50.127  1.00  2.87
ATOM    664  H   ARG A  39     47.824  68.972  50.316  1.00  2.87
ATOM    665  CA  ARG A  39     46.066  67.995  49.504  1.00  2.87
ATOM    666  HA  ARG A  39     45.590  67.102  49.917  1.00  2.87
ATOM    667  CB  ARG A  39     45.075  69.123  49.904  1.00  2.87
ATOM    668  HB1 ARG A  39     44.583  68.836  50.831  1.00  2.87
ATOM    669  HB2 ARG A  39     44.282  69.142  49.155  1.00  2.87
ATOM    670  CG  ARG A  39     45.600  70.557  50.053  1.00  2.87
ATOM    671  HG1 ARG A  39     44.827  71.255  49.728  1.00  2.87
ATOM    672  HG2 ARG A  39     46.452  70.671  49.392  1.00  2.87
ATOM    673  CD  ARG A  39     45.958  70.885  51.514  1.00  2.87
ATOM    674  HD1 ARG A  39     46.560  70.064  51.903  1.00  2.87
ATOM    675  HD2 ARG A  39     45.045  70.945  52.112  1.00  2.87
ATOM    676  NE  ARG A  39     46.764  72.112  51.660  1.00  2.87
ATOM    677  HE  ARG A  39     47.747  71.996  51.434  1.00  2.87
ATOM    678  CZ  ARG A  39     46.419  73.295  52.127  1.00  2.87
ATOM    679  NH1 ARG A  39     47.329  74.197  52.335  1.00  2.87
ATOM    680 1HH1 ARG A  39     48.308  73.963  52.185  1.00  2.87
ATOM    681 2HH1 ARG A  39     47.107  75.079  52.755  1.00  2.87
ATOM    682  NH2 ARG A  39     45.191  73.639  52.376  1.00  2.87
ATOM    683 1HH2 ARG A  39     44.450  73.008  52.127  1.00  2.87
ATOM    684 2HH2 ARG A  39     44.946  74.601  52.529  1.00  2.87
ATOM    685  C   ARG A  39     46.072  67.758  47.990  1.00  2.87
ATOM    686  O   ARG A  39     46.216  68.681  47.195  1.00  2.87
```

```
ATOM    687  N   ARG A  40      45.797  66.504  47.625  1.00  3.49
ATOM    688  H   ARG A  40      45.964  65.822  48.356  1.00  3.49
ATOM    689  CA  ARG A  40      44.780  66.089  46.637  1.00  3.49
ATOM    690  HA  ARG A  40      44.666  65.017  46.792  1.00  3.49
ATOM    691  CB  ARG A  40      43.428  66.763  46.966  1.00  3.49
ATOM    692  HB1 ARG A  40      43.351  67.722  46.452  1.00  3.49
ATOM    693  HB2 ARG A  40      43.387  66.968  48.038  1.00  3.49
ATOM    694  CG  ARG A  40      42.215  65.877  46.643  1.00  3.49
ATOM    695  HG1 ARG A  40      41.636  65.807  47.556  1.00  3.49
ATOM    696  HG2 ARG A  40      42.534  64.868  46.381  1.00  3.49
ATOM    697  CD  ARG A  40      41.270  66.406  45.559  1.00  3.49
ATOM    698  HD1 ARG A  40      41.768  66.376  44.587  1.00  3.49
ATOM    699  HD2 ARG A  40      41.026  67.448  45.775  1.00  3.49
ATOM    700  NE  ARG A  40      40.023  65.607  45.552  1.00  3.49
ATOM    701  HE  ARG A  40      39.249  65.953  46.096  1.00  3.49
ATOM    702  CZ  ARG A  40      39.828  64.450  44.942  1.00  3.49
ATOM    703  NH1 ARG A  40      38.756  63.748  45.171  1.00  3.49
ATOM    704 1HH1 ARG A  40      38.120  64.015  45.917  1.00  3.49
ATOM    705 2HH1 ARG A  40      38.610  62.865  44.715  1.00  3.49
ATOM    706  NH2 ARG A  40      40.699  63.966  44.100  1.00  3.49
ATOM    707 1HH2 ARG A  40      41.527  64.506  43.906  1.00  3.49
ATOM    708 2HH2 ARG A  40      40.550  63.085  43.644  1.00  3.49
ATOM    709  C   ARG A  40      45.118  66.208  45.173  1.00  3.49
ATOM    710  O   ARG A  40      44.334  65.722  44.355  1.00  3.49
ATOM    711  N   LEU A  41      46.259  66.813  44.879  1.00  5.30
ATOM    712  H   LEU A  41      46.800  67.212  45.630  1.00  5.30
ATOM    713  CA  LEU A  41      46.780  66.850  43.535  1.00  5.30
ATOM    714  HA  LEU A  41      46.052  66.358  42.890  1.00  5.30
ATOM    715  CB  LEU A  41      46.849  68.287  42.954  1.00  5.30
ATOM    716  HB1 LEU A  41      47.211  68.211  41.929  1.00  5.30
ATOM    717  HB2 LEU A  41      47.592  68.851  43.489  1.00  5.30
ATOM    718  CG  LEU A  41      45.577  69.164  42.879  1.00  5.30
ATOM    719  HG  LEU A  41      45.834  70.043  42.292  1.00  5.30
ATOM    720  CD1 LEU A  41      44.417  68.475  42.160  1.00  5.30
ATOM    721 1HD1 LEU A  41      43.608  69.191  42.017  1.00  5.30
ATOM    722 2HD1 LEU A  41      44.752  68.135  41.180  1.00  5.30
ATOM    723 3HD1 LEU A  41      44.050  67.630  42.735  1.00  5.30
ATOM    724  CD2 LEU A  41      45.076  69.678  44.229  1.00  5.30
ATOM    725 1HD2 LEU A  41      45.908  70.116  44.782  1.00  5.30
ATOM    726 2HD2 LEU A  41      44.337  70.460  44.058  1.00  5.30
ATOM    727 3HD2 LEU A  41      44.623  68.885  44.812  1.00  5.30
ATOM    728  C   LEU A  41      48.014  65.960  43.343  1.00  5.30
ATOM    729  O   LEU A  41      48.497  65.345  44.295  1.00  5.30
ATOM    730  N   GLY A  42      48.477  65.862  42.099  1.00  7.28
ATOM    731  H   GLY A  42      48.045  66.449  41.403  1.00  7.28
ATOM    732  CA  GLY A  42      49.707  65.173  41.711  1.00  7.28
ATOM    733  HA1 GLY A  42      49.491  64.482  40.896  1.00  7.28
ATOM    734  HA2 GLY A  42      50.115  64.607  42.549  1.00  7.28
ATOM    735  C   GLY A  42      50.767  66.173  41.234  1.00  7.28
ATOM    736  O   GLY A  42      51.841  66.226  41.837  1.00  7.28
ATOM    737  N   PRO A  43      50.418  67.080  40.297  1.00  6.70
ATOM    738  CD  PRO A  43      49.381  66.944  39.274  1.00  6.70
ATOM    739  HD1 PRO A  43      48.459  66.512  39.656  1.00  6.70
```

```
ATOM    740  HD2 PRO A  43      49.765  66.327  38.461  1.00  6.70
ATOM    741  CG  PRO A  43      49.109  68.357  38.753  1.00  6.70
ATOM    742  HG1 PRO A  43      48.329  68.830  39.352  1.00  6.70
ATOM    743  HG2 PRO A  43      48.837  68.357  37.697  1.00  6.70
ATOM    744  CB  PRO A  43      50.442  69.058  38.997  1.00  6.70
ATOM    745  HB1 PRO A  43      50.345  70.140  39.013  1.00  6.70
ATOM    746  HB2 PRO A  43      51.148  68.767  38.218  1.00  6.70
ATOM    747  CA  PRO A  43      50.893  68.463  40.336  1.00  6.70
ATOM    748  HA  PRO A  43      51.982  68.491  40.385  1.00  6.70
ATOM    749  C   PRO A  43      50.333  69.199  41.576  1.00  6.70
ATOM    750  O   PRO A  43      49.805  68.543  42.464  1.00  6.70
ATOM    751  N   ASP A  44      50.448  70.529  41.664  1.00  5.72
ATOM    752  H   ASP A  44      50.904  70.979  40.876  1.00  5.72
ATOM    753  CA  ASP A  44      49.563  71.482  42.391  1.00  5.72
ATOM    754  HA  ASP A  44      50.134  72.408  42.482  1.00  5.72
ATOM    755  CB  ASP A  44      48.375  71.808  41.462  1.00  5.72
ATOM    756  HB1 ASP A  44      47.826  72.665  41.856  1.00  5.72
ATOM    757  HB2 ASP A  44      47.706  70.948  41.445  1.00  5.72
ATOM    758  CG  ASP A  44      48.809  72.121  40.026  1.00  5.72
ATOM    759  OD1 ASP A  44      48.110  71.655  39.100  1.00  5.72
ATOM    760  OD2 ASP A  44      49.898  72.720  39.874  1.00  5.72
ATOM    761  C   ASP A  44      49.073  71.167  43.830  1.00  5.72
ATOM    762  O   ASP A  44      48.087  71.712  44.328  1.00  5.72
ATOM    763  N   SER A  45      49.747  70.254  44.515  1.00  5.78
ATOM    764  H   SER A  45      50.510  69.835  44.005  1.00  5.78
ATOM    765  CA  SER A  45      49.173  69.424  45.576  1.00  5.78
ATOM    766  HA  SER A  45      48.108  69.288  45.405  1.00  5.78
ATOM    767  CB  SER A  45      49.839  68.042  45.543  1.00  5.78
ATOM    768  HB1 SER A  45      49.437  67.507  44.695  1.00  5.78
ATOM    769  HB2 SER A  45      49.599  67.477  46.445  1.00  5.78
ATOM    770  OG  SER A  45      51.246  68.119  45.357  1.00  5.78
ATOM    771  HG  SER A  45      51.382  67.853  44.433  1.00  5.78
ATOM    772  C   SER A  45      49.276  70.055  46.949  1.00  5.78
ATOM    773  O   SER A  45      50.062  69.592  47.785  1.00  5.78
ATOM    774  N   LEU A  46      48.533  71.153  47.129  1.00  4.88
ATOM    775  H   LEU A  46      47.964  71.430  46.332  1.00  4.88
ATOM    776  CA  LEU A  46      48.802  72.171  48.139  1.00  4.88
ATOM    777  HA  LEU A  46      49.614  72.730  47.676  1.00  4.88
ATOM    778  CB  LEU A  46      47.649  73.188  48.259  1.00  4.88
ATOM    779  HB1 LEU A  46      46.841  72.755  48.840  1.00  4.88
ATOM    780  HB2 LEU A  46      47.260  73.391  47.259  1.00  4.88
ATOM    781  CG  LEU A  46      48.052  74.537  48.894  1.00  4.88
ATOM    782  HG  LEU A  46      48.442  74.353  49.892  1.00  4.88
ATOM    783  CD1 LEU A  46      49.112  75.305  48.098  1.00  4.88
ATOM    784 1HD1 LEU A  46      49.267  76.290  48.540  1.00  4.88
ATOM    785 2HD1 LEU A  46      50.065  74.781  48.114  1.00  4.88
ATOM    786 3HD1 LEU A  46      48.784  75.430  47.065  1.00  4.88
ATOM    787  CD2 LEU A  46      46.823  75.439  49.003  1.00  4.88
ATOM    788 1HD2 LEU A  46      47.089  76.375  49.498  1.00  4.88
ATOM    789 2HD2 LEU A  46      46.435  75.665  48.008  1.00  4.88
ATOM    790 3HD2 LEU A  46      46.050  74.942  49.585  1.00  4.88
ATOM    791  C   LEU A  46      49.386  71.607  49.433  1.00  4.88
ATOM    792  O   LEU A  46      48.756  70.829  50.162  1.00  4.88
```

```
ATOM    793  N   SER A  47      50.657  71.949  49.654  1.00  4.21
ATOM    794  H   SER A  47      51.155  72.498  48.972  1.00  4.21
ATOM    795  CA  SER A  47      51.356  71.616  50.883  1.00  4.21
ATOM    796  HA  SER A  47      51.478  70.535  50.944  1.00  4.21
ATOM    797  CB  SER A  47      52.735  72.266  50.956  1.00  4.21
ATOM    798  HB1 SER A  47      53.218  71.988  51.895  1.00  4.21
ATOM    799  HB2 SER A  47      52.633  73.353  50.916  1.00  4.21
ATOM    800  OG  SER A  47      53.517  71.817  49.869  1.00  4.21
ATOM    801  HG  SER A  47      54.378  72.244  49.946  1.00  4.21
ATOM    802  C   SER A  47      50.506  72.070  52.065  1.00  4.21
ATOM    803  O   SER A  47      49.944  73.170  52.000  1.00  4.21
ATOM    804  N   PHE A  48      50.311  71.230  53.081  1.00  3.35
ATOM    805  H   PHE A  48      50.763  70.319  53.114  1.00  3.35
ATOM    806  CA  PHE A  48      49.468  71.624  54.198  1.00  3.35
ATOM    807  HA  PHE A  48      48.746  72.359  53.854  1.00  3.35
ATOM    808  CB  PHE A  48      48.607  70.474  54.763  1.00  3.35
ATOM    809  HB1 PHE A  48      49.245  69.602  54.868  1.00  3.35
ATOM    810  HB2 PHE A  48      47.822  70.228  54.049  1.00  3.35
ATOM    811  CG  PHE A  48      47.975  70.798  56.120  1.00  3.35
ATOM    812  CD1 PHE A  48      47.277  72.009  56.310  1.00  3.35
ATOM    813  HD1 PHE A  48      47.129  72.685  55.484  1.00  3.35
ATOM    814  CE1 PHE A  48      46.872  72.400  57.600  1.00  3.35
ATOM    815  HE1 PHE A  48      46.436  73.367  57.764  1.00  3.35
ATOM    816  CZ  PHE A  48      47.105  71.567  58.704  1.00  3.35
ATOM    817  HZ  PHE A  48      46.826  71.885  59.700  1.00  3.35
ATOM    818  CE2 PHE A  48      47.789  70.358  58.523  1.00  3.35
ATOM    819  HE2 PHE A  48      48.054  69.758  59.384  1.00  3.35
ATOM    820  CD2 PHE A  48      48.236  69.987  57.245  1.00  3.35
ATOM    821  HD2 PHE A  48      48.860  69.110  57.167  1.00  3.35
ATOM    822  C   PHE A  48      50.218  72.339  55.311  1.00  3.35
ATOM    823  O   PHE A  48      50.151  73.561  55.434  1.00  3.35
ATOM    824  N   ASP A  49      50.933  71.539  56.071  1.00  2.80
ATOM    825  H   ASP A  49      50.908  70.543  55.899  1.00  2.80
ATOM    826  CA  ASP A  49      51.942  71.917  57.015  1.00  2.80
ATOM    827  HA  ASP A  49      52.105  72.995  56.983  1.00  2.80
ATOM    828  CB  ASP A  49      51.503  71.537  58.447  1.00  2.80
ATOM    829  HB1 ASP A  49      50.580  72.076  58.666  1.00  2.80
ATOM    830  HB2 ASP A  49      52.259  71.896  59.147  1.00  2.80
ATOM    831  CG  ASP A  49      51.256  70.040  58.730  1.00  2.80
ATOM    832  OD1 ASP A  49      51.440  69.201  57.815  1.00  2.80
ATOM    833  OD2 ASP A  49      50.858  69.744  59.884  1.00  2.80
ATOM    834  C   ASP A  49      53.238  71.224  56.551  1.00  2.80
ATOM    835  O   ASP A  49      53.265  70.423  55.606  1.00  2.80
ATOM    836  N   PHE A  50      54.325  71.523  57.241  1.00  2.55
ATOM    837  H   PHE A  50      54.231  72.239  57.948  1.00  2.55
ATOM    838  CA  PHE A  50      55.285  70.481  57.562  1.00  2.55
ATOM    839  HA  PHE A  50      54.938  69.508  57.206  1.00  2.55
ATOM    840  CB  PHE A  50      56.675  70.787  56.973  1.00  2.55
ATOM    841  HB1 PHE A  50      57.353  69.987  57.270  1.00  2.55
ATOM    842  HB2 PHE A  50      57.042  71.701  57.441  1.00  2.55
ATOM    843  CG  PHE A  50      56.797  70.947  55.463  1.00  2.55
ATOM    844  CD1 PHE A  50      57.654  71.939  54.949  1.00  2.55
ATOM    845  HD1 PHE A  50      58.204  72.579  55.625  1.00  2.55
```

```
ATOM    846  CE1 PHE A  50      57.817  72.095  53.561  1.00  2.55
ATOM    847  HE1 PHE A  50      58.491  72.851  53.183  1.00  2.55
ATOM    848  CZ  PHE A  50      57.122  71.256  52.675  1.00  2.55
ATOM    849  HZ  PHE A  50      57.257  71.371  51.610  1.00  2.55
ATOM    850  CE2 PHE A  50      56.262  70.266  53.179  1.00  2.55
ATOM    851  HE2 PHE A  50      55.709  69.639  52.498  1.00  2.55
ATOM    852  CD2 PHE A  50      56.113  70.099  54.570  1.00  2.55
ATOM    853  HD2 PHE A  50      55.451  69.335  54.951  1.00  2.55
ATOM    854  C   PHE A  50     55.343  70.443  59.074  1.00  2.55
ATOM    855  O   PHE A  50     55.153  71.475  59.730  1.00  2.55
ATOM    856  N   GLY A  51     55.713  69.300  59.633  1.00  2.73
ATOM    857  H   GLY A  51     55.845  68.459  59.082  1.00  2.73
ATOM    858  CA  GLY A  51      56.110  69.340  61.044  1.00  2.73
ATOM    859  HA1 GLY A  51      55.233  69.190  61.674  1.00  2.73
ATOM    860  HA2 GLY A  51      56.559  70.302  61.296  1.00  2.73
ATOM    861  C   GLY A  51     57.127  68.252  61.342  1.00  2.73
ATOM    862  O   GLY A  51     57.310  67.300  60.569  1.00  2.73
ATOM    863  N   HIS A  52     57.759  68.393  62.504  1.00  2.53
ATOM    864  H   HIS A  52     57.530  69.182  63.093  1.00  2.53
ATOM    865  CA  HIS A  52     58.443  67.288  63.151  1.00  2.53
ATOM    866  HA  HIS A  52     58.295  66.398  62.542  1.00  2.53
ATOM    867  CB  HIS A  52     59.958  67.487  63.231  1.00  2.53
ATOM    868  HB1 HIS A  52      60.317  67.887  62.282  1.00  2.53
ATOM    869  HB2 HIS A  52      60.398  66.500  63.356  1.00  2.53
ATOM    870  CG  HIS A  52     60.447  68.360  64.362  1.00  2.53
ATOM    871  ND1 HIS A  52      60.593  67.988  65.708  1.00  2.53
ATOM    872  CE1 HIS A  52      61.045  69.080  66.342  1.00  2.53
ATOM    873  HE1 HIS A  52      61.229  69.145  67.405  1.00  2.53
ATOM    874  NE2 HIS A  52      61.208  70.087  65.471  1.00  2.53
ATOM    875  HE2 HIS A  52      61.435  71.058  65.687  1.00  2.53
ATOM    876  CD2 HIS A  52      60.837  69.654  64.223  1.00  2.53
ATOM    877  HD2 HIS A  52      60.837  70.231  63.310  1.00  2.53
ATOM    878  C   HIS A  52     57.829  66.974  64.495  1.00  2.53
ATOM    879  O   HIS A  52     57.204  67.832  65.121  1.00  2.53
ATOM    880  N   LEU A  53     58.042  65.737  64.942  1.00  2.39
ATOM    881  H   LEU A  53     58.537  65.070  64.358  1.00  2.39
ATOM    882  CA  LEU A  53      57.770  65.390  66.334  1.00  2.39
ATOM    883  HA  LEU A  53      57.949  66.274  66.950  1.00  2.39
ATOM    884  CB  LEU A  53      56.295  64.943  66.461  1.00  2.39
ATOM    885  HB1 LEU A  53      56.236  63.947  66.904  1.00  2.39
ATOM    886  HB2 LEU A  53      55.869  64.862  65.465  1.00  2.39
ATOM    887  CG  LEU A  53      55.417  65.907  67.277  1.00  2.39
ATOM    888  HG  LEU A  53      55.554  66.928  66.927  1.00  2.39
ATOM    889  CD1 LEU A  53      53.946  65.533  67.097  1.00  2.39
ATOM    890 1HD1 LEU A  53      53.315  66.204  67.680  1.00  2.39
ATOM    891 2HD1 LEU A  53      53.668  65.637  66.047  1.00  2.39
ATOM    892 3HD1 LEU A  53      53.775  64.505  67.408  1.00  2.39
ATOM    893  CD2 LEU A  53      55.765  65.841  68.770  1.00  2.39
ATOM    894 1HD2 LEU A  53      55.091  66.492  69.325  1.00  2.39
ATOM    895 2HD2 LEU A  53      55.668  64.817  69.135  1.00  2.39
ATOM    896 3HD2 LEU A  53      56.782  66.196  68.928  1.00  2.39
ATOM    897  C   LEU A  53     58.724  64.291  66.824  1.00  2.39
ATOM    898  O   LEU A  53     59.362  63.556  66.058  1.00  2.39
```

```
ATOM    899  N   ARG A  54      58.748  64.198  68.155  1.00  2.22
ATOM    900  H   ARG A  54      58.084  64.785  68.638  1.00  2.22
ATOM    901  CA  ARG A  54      59.704  63.503  69.018  1.00  2.22
ATOM    902  HA  ARG A  54      60.225  62.730  68.445  1.00  2.22
ATOM    903  CB  ARG A  54      60.727  64.566  69.503  1.00  2.22
ATOM    904  HB1 ARG A  54      60.228  65.537  69.530  1.00  2.22
ATOM    905  HB2 ARG A  54      61.524  64.645  68.761  1.00  2.22
ATOM    906  CG  ARG A  54      61.356  64.390  70.899  1.00  2.22
ATOM    907  HG1 ARG A  54      60.555  64.310  71.635  1.00  2.22
ATOM    908  HG2 ARG A  54      61.911  65.297  71.141  1.00  2.22
ATOM    909  CD  ARG A  54      62.321  63.204  71.034  1.00  2.22
ATOM    910  HD1 ARG A  54      63.340  63.562  70.918  1.00  2.22
ATOM    911  HD2 ARG A  54      62.125  62.484  70.238  1.00  2.22
ATOM    912  NE  ARG A  54      62.126  62.524  72.326  1.00  2.22
ATOM    913  HE  ARG A  54      61.382  61.831  72.334  1.00  2.22
ATOM    914  CZ  ARG A  54      62.651  62.747  73.517  1.00  2.22
ATOM    915  NH1 ARG A  54      62.061  62.245  74.561  1.00  2.22
ATOM    916 1HH1 ARG A  54      61.138  61.852  74.410  1.00  2.22
ATOM    917 2HH1 ARG A  54      62.428  62.368  75.482  1.00  2.22
ATOM    918  NH2 ARG A  54      63.719  63.457  73.728  1.00  2.22
ATOM    919 1HH2 ARG A  54      64.288  63.675  72.917  1.00  2.22
ATOM    920 2HH2 ARG A  54      64.065  63.604  74.652  1.00  2.22
ATOM    921  C   ARG A  54      58.933  62.838  70.152  1.00  2.22
ATOM    922  O   ARG A  54      58.001  63.428  70.700  1.00  2.22
ATOM    923  N   ASN A  55      59.320  61.607  70.491  1.00  2.25
ATOM    924  H   ASN A  55      60.119  61.201  70.030  1.00  2.25
ATOM    925  CA  ASN A  55      58.608  60.816  71.491  1.00  2.25
ATOM    926  HA  ASN A  55      57.556  60.847  71.196  1.00  2.25
ATOM    927  CB  ASN A  55      58.993  59.321  71.434  1.00  2.25
ATOM    928  HB1 ASN A  55      58.778  58.934  70.440  1.00  2.25
ATOM    929  HB2 ASN A  55      58.364  58.776  72.139  1.00  2.25
ATOM    930  CG  ASN A  55      60.434  58.979  71.753  1.00  2.25
ATOM    931  OD1 ASN A  55      61.281  59.847  71.892  1.00  2.25
ATOM    932  ND2 ASN A  55      60.745  57.707  71.856  1.00  2.25
ATOM    933 1HD2 ASN A  55      60.023  57.004  71.736  1.00  2.25
ATOM    934 2HD2 ASN A  55      61.689  57.439  72.111  1.00  2.25
ATOM    935  C   ASN A  55      58.639  61.423  72.894  1.00  2.25
ATOM    936  O   ASN A  55      59.581  62.089  73.315  1.00  2.25
ATOM    937  N   ARG A  56      57.571  61.178  73.636  1.00  2.51
ATOM    938  H   ARG A  56      56.874  60.537  73.257  1.00  2.51
ATOM    939  CA  ARG A  56      57.375  61.659  75.006  1.00  2.51
ATOM    940  HA  ARG A  56      58.263  62.181  75.364  1.00  2.51
ATOM    941  CB  ARG A  56      56.179  62.633  75.029  1.00  2.51
ATOM    942  HB1 ARG A  56      55.938  62.896  76.060  1.00  2.51
ATOM    943  HB2 ARG A  56      55.311  62.138  74.591  1.00  2.51
ATOM    944  CG  ARG A  56      56.487  63.925  74.246  1.00  2.51
ATOM    945  HG1 ARG A  56      56.912  63.672  73.275  1.00  2.51
ATOM    946  HG2 ARG A  56      57.227  64.514  74.790  1.00  2.51
ATOM    947  CD  ARG A  56      55.247  64.783  73.969  1.00  2.51
ATOM    948  HD1 ARG A  56      54.498  64.167  73.467  1.00  2.51
ATOM    949  HD2 ARG A  56      55.544  65.581  73.285  1.00  2.51
ATOM    950  NE  ARG A  56      54.656  65.379  75.186  1.00  2.51
ATOM    951  HE  ARG A  56      55.034  65.097  76.076  1.00  2.51
```

```
ATOM    952  CZ  ARG A  56      53.660  66.250  75.202  1.00  2.51
ATOM    953  NH1 ARG A  56      53.193  66.717  76.323  1.00  2.51
ATOM    954 1HH1 ARG A  56      53.584  66.434  77.205  1.00  2.51
ATOM    955 2HH1 ARG A  56      52.440  67.379  76.307  1.00  2.51
ATOM    956  NH2 ARG A  56      53.111  66.681  74.104  1.00  2.51
ATOM    957 1HH2 ARG A  56      53.473  66.359  73.224  1.00  2.51
ATOM    958 2HH2 ARG A  56      52.344  67.323  74.128  1.00  2.51
ATOM    959  C   ARG A  56      57.192  60.433  75.896  1.00  2.51
ATOM    960  O   ARG A  56      57.161  59.304  75.413  1.00  2.51
ATOM    961  N   THR A  57      57.102  60.652  77.200  1.00  3.44
ATOM    962  H   THR A  57      57.205  61.588  77.557  1.00  3.44
ATOM    963  CA  THR A  57      56.935  59.587  78.193  1.00  3.44
ATOM    964  HA  THR A  57      57.852  58.999  78.217  1.00  3.44
ATOM    965  CB  THR A  57      56.733  60.196  79.589  1.00  3.44
ATOM    966  HB  THR A  57      55.736  60.631  79.664  1.00  3.44
ATOM    967  CG2 THR A  57      56.923  59.167  80.701  1.00  3.44
ATOM    968 1HG2 THR A  57      56.783  59.644  81.671  1.00  3.44
ATOM    969 2HG2 THR A  57      56.184  58.372  80.601  1.00  3.44
ATOM    970 3HG2 THR A  57      57.924  58.736  80.652  1.00  3.44
ATOM    971  OG1 THR A  57      57.686  61.216  79.802  1.00  3.44
ATOM    972  HG1 THR A  57      57.674  61.430  80.739  1.00  3.44
ATOM    973  C   THR A  57      55.770  58.661  77.844  1.00  3.44
ATOM    974  O   THR A  57      54.626  59.103  77.814  1.00  3.44
ATOM    975  N   GLY A  58      56.071  57.402  77.509  1.00  2.96
ATOM    976  H   GLY A  58      57.043  57.135  77.500  1.00  2.96
ATOM    977  CA  GLY A  58      55.114  56.405  77.011  1.00  2.96
ATOM    978  HA1 GLY A  58      54.234  56.411  77.657  1.00  2.96
ATOM    979  HA2 GLY A  58      55.568  55.415  77.067  1.00  2.96
ATOM    980  C   GLY A  58      54.632  56.623  75.567  1.00  2.96
ATOM    981  O   GLY A  58      54.525  55.677  74.789  1.00  2.96
ATOM    982  N   CYS A  59      54.358  57.873  75.198  1.00  2.48
ATOM    983  H   CYS A  59      54.411  58.579  75.920  1.00  2.48
ATOM    984  CA  CYS A  59      53.806  58.274  73.910  1.00  2.48
ATOM    985  HA  CYS A  59      53.054  57.538  73.614  1.00  2.48
ATOM    986  CB  CYS A  59      53.086  59.619  74.077  1.00  2.48
ATOM    987  HB1 CYS A  59      52.642  59.891  73.120  1.00  2.48
ATOM    988  HB2 CYS A  59      53.804  60.383  74.370  1.00  2.48
ATOM    989  SG  CYS A  59      51.778  59.491  75.333  1.00  2.48
ATOM    990  HG  CYS A  59      51.233  60.701  75.170  1.00  2.48
ATOM    991  C   CYS A  59      54.851  58.328  72.788  1.00  2.48
ATOM    992  O   CYS A  59      55.463  59.374  72.529  1.00  2.48
ATOM    993  N   HIS A  60      55.038  57.190  72.116  1.00  2.13
ATOM    994  H   HIS A  60      54.587  56.357  72.466  1.00  2.13
ATOM    995  CA  HIS A  60      55.819  57.093  70.885  1.00  2.13
ATOM    996  HA  HIS A  60      56.826  57.394  71.164  1.00  2.13
ATOM    997  CB  HIS A  60      55.910  55.632  70.414  1.00  2.13
ATOM    998  HB1 HIS A  60      55.273  55.506  69.540  1.00  2.13
ATOM    999  HB2 HIS A  60      55.528  54.970  71.191  1.00  2.13
ATOM   1000  CG  HIS A  60      57.313  55.156  70.086  1.00  2.13
ATOM   1001  ND1 HIS A  60      58.526  55.681  70.570  1.00  2.13
ATOM   1002  CE1 HIS A  60      59.492  54.952  69.982  1.00  2.13
ATOM   1003  HE1 HIS A  60      60.553  55.126  70.103  1.00  2.13
ATOM   1004  NE2 HIS A  60      58.958  54.011  69.185  1.00  2.13
```

```
ATOM   1005  HE2 HIS A  60      59.458  53.412  68.547  1.00  2.13
ATOM   1006  CD2 HIS A  60      57.594  54.118  69.249  1.00  2.13
ATOM   1007  HD2 HIS A  60      56.877  53.555  68.670  1.00  2.13
ATOM   1008  C   HIS A  60     55.362  58.093  69.811  1.00  2.13
ATOM   1009  O   HIS A  60     54.182  58.465  69.701  1.00  2.13
ATOM   1010  N   ALA A  61     56.340  58.589  69.051  1.00  2.10
ATOM   1011  H   ALA A  61     57.220  58.091  69.033  1.00  2.10
ATOM   1012  CA  ALA A  61     56.153  59.722  68.144  1.00  2.10
ATOM   1013  HA  ALA A  61     55.780  60.567  68.722  1.00  2.10
ATOM   1014  CB  ALA A  61     57.504  60.119  67.552  1.00  2.10
ATOM   1015  HB1 ALA A  61     57.382  60.975  66.886  1.00  2.10
ATOM   1016  HB2 ALA A  61     58.199  60.384  68.344  1.00  2.10
ATOM   1017  HB3 ALA A  61     57.897  59.277  66.985  1.00  2.10
ATOM   1018  C   ALA A  61     55.127  59.462  67.027  1.00  2.10
ATOM   1019  O   ALA A  61     54.501  60.380  66.503  1.00  2.10
ATOM   1020  N   GLU A  62     54.934  58.195  66.704  1.00  2.14
ATOM   1021  H   GLU A  62     55.520  57.512  67.180  1.00  2.14
ATOM   1022  CA  GLU A  62     54.005  57.682  65.713  1.00  2.14
ATOM   1023  HA  GLU A  62     54.095  58.276  64.803  1.00  2.14
ATOM   1024  CB  GLU A  62     54.397  56.230  65.371  1.00  2.14
ATOM   1025  HB1 GLU A  62     53.754  55.901  64.558  1.00  2.14
ATOM   1026  HB2 GLU A  62     54.185  55.587  66.226  1.00  2.14
ATOM   1027  CG  GLU A  62     55.873  56.034  64.925  1.00  2.14
ATOM   1028  HG1 GLU A  62     56.088  56.744  64.124  1.00  2.14
ATOM   1029  HG2 GLU A  62     55.964  55.030  64.506  1.00  2.14
ATOM   1030  CD  GLU A  62     56.915  56.180  66.059  1.00  2.14
ATOM   1031  OE1 GLU A  62     56.504  55.988  67.229  1.00  2.14
ATOM   1032  OE2 GLU A  62     58.087  56.527  65.783  1.00  2.14
ATOM   1033  C   GLU A  62     52.547  57.800  66.188  1.00  2.14
ATOM   1034  O   GLU A  62     51.666  58.219  65.436  1.00  2.14
ATOM   1035  N   LEU A  63     52.310  57.534  67.477  1.00  2.22
ATOM   1036  H   LEU A  63     53.108  57.296  68.056  1.00  2.22
ATOM   1037  CA  LEU A  63     51.019  57.748  68.136  1.00  2.22
ATOM   1038  HA  LEU A  63     50.230  57.295  67.536  1.00  2.22
ATOM   1039  CB  LEU A  63     51.039  57.091  69.535  1.00  2.22
ATOM   1040  HB1 LEU A  63     50.025  57.128  69.936  1.00  2.22
ATOM   1041  HB2 LEU A  63     51.668  57.690  70.195  1.00  2.22
ATOM   1042  CG  LEU A  63     51.544  55.638  69.610  1.00  2.22
ATOM   1043  HG  LEU A  63     52.593  55.603  69.315  1.00  2.22
ATOM   1044  CD1 LEU A  63     51.438  55.129  71.046  1.00  2.22
ATOM   1045 1HD1 LEU A  63     51.832  54.115  71.099  1.00  2.22
ATOM   1046 2HD1 LEU A  63     52.024  55.767  71.707  1.00  2.22
ATOM   1047 3HD1 LEU A  63     50.397  55.129  71.369  1.00  2.22
ATOM   1048  CD2 LEU A  63     50.742  54.707  68.706  1.00  2.22
ATOM   1049 1HD2 LEU A  63     51.068  53.677  68.848  1.00  2.22
ATOM   1050 2HD2 LEU A  63     49.679  54.776  68.937  1.00  2.22
ATOM   1051 3HD2 LEU A  63     50.901  54.972  67.662  1.00  2.22
ATOM   1052  C   LEU A  63     50.706  59.243  68.295  1.00  2.22
ATOM   1053  O   LEU A  63     49.573  59.697  68.072  1.00  2.22
ATOM   1054  N   LEU A  64     51.732  60.021  68.663  1.00  2.13
ATOM   1055  H   LEU A  64     52.635  59.590  68.832  1.00  2.13
ATOM   1056  CA  LEU A  64     51.590  61.475  68.731  1.00  2.13
ATOM   1057  HA  LEU A  64     50.772  61.703  69.414  1.00  2.13
```

```
ATOM   1058  CB  LEU A  64      52.876  62.121  69.274  1.00  2.13
ATOM   1059  HB1 LEU A  64      52.785  63.203  69.181  1.00  2.13
ATOM   1060  HB2 LEU A  64      53.718  61.801  68.661  1.00  2.13
ATOM   1061  CG  LEU A  64      53.157  61.778  70.750  1.00  2.13
ATOM   1062  HG  LEU A  64      53.220  60.698  70.873  1.00  2.13
ATOM   1063  CD1 LEU A  64      54.492  62.385  71.177  1.00  2.13
ATOM   1064 1HD1 LEU A  64      54.724  62.075  72.194  1.00  2.13
ATOM   1065 2HD1 LEU A  64      55.282  62.015  70.524  1.00  2.13
ATOM   1066 3HD1 LEU A  64      54.456  63.471  71.105  1.00  2.13
ATOM   1067  CD2 LEU A  64      52.081  62.323  71.697  1.00  2.13
ATOM   1068 1HD2 LEU A  64      52.400  62.205  72.731  1.00  2.13
ATOM   1069 2HD2 LEU A  64      51.894  63.376  71.488  1.00  2.13
ATOM   1070 3HD2 LEU A  64      51.152  61.766  71.563  1.00  2.13
ATOM   1071  C   LEU A  64     51.155  62.050  67.381  1.00  2.13
ATOM   1072  O   LEU A  64     50.192  62.826  67.316  1.00  2.13
ATOM   1073  N   PHE A  65     51.792  61.609  66.294  1.00  2.20
ATOM   1074  H   PHE A  65     52.592  60.991  66.386  1.00  2.20
ATOM   1075  CA  PHE A  65      51.369  62.048  64.976  1.00  2.20
ATOM   1076  HA  PHE A  65      51.229  63.122  65.098  1.00  2.20
ATOM   1077  CB  PHE A  65      52.446  61.930  63.904  1.00  2.20
ATOM   1078  HB1 PHE A  65      52.073  61.286  63.112  1.00  2.20
ATOM   1079  HB2 PHE A  65      53.346  61.474  64.317  1.00  2.20
ATOM   1080  CG  PHE A  65      52.780  63.304  63.336  1.00  2.20
ATOM   1081  CD1 PHE A  65      54.027  63.901  63.588  1.00  2.20
ATOM   1082  HD1 PHE A  65      54.781  63.352  64.129  1.00  2.20
ATOM   1083  CE1 PHE A  65      54.294  65.205  63.128  1.00  2.20
ATOM   1084  HE1 PHE A  65      55.259  65.655  63.314  1.00  2.20
ATOM   1085  CZ  PHE A  65     53.313  65.920  62.419  1.00  2.20
ATOM   1086  HZ  PHE A  65     53.519  66.917  62.055  1.00  2.20
ATOM   1087  CE2 PHE A  65      52.067  65.328  62.157  1.00  2.20
ATOM   1088  HE2 PHE A  65      51.317  65.869  61.593  1.00  2.20
ATOM   1089  CD2 PHE A  65      51.804  64.028  62.624  1.00  2.20
ATOM   1090  HD2 PHE A  65      50.841  63.586  62.447  1.00  2.20
ATOM   1091  C   PHE A  65     49.969  61.564  64.557  1.00  2.20
ATOM   1092  O   PHE A  65     49.213  62.362  63.995  1.00  2.20
ATOM   1093  N   LEU A  66     49.563  60.339  64.918  1.00  2.34
ATOM   1094  H   LEU A  66     50.237  59.701  65.330  1.00  2.34
ATOM   1095  CA  LEU A  66      48.171  59.891  64.758  1.00  2.34
ATOM   1096  HA  LEU A  66      47.938  59.891  63.693  1.00  2.34
ATOM   1097  CB  LEU A  66      47.992  58.459  65.307  1.00  2.34
ATOM   1098  HB1 LEU A  66      46.927  58.298  65.486  1.00  2.34
ATOM   1099  HB2 LEU A  66      48.488  58.380  66.270  1.00  2.34
ATOM   1100  CG  LEU A  66      48.482  57.318  64.402  1.00  2.34
ATOM   1101  HG  LEU A  66      49.553  57.401  64.233  1.00  2.34
ATOM   1102  CD1 LEU A  66      48.195  55.979  65.085  1.00  2.34
ATOM   1103 1HD1 LEU A  66      48.604  55.175  64.481  1.00  2.34
ATOM   1104 2HD1 LEU A  66      48.672  55.955  66.065  1.00  2.34
ATOM   1105 3HD1 LEU A  66      47.119  55.841  65.200  1.00  2.34
ATOM   1106  CD2 LEU A  66      47.765  57.335  63.053  1.00  2.34
ATOM   1107 1HD2 LEU A  66      47.811  56.359  62.580  1.00  2.34
ATOM   1108 2HD2 LEU A  66      46.721  57.599  63.207  1.00  2.34
ATOM   1109 3HD2 LEU A  66      48.225  58.073  62.400  1.00  2.34
ATOM   1110  C   LEU A  66     47.117  60.816  65.391  1.00  2.34
```

```
ATOM   1111  O   LEU A  66     46.071  61.042  64.772  1.00  2.34
ATOM   1112  N   SER A  67     47.440  61.334  66.587  1.00  2.45
ATOM   1113  H   SER A  67     48.320  61.039  66.989  1.00  2.45
ATOM   1114  CA  SER A  67     46.622  62.300  67.353  1.00  2.45
ATOM   1115  HA  SER A  67     45.600  61.925  67.417  1.00  2.45
ATOM   1116  CB  SER A  67     47.184  62.462  68.771  1.00  2.45
ATOM   1117  HB1 SER A  67     46.457  63.003  69.378  1.00  2.45
ATOM   1118  HB2 SER A  67     48.101  63.050  68.734  1.00  2.45
ATOM   1119  OG  SER A  67     47.474  61.223  69.389  1.00  2.45
ATOM   1120  HG  SER A  67     48.098  60.723  68.840  1.00  2.45
ATOM   1121  C   SER A  67     46.568  63.714  66.729  1.00  2.45
ATOM   1122  O   SER A  67     45.576  64.460  66.817  1.00  2.45
ATOM   1123  N   TYR A  68     47.687  64.114  66.115  1.00  2.51
ATOM   1124  H   TYR A  68     48.485  63.491  66.104  1.00  2.51
ATOM   1125  CA  TYR A  68     47.746  65.358  65.351  1.00  2.51
ATOM   1126  HA  TYR A  68     47.365  66.154  65.988  1.00  2.51
ATOM   1127  CB  TYR A  68     49.213  65.672  64.985  1.00  2.51
ATOM   1128  HB1 TYR A  68     49.323  65.695  63.900  1.00  2.51
ATOM   1129  HB2 TYR A  68     49.862  64.872  65.340  1.00  2.51
ATOM   1130  CG  TYR A  68     49.752  66.971  65.558  1.00  2.51
ATOM   1131  CD1 TYR A  68     49.688  67.208  66.947  1.00  2.51
ATOM   1132  HD1 TYR A  68     49.249  66.466  67.605  1.00  2.51
ATOM   1133  CE1 TYR A  68     50.213  68.400  67.482  1.00  2.51
ATOM   1134  HE1 TYR A  68     50.171  68.595  68.543  1.00  2.51
ATOM   1135  CZ  TYR A  68     50.827  69.349  66.636  1.00  2.51
ATOM   1136  OH  TYR A  68     51.302  70.502  67.177  1.00  2.51
ATOM   1137  HH  TYR A  68     51.608  71.147  66.526  1.00  2.51
ATOM   1138  CE2 TYR A  68     50.908  69.102  65.248  1.00  2.51
ATOM   1139  HE2 TYR A  68     51.404  69.811  64.597  1.00  2.51
ATOM   1140  CD2 TYR A  68     50.363  67.919  64.713  1.00  2.51
ATOM   1141  HD2 TYR A  68     50.443  67.726  63.649  1.00  2.51
ATOM   1142  C   TYR A  68     46.850  65.309  64.113  1.00  2.51
ATOM   1143  O   TYR A  68     45.969  66.166  63.984  1.00  2.51
ATOM   1144  N   LEU A  69     47.023  64.277  63.277  1.00  2.54
ATOM   1145  H   LEU A  69     47.740  63.606  63.535  1.00  2.54
ATOM   1146  CA  LEU A  69     45.964  63.765  62.405  1.00  2.54
ATOM   1147  HA  LEU A  69     45.660  64.502  61.683  1.00  2.54
ATOM   1148  CB  LEU A  69     46.449  62.487  61.672  1.00  2.54
ATOM   1149  HB1 LEU A  69     45.563  61.947  61.342  1.00  2.54
ATOM   1150  HB2 LEU A  69     46.959  61.843  62.389  1.00  2.54
ATOM   1151  CG  LEU A  69     47.340  62.666  60.426  1.00  2.54
ATOM   1152  HG  LEU A  69     46.819  63.306  59.715  1.00  2.54
ATOM   1153  CD1 LEU A  69     48.713  63.257  60.723  1.00  2.54
ATOM   1154 1HD1 LEU A  69     49.278  63.377  59.799  1.00  2.54
ATOM   1155 2HD1 LEU A  69     48.621  64.236  61.188  1.00  2.54
ATOM   1156 3HD1 LEU A  69     49.253  62.591  61.393  1.00  2.54
ATOM   1157  CD2 LEU A  69     47.597  61.309  59.762  1.00  2.54
ATOM   1158 1HD2 LEU A  69     48.210  60.683  60.411  1.00  2.54
ATOM   1159 2HD2 LEU A  69     46.654  60.809  59.580  1.00  2.54
ATOM   1160 3HD2 LEU A  69     48.107  61.455  58.812  1.00  2.54
ATOM   1161  C   LEU A  69     44.746  63.431  63.288  1.00  2.54
ATOM   1162  O   LEU A  69     44.807  63.346  64.497  1.00  2.54
ATOM   1163  N   GLY A  70     43.578  63.336  62.700  1.00  3.14
```

```
ATOM   1164  H   GLY A  70     43.533  63.534  61.715  1.00  3.14
ATOM   1165  CA  GLY A  70     42.321  63.406  63.447  1.00  3.14
ATOM   1166  HA1 GLY A  70     42.279  62.561  64.135  1.00  3.14
ATOM   1167  HA2 GLY A  70     41.510  63.307  62.738  1.00  3.14
ATOM   1168  C   GLY A  70     42.041  64.680  64.270  1.00  3.14
ATOM   1169  O   GLY A  70     40.878  64.884  64.620  1.00  3.14
ATOM   1170  N   ALA A  71     43.017  65.576  64.489  1.00  3.09
ATOM   1171  H   ALA A  71     43.980  65.296  64.362  1.00  3.09
ATOM   1172  CA  ALA A  71     42.724  66.988  64.769  1.00  3.09
ATOM   1173  HA  ALA A  71     41.658  67.167  64.646  1.00  3.09
ATOM   1174  CB  ALA A  71     43.080  67.236  66.240  1.00  3.09
ATOM   1175  HB1 ALA A  71     44.142  67.061  66.402  1.00  3.09
ATOM   1176  HB2 ALA A  71     42.821  68.254  66.524  1.00  3.09
ATOM   1177  HB3 ALA A  71     42.511  66.553  66.869  1.00  3.09
ATOM   1178  C   ALA A  71     43.427  67.949  63.795  1.00  3.09
ATOM   1179  O   ALA A  71     43.822  69.044  64.220  1.00  3.09
ATOM   1180  N   LEU A  72     43.686  67.546  62.547  1.00  3.16
ATOM   1181  H   LEU A  72     43.327  66.641  62.225  1.00  3.16
ATOM   1182  CA  LEU A  72     44.627  68.277  61.707  1.00  3.16
ATOM   1183  HA  LEU A  72     45.223  68.813  62.438  1.00  3.16
ATOM   1184  CB  LEU A  72     45.739  67.414  61.044  1.00  3.16
ATOM   1185  HB1 LEU A  72     45.902  66.560  61.657  1.00  3.16
ATOM   1186  HB2 LEU A  72     46.630  68.032  61.138  1.00  3.16
ATOM   1187  CG  LEU A  72     45.641  66.982  59.589  1.00  3.16
ATOM   1188  HG  LEU A  72     45.429  67.840  58.968  1.00  3.16
ATOM   1189  CD1 LEU A  72     47.001  66.464  59.100  1.00  3.16
ATOM   1190 1HD1 LEU A  72     47.705  67.297  59.132  1.00  3.16
ATOM   1191 2HD1 LEU A  72     47.336  65.662  59.751  1.00  3.16
ATOM   1192 3HD1 LEU A  72     46.906  66.100  58.087  1.00  3.16
ATOM   1193  CD2 LEU A  72     44.620  65.874  59.327  1.00  3.16
ATOM   1194 1HD2 LEU A  72     44.955  64.932  59.738  1.00  3.16
ATOM   1195 2HD2 LEU A  72     43.690  66.153  59.812  1.00  3.16
ATOM   1196 3HD2 LEU A  72     44.453  65.770  58.256  1.00  3.16
ATOM   1197  C   LEU A  72     44.146  69.430  60.848  1.00  3.16
ATOM   1198  O   LEU A  72     44.736  70.511  60.913  1.00  3.16
ATOM   1199  N   CYS A  73     43.172  69.217  59.962  1.00  6.69
ATOM   1200  H   CYS A  73     42.697  68.317  60.003  1.00  6.69
ATOM   1201  CA  CYS A  73     43.283  69.862  58.674  1.00  6.69
ATOM   1202  HA  CYS A  73     44.222  70.421  58.652  1.00  6.69
ATOM   1203  CB  CYS A  73     43.412  68.842  57.537  1.00  6.69
ATOM   1204  HB1 CYS A  73     42.548  68.824  56.904  1.00  6.69
ATOM   1205  HB2 CYS A  73     43.529  67.833  57.924  1.00  6.69
ATOM   1206  SG  CYS A  73     44.892  69.196  56.535  1.00  6.69
ATOM   1207  HG  CYS A  73     45.804  69.245  57.513  1.00  6.69
ATOM   1208  C   CYS A  73     42.251  71.015  58.348  1.00  6.69
ATOM   1209  O   CYS A  73     41.052  70.717  58.500  1.00  6.69
ATOM   1210  N   PRO A  74     42.554  72.084  57.590  1.00  8.59
ATOM   1211  CD  PRO A  74     42.498  71.820  56.155  1.00  8.59
ATOM   1212  HD1 PRO A  74     41.599  72.279  55.787  1.00  8.59
ATOM   1213  HD2 PRO A  74     42.459  70.765  55.948  1.00  8.59
ATOM   1214  CG  PRO A  74     43.745  72.444  55.566  1.00  8.59
ATOM   1215  HG1 PRO A  74     43.612  72.637  54.503  1.00  8.59
ATOM   1216  HG2 PRO A  74     44.615  71.813  55.717  1.00  8.59
```

```
ATOM   1217  CB  PRO A  74      43.858  73.720  56.372  1.00  8.59
ATOM   1218  HB1 PRO A  74      43.184  74.386  55.886  1.00  8.59
ATOM   1219  HB2 PRO A  74      44.869  74.081  56.367  1.00  8.59
ATOM   1220  CA  PRO A  74      43.388  73.345  57.783  1.00  8.59
ATOM   1221  HA  PRO A  74      44.269  72.974  58.284  1.00  8.59
ATOM   1222  C   PRO A  74      42.893  74.366  58.750  1.00  8.59
ATOM   1223  O   PRO A  74      41.821  74.907  58.493  1.00  8.59
ATOM   1224  N   GLY A  75      43.718  74.758  59.704  1.00  9.87
ATOM   1225  H   GLY A  75      44.483  74.156  59.920  1.00  9.87
ATOM   1226  CA  GLY A  75      43.784  76.154  60.134  1.00  9.87
ATOM   1227  HA1 GLY A  75      44.691  76.312  60.716  1.00  9.87
ATOM   1228  HA2 GLY A  75      42.902  76.375  60.722  1.00  9.87
ATOM   1229  C   GLY A  75      43.813  77.084  58.908  1.00  9.87
ATOM   1230  O   GLY A  75      43.134  78.104  58.878  1.00  9.87
ATOM   1231  N   LEU A  76      44.505  76.656  57.845  1.00 10.57
ATOM   1232  H   LEU A  76      45.069  75.830  57.963  1.00 10.57
ATOM   1233  CA  LEU A  76      44.530  77.309  56.538  1.00 10.57
ATOM   1234  HA  LEU A  76      44.669  78.369  56.753  1.00 10.57
ATOM   1235  CB  LEU A  76      45.792  76.852  55.772  1.00 10.57
ATOM   1236  HB1 LEU A  76      45.667  75.826  55.451  1.00 10.57
ATOM   1237  HB2 LEU A  76      46.616  76.877  56.473  1.00 10.57
ATOM   1238  CG  LEU A  76      46.183  77.662  54.525  1.00 10.57
ATOM   1239  HG  LEU A  76      45.499  77.428  53.709  1.00 10.57
ATOM   1240  CD1 LEU A  76      46.217  79.177  54.730  1.00 10.57
ATOM   1241 1HD1 LEU A  76      45.212  79.556  54.909  1.00 10.57
ATOM   1242 2HD1 LEU A  76      46.857  79.429  55.574  1.00 10.57
ATOM   1243 3HD1 LEU A  76      46.597  79.663  53.832  1.00 10.57
ATOM   1244  CD2 LEU A  76      47.585  77.216  54.101  1.00 10.57
ATOM   1245 1HD2 LEU A  76      47.828  77.629  53.123  1.00 10.57
ATOM   1246 2HD2 LEU A  76      48.323  77.575  54.820  1.00 10.57
ATOM   1247 3HD2 LEU A  76      47.650  76.133  54.075  1.00 10.57
ATOM   1248  C   LEU A  76      43.240  77.266  55.694  1.00 10.57
ATOM   1249  O   LEU A  76      43.130  77.974  54.706  1.00 10.57
ATOM   1250  N   TRP A  77      42.261  76.461  56.104  1.00 10.15
ATOM   1251  H   TRP A  77      42.420  75.949  56.958  1.00 10.15
ATOM   1252  CA  TRP A  77      40.914  76.309  55.541  1.00 10.15
ATOM   1253  HA  TRP A  77      40.827  76.917  54.642  1.00 10.15
ATOM   1254  CB  TRP A  77      40.535  74.869  55.177  1.00 10.15
ATOM   1255  HB1 TRP A  77      39.452  74.857  55.074  1.00 10.15
ATOM   1256  HB2 TRP A  77      40.755  74.242  56.038  1.00 10.15
ATOM   1257  CG  TRP A  77      41.016  74.173  53.926  1.00 10.15
ATOM   1258  CD1 TRP A  77      40.611  72.922  53.629  1.00 10.15
ATOM   1259  HD1 TRP A  77      39.941  72.341  54.251  1.00 10.15
ATOM   1260  NE1 TRP A  77      41.183  72.484  52.461  1.00 10.15
ATOM   1261  HE1 TRP A  77      41.005  71.561  52.093  1.00 10.15
ATOM   1262  CE2 TRP A  77      41.904  73.483  51.858  1.00 10.15
ATOM   1263  CZ2 TRP A  77      42.577  73.544  50.629  1.00 10.15
ATOM   1264  HZ2 TRP A  77      42.576  72.691  49.963  1.00 10.15
ATOM   1265  CH2 TRP A  77      43.180  74.751  50.242  1.00 10.15
ATOM   1266  HH2 TRP A  77      43.647  74.842  49.270  1.00 10.15
ATOM   1267  CZ3 TRP A  77      43.120  75.862  51.102  1.00 10.15
ATOM   1268  HZ3 TRP A  77      43.542  76.808  50.785  1.00 10.15
ATOM   1269  CE3 TRP A  77      42.449  75.781  52.338  1.00 10.15
```

```
ATOM   1270  HE3 TRP A  77      42.343  76.671  52.931  1.00 10.15
ATOM   1271  CD2 TRP A  77      41.819  74.590  52.766  1.00 10.15
ATOM   1272  C   TRP A  77      39.883  76.892  56.545  1.00 10.15
ATOM   1273  O   TRP A  77      38.687  76.928  56.267  1.00 10.15
ATOM   1274  N   GLY A  78      40.340  77.264  57.748  1.00  9.88
ATOM   1275  H   GLY A  78      41.341  77.381  57.803  1.00  9.88
ATOM   1276  CA  GLY A  78      39.733  76.925  59.048  1.00  9.88
ATOM   1277  HA1 GLY A  78      39.226  77.812  59.420  1.00  9.88
ATOM   1278  HA2 GLY A  78      40.534  76.684  59.741  1.00  9.88
ATOM   1279  C   GLY A  78      38.737  75.771  59.089  1.00  9.88
ATOM   1280  O   GLY A  78      37.749  75.844  59.817  1.00  9.88
ATOM   1281  N   CYS A  79      38.968  74.697  58.337  1.00  8.59
ATOM   1282  H   CYS A  79      39.836  74.656  57.829  1.00  8.59
ATOM   1283  CA  CYS A  79      38.172  73.491  58.474  1.00  8.59
ATOM   1284  HA  CYS A  79      37.118  73.749  58.348  1.00  8.59
ATOM   1285  CB  CYS A  79      38.568  72.469  57.411  1.00  8.59
ATOM   1286  HB1 CYS A  79      38.287  71.473  57.764  1.00  8.59
ATOM   1287  HB2 CYS A  79      39.648  72.477  57.256  1.00  8.59
ATOM   1288  SG  CYS A  79      37.694  72.807  55.854  1.00  8.59
ATOM   1289  HG  CYS A  79      36.462  72.643  56.321  1.00  8.59
ATOM   1290  C   CYS A  79      38.381  72.893  59.889  1.00  8.59
ATOM   1291  O   CYS A  79      39.450  72.395  60.218  1.00  8.59
ATOM   1292  N   ALA A  80      37.330  72.893  60.687  1.00  8.18
ATOM   1293  H   ALA A  80      36.527  73.440  60.453  1.00  8.18
ATOM   1294  CA  ALA A  80      36.987  71.661  61.402  1.00  8.18
ATOM   1295  HA  ALA A  80      37.893  71.152  61.734  1.00  8.18
ATOM   1296  CB  ALA A  80      36.137  72.016  62.614  1.00  8.18
ATOM   1297  HB1 ALA A  80      36.668  72.726  63.247  1.00  8.18
ATOM   1298  HB2 ALA A  80      35.193  72.465  62.308  1.00  8.18
ATOM   1299  HB3 ALA A  80      35.915  71.135  63.218  1.00  8.18
ATOM   1300  C   ALA A  80      36.242  70.707  60.403  1.00  8.18
ATOM   1301  O   ALA A  80      36.026  71.046  59.261  1.00  8.18
ATOM   1302  N   ASP A  81      35.852  69.522  60.900  1.00  6.65
ATOM   1303  H   ASP A  81      35.909  69.400  61.899  1.00  6.65
ATOM   1304  CA  ASP A  81      36.034  68.273  60.134  1.00  6.65
ATOM   1305  HA  ASP A  81      35.933  67.462  60.857  1.00  6.65
ATOM   1306  CB  ASP A  81      34.975  67.983  59.046  1.00  6.65
ATOM   1307  HB1 ASP A  81      34.982  68.771  58.292  1.00  6.65
ATOM   1308  HB2 ASP A  81      33.986  67.947  59.505  1.00  6.65
ATOM   1309  CG  ASP A  81      35.258  66.624  58.361  1.00  6.65
ATOM   1310  OD1 ASP A  81      34.862  66.396  57.195  1.00  6.65
ATOM   1311  OD2 ASP A  81      35.860  65.738  59.010  1.00  6.65
ATOM   1312  C   ASP A  81      37.444  68.198  59.582  1.00  6.65
ATOM   1313  O   ASP A  81      37.712  68.513  58.423  1.00  6.65
ATOM   1314  N   ASP A  82      38.322  67.775  60.492  1.00  3.72
ATOM   1315  H   ASP A  82      37.958  67.454  61.376  1.00  3.72
ATOM   1316  CA  ASP A  82      39.773  67.825  60.414  1.00  3.72
ATOM   1317  HA  ASP A  82      40.046  68.881  60.383  1.00  3.72
ATOM   1318  CB  ASP A  82      40.349  67.216  61.700  1.00  3.72
ATOM   1319  HB1 ASP A  82      39.646  67.320  62.528  1.00  3.72
ATOM   1320  HB2 ASP A  82      41.230  67.793  61.964  1.00  3.72
ATOM   1321  CG  ASP A  82      40.741  65.740  61.539  1.00  3.72
ATOM   1322  OD1 ASP A  82      39.835  64.875  61.479  1.00  3.72
```

```
ATOM   1323  OD2 ASP A  82      41.957  65.453  61.456  1.00  3.72
ATOM   1324  C   ASP A  82      40.395  67.187  59.170  1.00  3.72
ATOM   1325  O   ASP A  82      41.610  67.152  59.057  1.00  3.72
ATOM   1326  N   ARG A  83      39.615  66.637  58.248  1.00  3.24
ATOM   1327  H   ARG A  83      38.620  66.740  58.392  1.00  3.24
ATOM   1328  CA  ARG A  83      40.137  65.845  57.145  1.00  3.24
ATOM   1329  HA  ARG A  83      41.177  66.130  56.971  1.00  3.24
ATOM   1330  CB  ARG A  83      40.101  64.358  57.566  1.00  3.24
ATOM   1331  HB1 ARG A  83      39.893  63.722  56.705  1.00  3.24
ATOM   1332  HB2 ARG A  83      39.287  64.206  58.279  1.00  3.24
ATOM   1333  CG  ARG A  83      41.433  63.882  58.173  1.00  3.24
ATOM   1334  HG1 ARG A  83      41.776  64.581  58.933  1.00  3.24
ATOM   1335  HG2 ARG A  83      42.184  63.836  57.385  1.00  3.24
ATOM   1336  CD  ARG A  83      41.279  62.495  58.810  1.00  3.24
ATOM   1337  HD1 ARG A  83      42.266  62.104  59.050  1.00  3.24
ATOM   1338  HD2 ARG A  83      40.779  61.837  58.101  1.00  3.24
ATOM   1339  NE  ARG A  83      40.501  62.592  60.047  1.00  3.24
ATOM   1340  HE  ARG A  83      40.288  63.544  60.351  1.00  3.24
ATOM   1341  CZ  ARG A  83      40.157  61.674  60.925  1.00  3.24
ATOM   1342  NH1 ARG A  83      39.492  62.076  61.966  1.00  3.24
ATOM   1343 1HH1 ARG A  83      39.394  63.092  62.079  1.00  3.24
ATOM   1344 2HH1 ARG A  83      39.272  61.458  62.716  1.00  3.24
ATOM   1345  NH2 ARG A  83      40.444  60.402  60.811  1.00  3.24
ATOM   1346 1HH2 ARG A  83      40.989  60.062  60.037  1.00  3.24
ATOM   1347 2HH2 ARG A  83      40.246  59.766  61.559  1.00  3.24
ATOM   1348  C   ARG A  83      39.505  66.007  55.785  1.00  3.24
ATOM   1349  O   ARG A  83      40.014  65.435  54.821  1.00  3.24
ATOM   1350  N   ASN A  84      38.429  66.770  55.683  1.00  3.60
ATOM   1351  H   ASN A  84      38.088  67.198  56.534  1.00  3.60
ATOM   1352  CA  ASN A  84      37.812  67.115  54.405  1.00  3.60
ATOM   1353  HA  ASN A  84      36.844  67.556  54.653  1.00  3.60
ATOM   1354  CB  ASN A  84      38.630  68.215  53.697  1.00  3.60
ATOM   1355  HB1 ASN A  84      37.989  68.704  52.966  1.00  3.60
ATOM   1356  HB2 ASN A  84      39.423  67.710  53.157  1.00  3.60
ATOM   1357  CG  ASN A  84      39.296  69.301  54.539  1.00  3.60
ATOM   1358  OD1 ASN A  84      40.295  69.861  54.112  1.00  3.60
ATOM   1359  ND2 ASN A  84      38.825  69.669  55.711  1.00  3.60
ATOM   1360 1HD2 ASN A  84      38.048  69.213  56.161  1.00  3.60
ATOM   1361 2HD2 ASN A  84      39.308  70.407  56.205  1.00  3.60
ATOM   1362  C   ASN A  84      37.495  65.888  53.528  1.00  3.60
ATOM   1363  O   ASN A  84      37.825  65.831  52.339  1.00  3.60
ATOM   1364  N   ARG A  85      36.852  64.888  54.141  1.00  3.36
ATOM   1365  H   ARG A  85      36.657  65.025  55.124  1.00  3.36
ATOM   1366  CA  ARG A  85      36.320  63.678  53.491  1.00  3.36
ATOM   1367  HA  ARG A  85      37.154  63.000  53.317  1.00  3.36
ATOM   1368  CB  ARG A  85      35.336  63.006  54.476  1.00  3.36
ATOM   1369  HB1 ARG A  85      35.878  62.795  55.400  1.00  3.36
ATOM   1370  HB2 ARG A  85      35.010  62.052  54.060  1.00  3.36
ATOM   1371  CG  ARG A  85      34.087  63.859  54.810  1.00  3.36
ATOM   1372  HG1 ARG A  85      33.409  63.828  53.958  1.00  3.36
ATOM   1373  HG2 ARG A  85      34.366  64.899  54.980  1.00  3.36
ATOM   1374  CD  ARG A  85      33.311  63.364  56.034  1.00  3.36
ATOM   1375  HD1 ARG A  85      33.132  62.293  55.925  1.00  3.36
```

```
ATOM   1376  HD2 ARG A  85      32.347  63.877  56.057  1.00  3.36
ATOM   1377  NE  ARG A  85      34.019  63.650  57.297  1.00  3.36
ATOM   1378  HE  ARG A  85      34.591  64.498  57.336  1.00  3.36
ATOM   1379  CZ  ARG A  85      33.880  63.013  58.442  1.00  3.36
ATOM   1380  NH1 ARG A  85      34.605  63.333  59.471  1.00  3.36
ATOM   1381 1HH1 ARG A  85      35.222  64.146  59.383  1.00  3.36
ATOM   1382 2HH1 ARG A  85      34.462  62.921  60.368  1.00  3.36
ATOM   1383  NH2 ARG A  85      33.008  62.051  58.589  1.00  3.36
ATOM   1384 1HH2 ARG A  85      32.388  61.845  57.829  1.00  3.36
ATOM   1385 2HH2 ARG A  85      32.873  61.613  59.479  1.00  3.36
ATOM   1386  C   ARG A  85     35.685  64.002  52.130  1.00  3.36
ATOM   1387  O   ARG A  85     34.807  64.865  52.086  1.00  3.36
ATOM   1388  N   ARG A  86     36.177  63.358  51.060  1.00  3.32
ATOM   1389  H   ARG A  86     36.883  62.660  51.285  1.00  3.32
ATOM   1390  CA  ARG A  86      35.969  63.561  49.595  1.00  3.32
ATOM   1391  HA  ARG A  86      35.790  62.583  49.147  1.00  3.32
ATOM   1392  CB  ARG A  86      34.795  64.482  49.199  1.00  3.32
ATOM   1393  HB1 ARG A  86      34.843  64.671  48.125  1.00  3.32
ATOM   1394  HB2 ARG A  86      34.930  65.450  49.681  1.00  3.32
ATOM   1395  CG  ARG A  86      33.397  63.889  49.482  1.00  3.32
ATOM   1396  HG1 ARG A  86      33.406  63.329  50.417  1.00  3.32
ATOM   1397  HG2 ARG A  86      33.137  63.196  48.681  1.00  3.32
ATOM   1398  CD  ARG A  86      32.324  64.988  49.575  1.00  3.32
ATOM   1399  HD1 ARG A  86      31.345  64.518  49.692  1.00  3.32
ATOM   1400  HD2 ARG A  86      32.333  65.576  48.655  1.00  3.32
ATOM   1401  NE  ARG A  86      32.607  65.845  50.738  1.00  3.32
ATOM   1402  HE  ARG A  86      33.349  65.507  51.339  1.00  3.32
ATOM   1403  CZ  ARG A  86      32.130  67.022  51.069  1.00  3.32
ATOM   1404  NH1 ARG A  86      32.664  67.654  52.072  1.00  3.32
ATOM   1405 1HH1 ARG A  86      33.468  67.243  52.522  1.00  3.32
ATOM   1406 2HH1 ARG A  86      32.302  68.537  52.384  1.00  3.32
ATOM   1407  NH2 ARG A  86      31.149  67.578  50.413  1.00  3.32
ATOM   1408 1HH2 ARG A  86      30.766  67.081  49.630  1.00  3.32
ATOM   1409 2HH2 ARG A  86      30.800  68.488  50.662  1.00  3.32
ATOM   1410  C   ARG A  86     37.246  64.071  48.910  1.00  3.32
ATOM   1411  O   ARG A  86     37.387  63.962  47.691  1.00  3.32
ATOM   1412  N   LEU A  87     38.191  64.622  49.672  1.00  3.23
ATOM   1413  H   LEU A  87     37.998  64.766  50.658  1.00  3.23
ATOM   1414  CA  LEU A  87      39.561  64.835  49.218  1.00  3.23
ATOM   1415  HA  LEU A  87      39.547  65.138  48.175  1.00  3.23
ATOM   1416  CB  LEU A  87      40.210  65.974  50.051  1.00  3.23
ATOM   1417  HB1 LEU A  87      41.255  66.078  49.768  1.00  3.23
ATOM   1418  HB2 LEU A  87      40.201  65.689  51.106  1.00  3.23
ATOM   1419  CG  LEU A  87      39.571  67.374  49.919  1.00  3.23
ATOM   1420  HG  LEU A  87      38.616  67.392  50.441  1.00  3.23
ATOM   1421  CD1 LEU A  87      40.494  68.433  50.531  1.00  3.23
ATOM   1422 1HD1 LEU A  87      39.974  69.390  50.575  1.00  3.23
ATOM   1423 2HD1 LEU A  87      40.780  68.140  51.537  1.00  3.23
ATOM   1424 3HD1 LEU A  87      41.402  68.540  49.936  1.00  3.23
ATOM   1425  CD2 LEU A  87      39.330  67.800  48.468  1.00  3.23
ATOM   1426 1HD2 LEU A  87      38.924  68.813  48.454  1.00  3.23
ATOM   1427 2HD2 LEU A  87      40.266  67.792  47.917  1.00  3.23
ATOM   1428 3HD2 LEU A  87      38.595  67.146  48.002  1.00  3.23
```

```
ATOM   1429  C   LEU A  87     40.387  63.526  49.275  1.00  3.23
ATOM   1430  O   LEU A  87     39.880  62.432  49.510  1.00  3.23
ATOM   1431  N   SER A  88     41.696  63.637  49.068  1.00  3.05
ATOM   1432  H   SER A  88     42.065  64.548  48.840  1.00  3.05
ATOM   1433  CA  SER A  88     42.690  62.710  49.606  1.00  3.05
ATOM   1434  HA  SER A  88     42.317  62.336  50.556  1.00  3.05
ATOM   1435  CB  SER A  88     42.937  61.526  48.667  1.00  3.05
ATOM   1436  HB1 SER A  88     43.907  61.078  48.887  1.00  3.05
ATOM   1437  HB2 SER A  88     42.931  61.863  47.630  1.00  3.05
ATOM   1438  OG  SER A  88     41.938  60.545  48.873  1.00  3.05
ATOM   1439  HG  SER A  88     41.085  61.013  48.938  1.00  3.05
ATOM   1440  C   SER A  88     43.990  63.469  49.889  1.00  3.05
ATOM   1441  O   SER A  88     44.248  64.467  49.216  1.00  3.05
ATOM   1442  N   TYR A  89     44.778  63.041  50.883  1.00  2.52
ATOM   1443  H   TYR A  89     44.467  62.222  51.393  1.00  2.52
ATOM   1444  CA  TYR A  89     45.613  64.011  51.631  1.00  2.52
ATOM   1445  HA  TYR A  89     45.640  64.951  51.083  1.00  2.52
ATOM   1446  CB  TYR A  89     44.872  64.285  52.971  1.00  2.52
ATOM   1447  HB1 TYR A  89     45.514  64.043  53.819  1.00  2.52
ATOM   1448  HB2 TYR A  89     44.014  63.617  53.061  1.00  2.52
ATOM   1449  CG  TYR A  89     44.381  65.708  53.168  1.00  2.52
ATOM   1450  CD1 TYR A  89     43.022  65.963  53.442  1.00  2.52
ATOM   1451  HD1 TYR A  89     42.318  65.147  53.527  1.00  2.52
ATOM   1452  CE1 TYR A  89     42.570  67.283  53.632  1.00  2.52
ATOM   1453  HE1 TYR A  89     41.534  67.467  53.859  1.00  2.52
ATOM   1454  CZ  TYR A  89     43.476  68.361  53.546  1.00  2.52
ATOM   1455  OH  TYR A  89     43.046  69.650  53.642  1.00  2.52
ATOM   1456  HH  TYR A  89     42.110  69.694  53.894  1.00  2.52
ATOM   1457  CE2 TYR A  89     44.847  68.094  53.347  1.00  2.52
ATOM   1458  HE2 TYR A  89     45.551  68.906  53.392  1.00  2.52
ATOM   1459  CD2 TYR A  89     45.297  66.775  53.151  1.00  2.52
ATOM   1460  HD2 TYR A  89     46.352  66.576  53.014  1.00  2.52
ATOM   1461  C   TYR A  89     47.101  63.642  51.876  1.00  2.52
ATOM   1462  O   TYR A  89     47.710  64.212  52.780  1.00  2.52
ATOM   1463  N   SER A  90     47.665  62.700  51.100  1.00  2.40
ATOM   1464  H   SER A  90     47.101  62.417  50.315  1.00  2.40
ATOM   1465  CA  SER A  90     48.924  61.936  51.330  1.00  2.40
ATOM   1466  HA  SER A  90     48.607  61.082  51.916  1.00  2.40
ATOM   1467  CB  SER A  90     49.475  61.382  50.008  1.00  2.40
ATOM   1468  HB1 SER A  90     49.891  62.203  49.420  1.00  2.40
ATOM   1469  HB2 SER A  90     48.657  60.930  49.444  1.00  2.40
ATOM   1470  OG  SER A  90     50.477  60.396  50.219  1.00  2.40
ATOM   1471  HG  SER A  90     50.163  59.746  50.866  1.00  2.40
ATOM   1472  C   SER A  90     50.054  62.606  52.145  1.00  2.40
ATOM   1473  O   SER A  90     50.351  63.775  51.980  1.00  2.40
ATOM   1474  N   VAL A  91     50.706  61.815  53.000  1.00  2.05
ATOM   1475  H   VAL A  91     50.568  60.824  52.876  1.00  2.05
ATOM   1476  CA  VAL A  91     51.274  62.179  54.298  1.00  2.05
ATOM   1477  HA  VAL A  91     51.352  63.245  54.407  1.00  2.05
ATOM   1478  CB  VAL A  91     50.309  61.744  55.421  1.00  2.05
ATOM   1479  HB  VAL A  91     50.194  60.662  55.400  1.00  2.05
ATOM   1480  CG1 VAL A  91     50.822  62.176  56.792  1.00  2.05
ATOM   1481 1HG1 VAL A  91     50.960  63.256  56.811  1.00  2.05
```

```
ATOM   1482 2HG1 VAL A  91      50.113  61.887  57.566  1.00  2.05
ATOM   1483 3HG1 VAL A  91      51.776  61.705  57.000  1.00  2.05
ATOM   1484  CG2 VAL A  91      48.916  62.361  55.263  1.00  2.05
ATOM   1485 1HG2 VAL A  91      48.303  62.128  56.132  1.00  2.05
ATOM   1486 2HG2 VAL A  91      48.993  63.443  55.163  1.00  2.05
ATOM   1487 3HG2 VAL A  91      48.418  61.961  54.384  1.00  2.05
ATOM   1488  C   VAL A  91     52.637  61.485  54.399  1.00  2.05
ATOM   1489  O   VAL A  91     52.716  60.378  54.914  1.00  2.05
ATOM   1490  N   THR A  92     53.720  62.030  53.842  1.00  1.94
ATOM   1491  H   THR A  92     53.626  62.949  53.424  1.00  1.94
ATOM   1492  CA  THR A  92      55.053  61.421  54.101  1.00  1.94
ATOM   1493  HA  THR A  92      54.973  60.365  53.880  1.00  1.94
ATOM   1494  CB  THR A  92      56.199  61.953  53.216  1.00  1.94
ATOM   1495  HB  THR A  92      56.623  62.853  53.657  1.00  1.94
ATOM   1496  CG2 THR A  92      57.326  60.939  52.998  1.00  1.94
ATOM   1497 1HG2 THR A  92      58.013  61.314  52.239  1.00  1.94
ATOM   1498 2HG2 THR A  92      57.895  60.789  53.914  1.00  1.94
ATOM   1499 3HG2 THR A  92      56.920  59.985  52.663  1.00  1.94
ATOM   1500  OG1 THR A  92      55.725  62.247  51.920  1.00  1.94
ATOM   1501  HG1 THR A  92      55.139  63.004  51.989  1.00  1.94
ATOM   1502  C   THR A  92     55.415  61.508  55.609  1.00  1.94
ATOM   1503  O   THR A  92     54.740  62.220  56.343  1.00  1.94
ATOM   1504  N   TRP A  93     56.437  60.798  56.101  1.00  1.88
ATOM   1505  H   TRP A  93     56.885  60.154  55.463  1.00  1.88
ATOM   1506  CA  TRP A  93      56.822  60.655  57.508  1.00  1.88
ATOM   1507  HA  TRP A  93      56.899  61.638  57.920  1.00  1.88
ATOM   1508  CB  TRP A  93      55.768  60.083  58.463  1.00  1.88
ATOM   1509  HB1 TRP A  93      55.185  60.939  58.785  1.00  1.88
ATOM   1510  HB2 TRP A  93      56.288  59.715  59.338  1.00  1.88
ATOM   1511  CG  TRP A  93      54.754  59.056  58.106  1.00  1.88
ATOM   1512  CD1 TRP A  93      53.596  59.337  57.487  1.00  1.88
ATOM   1513  HD1 TRP A  93      53.351  60.324  57.137  1.00  1.88
ATOM   1514  NE1 TRP A  93      52.761  58.244  57.508  1.00  1.88
ATOM   1515  HE1 TRP A  93      51.805  58.280  57.190  1.00  1.88
ATOM   1516  CE2 TRP A  93      53.300  57.232  58.265  1.00  1.88
ATOM   1517  CZ2 TRP A  93      52.814  55.987  58.682  1.00  1.88
ATOM   1518  HZ2 TRP A  93      51.857  55.623  58.350  1.00  1.88
ATOM   1519  CH2 TRP A  93      53.584  55.221  59.568  1.00  1.88
ATOM   1520  HH2 TRP A  93      53.236  54.246  59.867  1.00  1.88
ATOM   1521  CZ3 TRP A  93      54.813  55.712  60.038  1.00  1.88
ATOM   1522  HZ3 TRP A  93      55.401  55.126  60.729  1.00  1.88
ATOM   1523  CE3 TRP A  93      55.302  56.953  59.591  1.00  1.88
ATOM   1524  HE3 TRP A  93      56.252  57.324  59.943  1.00  1.88
ATOM   1525  CD2 TRP A  93      54.570  57.732  58.677  1.00  1.88
ATOM   1526  C   TRP A  93     58.216  60.023  57.679  1.00  1.88
ATOM   1527  O   TRP A  93     58.382  58.817  57.847  1.00  1.88
ATOM   1528  N   PHE A  94     59.253  60.856  57.603  1.00  1.89
ATOM   1529  H   PHE A  94     59.056  61.826  57.395  1.00  1.89
ATOM   1530  CA  PHE A  94      60.655  60.451  57.697  1.00  1.89
ATOM   1531  HA  PHE A  94      60.761  59.494  57.191  1.00  1.89
ATOM   1532  CB  PHE A  94      61.556  61.436  56.919  1.00  1.89
ATOM   1533  HB1 PHE A  94      62.573  61.045  56.944  1.00  1.89
ATOM   1534  HB2 PHE A  94      61.587  62.386  57.442  1.00  1.89
```

```
ATOM   1535  CG  PHE A  94      61.193  61.742  55.466  1.00  1.89
ATOM   1536  CD1 PHE A  94      60.222  62.718  55.169  1.00  1.89
ATOM   1537  HD1 PHE A  94      59.717  63.226  55.965  1.00  1.89
ATOM   1538  CE1 PHE A  94      59.933  63.077  53.843  1.00  1.89
ATOM   1539  HE1 PHE A  94      59.164  63.803  53.631  1.00  1.89
ATOM   1540  CZ  PHE A  94      60.673  62.521  52.792  1.00  1.89
ATOM   1541  HZ  PHE A  94      60.487  62.832  51.773  1.00  1.89
ATOM   1542  CE2 PHE A  94      61.684  61.591  53.076  1.00  1.89
ATOM   1543  HE2 PHE A  94      62.279  61.208  52.262  1.00  1.89
ATOM   1544  CD2 PHE A  94      61.920  61.173  54.401  1.00  1.89
ATOM   1545  HD2 PHE A  94      62.694  60.451  54.608  1.00  1.89
ATOM   1546  C   PHE A  94      61.113  60.235  59.166  1.00  1.89
ATOM   1547  O   PHE A  94      61.502  61.192  59.817  1.00  1.89
ATOM   1548  N   CYS A  95      60.992  59.005  59.681  1.00  1.89
ATOM   1549  H   CYS A  95      60.616  58.330  59.028  1.00  1.89
ATOM   1550  CA  CYS A  95      61.265  58.418  61.001  1.00  1.89
ATOM   1551  HA  CYS A  95      60.730  58.978  61.754  1.00  1.89
ATOM   1552  CB  CYS A  95      60.705  56.984  61.013  1.00  1.89
ATOM   1553  HB1 CYS A  95      60.748  56.591  62.032  1.00  1.89
ATOM   1554  HB2 CYS A  95      61.326  56.356  60.377  1.00  1.89
ATOM   1555  SG  CYS A  95      58.990  56.864  60.429  1.00  1.89
ATOM   1556  HG  CYS A  95      59.150  57.419  59.214  1.00  1.89
ATOM   1557  C   CYS A  95      62.760  58.295  61.347  1.00  1.89
ATOM   1558  O   CYS A  95      63.599  57.988  60.502  1.00  1.89
ATOM   1559  N   SER A  96      63.129  58.322  62.629  1.00  1.83
ATOM   1560  H   SER A  96      62.461  58.619  63.327  1.00  1.83
ATOM   1561  CA  SER A  96      64.446  57.799  63.014  1.00  1.83
ATOM   1562  HA  SER A  96      65.208  58.156  62.317  1.00  1.83
ATOM   1563  CB  SER A  96      64.835  58.266  64.415  1.00  1.83
ATOM   1564  HB1 SER A  96      64.048  57.960  65.093  1.00  1.83
ATOM   1565  HB2 SER A  96      64.938  59.348  64.433  1.00  1.83
ATOM   1566  OG  SER A  96      66.037  57.682  64.866  1.00  1.83
ATOM   1567  HG  SER A  96      66.771  58.183  64.474  1.00  1.83
ATOM   1568  C   SER A  96      64.467  56.272  63.004  1.00  1.83
ATOM   1569  O   SER A  96      65.373  55.652  62.445  1.00  1.83
ATOM   1570  N   TRP A  97      63.445  55.670  63.605  1.00  2.16
ATOM   1571  H   TRP A  97      62.710  56.236  64.004  1.00  2.16
ATOM   1572  CA  TRP A  97      63.290  54.235  63.772  1.00  2.16
ATOM   1573  HA  TRP A  97      64.080  53.685  63.262  1.00  2.16
ATOM   1574  CB  TRP A  97      63.294  53.893  65.269  1.00  2.16
ATOM   1575  HB1 TRP A  97      62.792  52.935  65.401  1.00  2.16
ATOM   1576  HB2 TRP A  97      62.684  54.626  65.800  1.00  2.16
ATOM   1577  CG  TRP A  97      64.614  53.763  65.958  1.00  2.16
ATOM   1578  CD1 TRP A  97      64.965  54.419  67.083  1.00  2.16
ATOM   1579  HD1 TRP A  97      64.334  55.144  67.586  1.00  2.16
ATOM   1580  NE1 TRP A  97      66.162  53.928  67.560  1.00  2.16
ATOM   1581  HE1 TRP A  97      66.515  54.170  68.472  1.00  2.16
ATOM   1582  CE2 TRP A  97      66.631  52.899  66.771  1.00  2.16
ATOM   1583  CZ2 TRP A  97      67.735  52.042  66.871  1.00  2.16
ATOM   1584  HZ2 TRP A  97      68.423  52.129  67.694  1.00  2.16
ATOM   1585  CH2 TRP A  97      67.924  51.052  65.892  1.00  2.16
ATOM   1586  HH2 TRP A  97      68.762  50.373  65.957  1.00  2.16
ATOM   1587  CZ3 TRP A  97      67.009  50.928  64.833  1.00  2.16
```

```
ATOM   1588  HZ3 TRP A  97      67.149  50.151  64.092  1.00  2.16
ATOM   1589  CE3 TRP A  97      65.897  51.787  64.749  1.00  2.16
ATOM   1590  HE3 TRP A  97      65.185  51.669  63.948  1.00  2.16
ATOM   1591  CD2 TRP A  97      65.677  52.793  65.714  1.00  2.16
ATOM   1592  C   TRP A  97     61.934  53.810  63.223  1.00  2.16
ATOM   1593  O   TRP A  97     60.935  54.496  63.433  1.00  2.16
ATOM   1594  N   SER A  98     61.903  52.684  62.518  1.00  2.32
ATOM   1595  H   SER A  98     62.777  52.220  62.304  1.00  2.32
ATOM   1596  CA  SER A  98     60.650  52.101  62.044  1.00  2.32
ATOM   1597  HA  SER A  98     60.157  52.865  61.446  1.00  2.32
ATOM   1598  CB  SER A  98     60.947  50.895  61.157  1.00  2.32
ATOM   1599  HB1 SER A  98     61.574  51.205  60.324  1.00  2.32
ATOM   1600  HB2 SER A  98     60.013  50.495  60.767  1.00  2.32
ATOM   1601  OG  SER A  98     61.621  49.897  61.892  1.00  2.32
ATOM   1602  HG  SER A  98     61.118  49.702  62.696  1.00  2.32
ATOM   1603  C   SER A  98     59.705  51.682  63.191  1.00  2.32
ATOM   1604  O   SER A  98     60.181  51.252  64.240  1.00  2.32
ATOM   1605  N   PRO A  99     58.372  51.731  63.004  1.00  2.38
ATOM   1606  CD  PRO A  99     57.684  51.847  61.717  1.00  2.38
ATOM   1607  HD1 PRO A  99      57.531  50.849  61.304  1.00  2.38
ATOM   1608  HD2 PRO A  99      58.218  52.466  60.997  1.00  2.38
ATOM   1609  CG  PRO A  99     56.331  52.487  62.014  1.00  2.38
ATOM   1610  HG1 PRO A  99      55.562  52.171  61.310  1.00  2.38
ATOM   1611  HG2 PRO A  99      56.437  53.572  62.009  1.00  2.38
ATOM   1612  CB  PRO A  99     56.044  52.018  63.433  1.00  2.38
ATOM   1613  HB1 PRO A  99     55.674  50.990  63.418  1.00  2.38
ATOM   1614  HB2 PRO A  99     55.334  52.676  63.934  1.00  2.38
ATOM   1615  CA  PRO A  99     57.432  52.079  64.076  1.00  2.38
ATOM   1616  HA  PRO A  99     57.628  53.127  64.311  1.00  2.38
ATOM   1617  C   PRO A  99     57.411  51.358  65.435  1.00  2.38
ATOM   1618  O   PRO A  99     57.066  51.984  66.431  1.00  2.38
ATOM   1619  N   CYS A 100     57.683  50.055  65.466  1.00  2.46
ATOM   1620  H   CYS A 100     58.033  49.669  64.605  1.00  2.46
ATOM   1621  CA  CYS A 100      57.343  49.075  66.513  1.00  2.46
ATOM   1622  HA  CYS A 100      58.091  48.290  66.387  1.00  2.46
ATOM   1623  CB  CYS A 100      57.539  49.571  67.961  1.00  2.46
ATOM   1624  HB1 CYS A 100      58.319  50.334  67.989  1.00  2.46
ATOM   1625  HB2 CYS A 100      57.861  48.736  68.582  1.00  2.46
ATOM   1626  SG  CYS A 100     56.000  50.228  68.668  1.00  2.46
ATOM   1627  HG  CYS A 100     56.029  51.368  67.950  1.00  2.46
ATOM   1628  C   CYS A 100     55.996  48.375  66.263  1.00  2.46
ATOM   1629  O   CYS A 100     55.179  48.855  65.483  1.00  2.46
ATOM   1630  N   ALA A 101     55.767  47.229  66.913  1.00  2.67
ATOM   1631  H   ALA A 101     56.457  46.894  67.566  1.00  2.67
ATOM   1632  CA  ALA A 101      54.581  46.401  66.669  1.00  2.67
ATOM   1633  HA  ALA A 101      54.556  46.131  65.612  1.00  2.67
ATOM   1634  CB  ALA A 101      54.739  45.109  67.480  1.00  2.67
ATOM   1635  HB1 ALA A 101      53.901  44.443  67.269  1.00  2.67
ATOM   1636  HB2 ALA A 101      55.661  44.604  67.192  1.00  2.67
ATOM   1637  HB3 ALA A 101      54.757  45.328  68.549  1.00  2.67
ATOM   1638  C   ALA A 101     53.247  47.097  66.986  1.00  2.67
ATOM   1639  O   ALA A 101     52.285  47.045  66.214  1.00  2.67
ATOM   1640  N   ASN A 102     53.220  47.783  68.124  1.00  2.54
```

```
ATOM   1641  H   ASN A 102      54.025  47.690  68.734  1.00  2.54
ATOM   1642  CA  ASN A 102      52.063  48.481  68.665  1.00  2.54
ATOM   1643  HA  ASN A 102      51.209  47.802  68.682  1.00  2.54
ATOM   1644  CB  ASN A 102      52.398  48.937  70.106  1.00  2.54
ATOM   1645  HB1 ASN A 102      51.484  49.317  70.561  1.00  2.54
ATOM   1646  HB2 ASN A 102      53.112  49.760  70.076  1.00  2.54
ATOM   1647  CG  ASN A 102      52.999  47.869  71.019  1.00  2.54
ATOM   1648  OD1 ASN A 102      53.898  47.119  70.669  1.00  2.54
ATOM   1649  ND2 ASN A 102      52.555  47.782  72.247  1.00  2.54
ATOM   1650 1HD2 ASN A 102      51.837  48.394  72.584  1.00  2.54
ATOM   1651 2HD2 ASN A 102      52.994  47.092  72.831  1.00  2.54
ATOM   1652  C   ASN A 102      51.709  49.682  67.782  1.00  2.54
ATOM   1653  O   ASN A 102      50.561  49.826  67.358  1.00  2.54
ATOM   1654  N   CYS A 103      52.712  50.501  67.453  1.00  2.33
ATOM   1655  H   CYS A 103      53.628  50.344  67.850  1.00  2.33
ATOM   1656  CA  CYS A 103      52.522  51.660  66.592  1.00  2.33
ATOM   1657  HA  CYS A 103      51.733  52.283  67.011  1.00  2.33
ATOM   1658  CB  CYS A 103      53.810  52.488  66.543  1.00  2.33
ATOM   1659  HB1 CYS A 103      53.689  53.297  65.823  1.00  2.33
ATOM   1660  HB2 CYS A 103      54.639  51.858  66.222  1.00  2.33
ATOM   1661  SG  CYS A 103      54.171  53.206  68.167  1.00  2.33
ATOM   1662  HG  CYS A 103      55.224  53.959  67.788  1.00  2.33
ATOM   1663  C   CYS A 103      52.112  51.279  65.179  1.00  2.33
ATOM   1664  O   CYS A 103      51.241  51.924  64.605  1.00  2.33
ATOM   1665  N   ALA A 104      52.699  50.214  64.636  1.00  2.46
ATOM   1666  H   ALA A 104      53.447  49.755  65.145  1.00  2.46
ATOM   1667  CA  ALA A 104      52.296  49.629  63.370  1.00  2.46
ATOM   1668  HA  ALA A 104      52.440  50.369  62.583  1.00  2.46
ATOM   1669  CB  ALA A 104      53.191  48.419  63.082  1.00  2.46
ATOM   1670  HB1 ALA A 104      52.843  47.918  62.182  1.00  2.46
ATOM   1671  HB2 ALA A 104      54.223  48.739  62.942  1.00  2.46
ATOM   1672  HB3 ALA A 104      53.146  47.705  63.901  1.00  2.46
ATOM   1673  C   ALA A 104      50.823  49.248  63.381  1.00  2.46
ATOM   1674  O   ALA A 104      50.081  49.688  62.513  1.00  2.46
ATOM   1675  N   THR A 105      50.376  48.540  64.418  1.00  2.57
ATOM   1676  H   THR A 105      51.052  48.211  65.099  1.00  2.57
ATOM   1677  CA  THR A 105      48.964  48.173  64.577  1.00  2.57
ATOM   1678  HA  THR A 105      48.664  47.562  63.726  1.00  2.57
ATOM   1679  CB  THR A 105      48.791  47.325  65.850  1.00  2.57
ATOM   1680  HB  THR A 105      49.151  47.880  66.714  1.00  2.57
ATOM   1681  CG2 THR A 105      47.345  46.902  66.122  1.00  2.57
ATOM   1682 1HG2 THR A 105      47.331  46.201  66.957  1.00  2.57
ATOM   1683 2HG2 THR A 105      46.748  47.774  66.382  1.00  2.57
ATOM   1684 3HG2 THR A 105      46.935  46.417  65.235  1.00  2.57
ATOM   1685  OG1 THR A 105      49.540  46.135  65.738  1.00  2.57
ATOM   1686  HG1 THR A 105      50.484  46.340  65.822  1.00  2.57
ATOM   1687  C   THR A 105      48.042  49.387  64.622  1.00  2.57
ATOM   1688  O   THR A 105      47.000  49.381  63.966  1.00  2.57
ATOM   1689  N   THR A 106      48.394  50.447  65.351  1.00  2.48
ATOM   1690  H   THR A 106      49.248  50.432  65.904  1.00  2.48
ATOM   1691  CA  THR A 106      47.556  51.656  65.382  1.00  2.48
ATOM   1692  HA  THR A 106      46.523  51.348  65.545  1.00  2.48
ATOM   1693  CB  THR A 106      47.926  52.587  66.546  1.00  2.48
```

```
ATOM   1694  HB  THR A 106      47.429  53.548  66.416  1.00  2.48
ATOM   1695  CG2 THR A 106      47.494  51.992  67.887  1.00  2.48
ATOM   1696 1HG2 THR A 106      47.781  52.674  68.688  1.00  2.48
ATOM   1697 2HG2 THR A 106      46.412  51.862  67.903  1.00  2.48
ATOM   1698 3HG2 THR A 106      47.985  51.033  68.050  1.00  2.48
ATOM   1699  OG1 THR A 106      49.316  52.782  66.627  1.00  2.48
ATOM   1700  HG1 THR A 106      49.644  53.110  65.782  1.00  2.48
ATOM   1701  C   THR A 106     47.547  52.446  64.075  1.00  2.48
ATOM   1702  O   THR A 106     46.516  53.007  63.720  1.00  2.48
ATOM   1703  N   LEU A 107     48.657  52.460  63.337  1.00  2.42
ATOM   1704  H   LEU A 107     49.466  51.957  63.688  1.00  2.42
ATOM   1705  CA  LEU A 107      48.802  53.142  62.048  1.00  2.42
ATOM   1706  HA  LEU A 107      48.389  54.146  62.137  1.00  2.42
ATOM   1707  CB  LEU A 107      50.304  53.248  61.719  1.00  2.42
ATOM   1708  HB1 LEU A 107      50.433  53.374  60.642  1.00  2.42
ATOM   1709  HB2 LEU A 107      50.779  52.312  62.006  1.00  2.42
ATOM   1710  CG  LEU A 107      51.008  54.417  62.443  1.00  2.42
ATOM   1711  HG  LEU A 107      50.604  54.512  63.450  1.00  2.42
ATOM   1712  CD1 LEU A 107      52.520  54.187  62.574  1.00  2.42
ATOM   1713 1HD1 LEU A 107      52.874  53.493  61.817  1.00  2.42
ATOM   1714 2HD1 LEU A 107      53.061  55.130  62.483  1.00  2.42
ATOM   1715 3HD1 LEU A 107      52.737  53.762  63.551  1.00  2.42
ATOM   1716  CD2 LEU A 107      50.752  55.731  61.705  1.00  2.42
ATOM   1717 1HD2 LEU A 107      51.221  55.712  60.730  1.00  2.42
ATOM   1718 2HD2 LEU A 107      49.689  55.893  61.567  1.00  2.42
ATOM   1719 3HD2 LEU A 107      51.160  56.559  62.287  1.00  2.42
ATOM   1720  C   LEU A 107     47.991  52.455  60.939  1.00  2.42
ATOM   1721  O   LEU A 107     47.259  53.098  60.183  1.00  2.42
ATOM   1722  N   THR A 108     48.057  51.129  60.945  1.00  2.63
ATOM   1723  H   THR A 108     48.760  50.715  61.550  1.00  2.63
ATOM   1724  CA  THR A 108      47.234  50.186  60.191  1.00  2.63
ATOM   1725  HA  THR A 108      47.418  50.325  59.125  1.00  2.63
ATOM   1726  CB  THR A 108      47.689  48.767  60.573  1.00  2.63
ATOM   1727  HB  THR A 108      47.916  48.728  61.637  1.00  2.63
ATOM   1728  CG2 THR A 108      46.671  47.677  60.292  1.00  2.63
ATOM   1729 1HG2 THR A 108      46.241  47.819  59.305  1.00  2.63
ATOM   1730 2HG2 THR A 108      47.168  46.712  60.347  1.00  2.63
ATOM   1731 3HG2 THR A 108      45.884  47.698  61.043  1.00  2.63
ATOM   1732  OG1 THR A 108      48.861  48.479  59.853  1.00  2.63
ATOM   1733  HG1 THR A 108      49.117  47.567  60.025  1.00  2.63
ATOM   1734  C   THR A 108     45.737  50.378  60.429  1.00  2.63
ATOM   1735  O   THR A 108     44.971  50.516  59.467  1.00  2.63
ATOM   1736  N   ARG A 109     45.310  50.406  61.703  1.00  2.73
ATOM   1737  H   ARG A 109     45.994  50.264  62.441  1.00  2.73
ATOM   1738  CA  ARG A 109      43.914  50.695  62.072  1.00  2.73
ATOM   1739  HA  ARG A 109      43.255  49.989  61.566  1.00  2.73
ATOM   1740  CB  ARG A 109      43.697  50.573  63.596  1.00  2.73
ATOM   1741  HB1 ARG A 109      42.728  51.012  63.841  1.00  2.73
ATOM   1742  HB2 ARG A 109      44.465  51.149  64.114  1.00  2.73
ATOM   1743  CG  ARG A 109      43.699  49.122  64.121  1.00  2.73
ATOM   1744  HG1 ARG A 109      44.638  48.640  63.858  1.00  2.73
ATOM   1745  HG2 ARG A 109      42.889  48.567  63.646  1.00  2.73
ATOM   1746  CD  ARG A 109      43.508  49.074  65.651  1.00  2.73
```

```
ATOM   1747  HD1 ARG A 109      42.488  49.388  65.880  1.00  2.73
ATOM   1748  HD2 ARG A 109      44.200  49.783  66.111  1.00  2.73
ATOM   1749  NE  ARG A 109      43.773  47.726  66.210  1.00  2.73
ATOM   1750  HE  ARG A 109      44.383  47.139  65.665  1.00  2.73
ATOM   1751  CZ  ARG A 109      43.377  47.235  67.378  1.00  2.73
ATOM   1752  NH1 ARG A 109      43.786  46.056  67.770  1.00  2.73
ATOM   1753 1HH1 ARG A 109      44.406  45.520  67.187  1.00  2.73
ATOM   1754 2HH1 ARG A 109      43.521  45.691  68.668  1.00  2.73
ATOM   1755  NH2 ARG A 109      42.586  47.894  68.182  1.00  2.73
ATOM   1756 1HH2 ARG A 109      42.233  48.792  67.902  1.00  2.73
ATOM   1757 2HH2 ARG A 109      42.295  47.498  69.057  1.00  2.73
ATOM   1758  C   ARG A 109     43.482  52.074  61.562  1.00  2.73
ATOM   1759  O   ARG A 109     42.408  52.200  60.981  1.00  2.73
ATOM   1760  N   PHE A 110     44.338  53.082  61.695  1.00  2.78
ATOM   1761  H   PHE A 110     45.194  52.943  62.218  1.00  2.78
ATOM   1762  CA  PHE A 110      44.035  54.419  61.224  1.00  2.78
ATOM   1763  HA  PHE A 110      43.095  54.710  61.696  1.00  2.78
ATOM   1764  CB  PHE A 110      45.104  55.405  61.697  1.00  2.78
ATOM   1765  HB1 PHE A 110      45.959  55.378  61.021  1.00  2.78
ATOM   1766  HB2 PHE A 110      45.459  55.103  62.682  1.00  2.78
ATOM   1767  CG  PHE A 110      44.592  56.823  61.834  1.00  2.78
ATOM   1768  CD1 PHE A 110      43.892  57.198  62.998  1.00  2.78
ATOM   1769  HD1 PHE A 110      43.690  56.467  63.767  1.00  2.78
ATOM   1770  CE1 PHE A 110      43.500  58.534  63.190  1.00  2.78
ATOM   1771  HE1 PHE A 110      43.010  58.831  64.107  1.00  2.78
ATOM   1772  CZ  PHE A 110     43.797  59.498  62.215  1.00  2.78
ATOM   1773  HZ  PHE A 110     43.528  60.530  62.387  1.00  2.78
ATOM   1774  CE2 PHE A 110      44.476  59.122  61.043  1.00  2.78
ATOM   1775  HE2 PHE A 110      44.726  59.871  60.313  1.00  2.78
ATOM   1776  CD2 PHE A 110      44.873  57.786  60.847  1.00  2.78
ATOM   1777  HD2 PHE A 110      45.429  57.505  59.965  1.00  2.78
ATOM   1778  C   PHE A 110     43.802  54.506  59.712  1.00  2.78
ATOM   1779  O   PHE A 110     42.845  55.161  59.283  1.00  2.78
ATOM   1780  N   LEU A 111     44.618  53.816  58.906  1.00  2.78
ATOM   1781  H   LEU A 111     45.415  53.335  59.313  1.00  2.78
ATOM   1782  CA  LEU A 111      44.385  53.719  57.463  1.00  2.78
ATOM   1783  HA  LEU A 111      44.299  54.733  57.072  1.00  2.78
ATOM   1784  CB  LEU A 111      45.597  53.047  56.792  1.00  2.78
ATOM   1785  HB1 LEU A 111      45.759  52.066  57.241  1.00  2.78
ATOM   1786  HB2 LEU A 111      46.477  53.655  57.004  1.00  2.78
ATOM   1787  CG  LEU A 111      45.473  52.865  55.265  1.00  2.78
ATOM   1788  HG  LEU A 111      44.762  52.067  55.053  1.00  2.78
ATOM   1789  CD1 LEU A 111      45.036  54.127  54.521  1.00  2.78
ATOM   1790 1HD1 LEU A 111      45.086  53.962  53.445  1.00  2.78
ATOM   1791 2HD1 LEU A 111      44.012  54.391  54.774  1.00  2.78
ATOM   1792 3HD1 LEU A 111      45.693  54.951  54.792  1.00  2.78
ATOM   1793  CD2 LEU A 111      46.833  52.492  54.691  1.00  2.78
ATOM   1794 1HD2 LEU A 111      46.724  52.255  53.633  1.00  2.78
ATOM   1795 2HD2 LEU A 111      47.536  53.315  54.805  1.00  2.78
ATOM   1796 3HD2 LEU A 111      47.214  51.612  55.206  1.00  2.78
ATOM   1797  C   LEU A 111     43.067  53.010  57.137  1.00  2.78
ATOM   1798  O   LEU A 111     42.218  53.561  56.434  1.00  2.78
ATOM   1799  N   ARG A 112      42.857  51.827  57.722  1.00  2.89
```

```
ATOM   1800  H   ARG A 112     43.604  51.458  58.304  1.00  2.89
ATOM   1801  CA  ARG A 112     41.628  51.028  57.565  1.00  2.89
ATOM   1802  HA  ARG A 112     41.420  50.938  56.498  1.00  2.89
ATOM   1803  CB  ARG A 112     41.921  49.591  58.108  1.00  2.89
ATOM   1804  HB1 ARG A 112     40.996  49.013  58.072  1.00  2.89
ATOM   1805  HB2 ARG A 112     42.216  49.673  59.155  1.00  2.89
ATOM   1806  CG  ARG A 112     43.016  48.777  57.340  1.00  2.89
ATOM   1807  HG1 ARG A 112     43.926  49.373  57.270  1.00  2.89
ATOM   1808  HG2 ARG A 112     42.658  48.596  56.326  1.00  2.89
ATOM   1809  CD  ARG A 112     43.389  47.409  57.991  1.00  2.89
ATOM   1810  HD1 ARG A 112     42.469  46.837  58.126  1.00  2.89
ATOM   1811  HD2 ARG A 112     43.805  47.617  58.977  1.00  2.89
ATOM   1812  NE  ARG A 112     44.357  46.581  57.197  1.00  2.89
ATOM   1813  HE  ARG A 112     44.349  46.724  56.200  1.00  2.89
ATOM   1814  CZ  ARG A 112     45.228  45.659  57.637  1.00  2.89
ATOM   1815  NH1 ARG A 112     46.080  45.048  56.859  1.00  2.89
ATOM   1816 1HH1 ARG A 112     46.122  45.228  55.864  1.00  2.89
ATOM   1817 2HH1 ARG A 112     46.786  44.453  57.299  1.00  2.89
ATOM   1818  NH2 ARG A 112     45.301  45.294  58.885  1.00  2.89
ATOM   1819 1HH2 ARG A 112     44.690  45.700  59.558  1.00  2.89
ATOM   1820 2HH2 ARG A 112     46.084  44.715  59.205  1.00  2.89
ATOM   1821  C   ARG A 112    40.352  51.673  58.150  1.00  2.89
ATOM   1822  O   ARG A 112    39.274  51.121  57.952  1.00  2.89
ATOM   1823  N   GLN A 113    40.453  52.835  58.805  1.00  2.90
ATOM   1824  H   GLN A 113    41.389  53.138  59.031  1.00  2.90
ATOM   1825  CA  GLN A 113     39.337  53.677  59.283  1.00  2.90
ATOM   1826  HA  GLN A 113     38.393  53.155  59.123  1.00  2.90
ATOM   1827  CB  GLN A 113     39.506  53.935  60.792  1.00  2.90
ATOM   1828  HB1 GLN A 113     38.793  54.700  61.103  1.00  2.90
ATOM   1829  HB2 GLN A 113     40.511  54.316  60.981  1.00  2.90
ATOM   1830  CG  GLN A 113     39.261  52.688  61.655  1.00  2.90
ATOM   1831  HG1 GLN A 113     39.906  51.872  61.333  1.00  2.90
ATOM   1832  HG2 GLN A 113     38.227  52.366  61.531  1.00  2.90
ATOM   1833  CD  GLN A 113     39.525  52.970  63.131  1.00  2.90
ATOM   1834  OE1 GLN A 113     40.575  52.676  63.684  1.00  2.90
ATOM   1835  NE2 GLN A 113     38.579  53.543  63.845  1.00  2.90
ATOM   1836 1HE2 GLN A 113     37.701  53.795  63.425  1.00  2.90
ATOM   1837 2HE2 GLN A 113     38.786  53.668  64.819  1.00  2.90
ATOM   1838  C   GLN A 113    39.232  55.012  58.541  1.00  2.90
ATOM   1839  O   GLN A 113    38.222  55.703  58.681  1.00  2.90
ATOM   1840  N   THR A 114    40.229  55.400  57.745  1.00  2.88
ATOM   1841  H   THR A 114    41.035  54.794  57.640  1.00  2.88
ATOM   1842  CA  THR A 114     40.262  56.725  57.116  1.00  2.88
ATOM   1843  HA  THR A 114     39.272  57.171  57.175  1.00  2.88
ATOM   1844  CB  THR A 114     41.217  57.686  57.861  1.00  2.88
ATOM   1845  HB  THR A 114     42.252  57.480  57.586  1.00  2.88
ATOM   1846  CG2 THR A 114     40.874  59.138  57.531  1.00  2.88
ATOM   1847 1HG2 THR A 114     41.550  59.785  58.080  1.00  2.88
ATOM   1848 2HG2 THR A 114     40.998  59.326  56.466  1.00  2.88
ATOM   1849 3HG2 THR A 114     39.847  59.355  57.825  1.00  2.88
ATOM   1850  OG1 THR A 114     41.127  57.588  59.263  1.00  2.88
ATOM   1851  HG1 THR A 114     41.615  56.782  59.498  1.00  2.88
ATOM   1852  C   THR A 114    40.635  56.642  55.630  1.00  2.88
```

```
ATOM   1853  O   THR A 114     41.716  57.090  55.237  1.00  2.88
ATOM   1854  N   PRO A 115     39.726  56.173  54.754  1.00  3.14
ATOM   1855  CD  PRO A 115     38.389  55.694  55.092  1.00  3.14
ATOM   1856  HD1 PRO A 115     37.848  56.395  55.727  1.00  3.14
ATOM   1857  HD2 PRO A 115     38.468  54.725  55.589  1.00  3.14
ATOM   1858  CG  PRO A 115     37.660  55.525  53.763  1.00  3.14
ATOM   1859  HG1 PRO A 115     37.230  56.479  53.454  1.00  3.14
ATOM   1860  HG2 PRO A 115     36.891  54.754  53.817  1.00  3.14
ATOM   1861  CB  PRO A 115     38.790  55.139  52.815  1.00  3.14
ATOM   1862  HB1 PRO A 115     38.534  55.364  51.779  1.00  3.14
ATOM   1863  HB2 PRO A 115     38.996  54.072  52.921  1.00  3.14
ATOM   1864  CA  PRO A 115     39.996  55.941  53.330  1.00  3.14
ATOM   1865  HA  PRO A 115     40.875  55.297  53.263  1.00  3.14
ATOM   1866  C   PRO A 115     40.303  57.179  52.456  1.00  3.14
ATOM   1867  O   PRO A 115     40.368  57.074  51.231  1.00  3.14
ATOM   1868  N   ASN A 116     40.546  58.336  53.076  1.00  3.04
ATOM   1869  H   ASN A 116     40.615  58.282  54.081  1.00  3.04
ATOM   1870  CA  ASN A 116     40.982  59.594  52.459  1.00  3.04
ATOM   1871  HA  ASN A 116     40.850  59.551  51.375  1.00  3.04
ATOM   1872  CB  ASN A 116     40.104  60.741  53.015  1.00  3.04
ATOM   1873  HB1 ASN A 116     40.444  61.683  52.583  1.00  3.04
ATOM   1874  HB2 ASN A 116     40.234  60.815  54.094  1.00  3.04
ATOM   1875  CG  ASN A 116     38.618  60.626  52.707  1.00  3.04
ATOM   1876  OD1 ASN A 116     38.078  61.270  51.825  1.00  3.04
ATOM   1877  ND2 ASN A 116     37.878  59.830  53.445  1.00  3.04
ATOM   1878 1HD2 ASN A 116     38.293  59.284  54.173  1.00  3.04
ATOM   1879 2HD2 ASN A 116     36.924  59.705  53.155  1.00  3.04
ATOM   1880  C   ASN A 116     42.477  59.902  52.736  1.00  3.04
ATOM   1881  O   ASN A 116     43.063  60.852  52.201  1.00  3.04
ATOM   1882  N   LEU A 117     43.115  59.090  53.586  1.00  2.58
ATOM   1883  H   LEU A 117     42.634  58.276  53.957  1.00  2.58
ATOM   1884  CA  LEU A 117     44.499  59.256  53.997  1.00  2.58
ATOM   1885  HA  LEU A 117     44.887  60.201  53.619  1.00  2.58
ATOM   1886  CB  LEU A 117     44.615  59.260  55.531  1.00  2.58
ATOM   1887  HB1 LEU A 117     45.668  59.136  55.789  1.00  2.58
ATOM   1888  HB2 LEU A 117     44.083  58.393  55.927  1.00  2.58
ATOM   1889  CG  LEU A 117     44.110  60.527  56.242  1.00  2.58
ATOM   1890  HG  LEU A 117     43.043  60.654  56.063  1.00  2.58
ATOM   1891  CD1 LEU A 117     44.361  60.349  57.734  1.00  2.58
ATOM   1892 1HD1 LEU A 117     44.074  61.249  58.273  1.00  2.58
ATOM   1893 2HD1 LEU A 117     43.790  59.500  58.109  1.00  2.58
ATOM   1894 3HD1 LEU A 117     45.418  60.151  57.902  1.00  2.58
ATOM   1895  CD2 LEU A 117     44.844  61.797  55.805  1.00  2.58
ATOM   1896 1HD2 LEU A 117     44.509  62.648  56.401  1.00  2.58
ATOM   1897 2HD2 LEU A 117     45.920  61.676  55.930  1.00  2.58
ATOM   1898 3HD2 LEU A 117     44.618  62.009  54.764  1.00  2.58
ATOM   1899  C   LEU A 117     45.374  58.156  53.430  1.00  2.58
ATOM   1900  O   LEU A 117     45.015  56.984  53.425  1.00  2.58
ATOM   1901  N   ARG A 118     46.549  58.554  52.940  1.00  2.41
ATOM   1902  H   ARG A 118     46.774  59.535  53.017  1.00  2.41
ATOM   1903  CA  ARG A 118     47.501  57.675  52.259  1.00  2.41
ATOM   1904  HA  ARG A 118     47.287  56.645  52.559  1.00  2.41
ATOM   1905  CB  ARG A 118     47.272  57.715  50.724  1.00  2.41
```

```
ATOM   1906  HB1 ARG A 118      48.067  57.155  50.230  1.00  2.41
ATOM   1907  HB2 ARG A 118      47.288  58.746  50.367  1.00  2.41
ATOM   1908  CG  ARG A 118     45.908  57.051  50.394  1.00  2.41
ATOM   1909  HG1 ARG A 118      45.119  57.612  50.889  1.00  2.41
ATOM   1910  HG2 ARG A 118      45.917  56.043  50.812  1.00  2.41
ATOM   1911  CD  ARG A 118     45.480  56.941  48.920  1.00  2.41
ATOM   1912  HD1 ARG A 118      46.304  56.507  48.349  1.00  2.41
ATOM   1913  HD2 ARG A 118      45.275  57.936  48.522  1.00  2.41
ATOM   1914  NE  ARG A 118     44.280  56.072  48.776  1.00  2.41
ATOM   1915  HE  ARG A 118     44.436  55.161  48.376  1.00  2.41
ATOM   1916  CZ  ARG A 118     43.060  56.290  49.254  1.00  2.41
ATOM   1917  NH1 ARG A 118      42.176  55.339  49.362  1.00  2.41
ATOM   1918 1HH1 ARG A 118      42.376  54.398  49.070  1.00  2.41
ATOM   1919 2HH1 ARG A 118      41.300  55.563  49.814  1.00  2.41
ATOM   1920  NH2 ARG A 118      42.649  57.459  49.646  1.00  2.41
ATOM   1921 1HH2 ARG A 118      43.170  58.292  49.431  1.00  2.41
ATOM   1922 2HH2 ARG A 118      41.722  57.537  50.044  1.00  2.41
ATOM   1923  C   ARG A 118     48.908  57.971  52.781  1.00  2.41
ATOM   1924  O   ARG A 118     49.525  58.998  52.474  1.00  2.41
ATOM   1925  N   LEU A 119     49.308  57.109  53.711  1.00  2.13
ATOM   1926  H   LEU A 119     48.723  56.303  53.872  1.00  2.13
ATOM   1927  CA  LEU A 119      50.464  57.254  54.593  1.00  2.13
ATOM   1928  HA  LEU A 119      50.647  58.309  54.802  1.00  2.13
ATOM   1929  CB  LEU A 119      50.125  56.510  55.905  1.00  2.13
ATOM   1930  HB1 LEU A 119      51.034  56.406  56.489  1.00  2.13
ATOM   1931  HB2 LEU A 119      49.798  55.503  55.645  1.00  2.13
ATOM   1932  CG  LEU A 119      49.035  57.163  56.786  1.00  2.13
ATOM   1933  HG  LEU A 119      48.208  57.482  56.152  1.00  2.13
ATOM   1934  CD1 LEU A 119      48.483  56.179  57.818  1.00  2.13
ATOM   1935 1HD1 LEU A 119      47.708  56.661  58.413  1.00  2.13
ATOM   1936 2HD1 LEU A 119      48.045  55.322  57.309  1.00  2.13
ATOM   1937 3HD1 LEU A 119      49.276  55.828  58.480  1.00  2.13
ATOM   1938  CD2 LEU A 119      49.549  58.374  57.566  1.00  2.13
ATOM   1939 1HD2 LEU A 119      48.706  58.884  58.032  1.00  2.13
ATOM   1940 2HD2 LEU A 119      50.229  58.053  58.357  1.00  2.13
ATOM   1941 3HD2 LEU A 119      50.052  59.072  56.904  1.00  2.13
ATOM   1942  C   LEU A 119     51.731  56.652  53.938  1.00  2.13
ATOM   1943  O   LEU A 119     51.728  55.537  53.378  1.00  2.13
ATOM   1944  N   ARG A 120     52.819  57.422  54.038  1.00  2.13
ATOM   1945  H   ARG A 120     52.761  58.305  54.538  1.00  2.13
ATOM   1946  CA  ARG A 120     54.098  57.083  53.411  1.00  2.13
ATOM   1947  HA  ARG A 120     54.093  56.058  53.033  1.00  2.13
ATOM   1948  CB  ARG A 120     54.405  58.049  52.242  1.00  2.13
ATOM   1949  HB1 ARG A 120      55.490  58.089  52.132  1.00  2.13
ATOM   1950  HB2 ARG A 120      54.066  59.049  52.510  1.00  2.13
ATOM   1951  CG  ARG A 120     53.849  57.720  50.852  1.00  2.13
ATOM   1952  HG1 ARG A 120      52.762  57.815  50.848  1.00  2.13
ATOM   1953  HG2 ARG A 120      54.136  56.709  50.576  1.00  2.13
ATOM   1954  CD  ARG A 120     54.484  58.721  49.865  1.00  2.13
ATOM   1955  HD1 ARG A 120      55.570  58.700  49.982  1.00  2.13
ATOM   1956  HD2 ARG A 120      54.150  59.726  50.131  1.00  2.13
ATOM   1957  NE  ARG A 120     54.152  58.467  48.449  1.00  2.13
ATOM   1958  HE  ARG A 120     53.451  59.068  48.049  1.00  2.13
```

```
ATOM   1959  CZ  ARG A 120      54.748  57.610  47.635  1.00  2.13
ATOM   1960  NH1 ARG A 120      54.477  57.603  46.360  1.00  2.13
ATOM   1961 1HH1 ARG A 120      53.855  58.291  45.977  1.00  2.13
ATOM   1962 2HH1 ARG A 120      54.856  56.879  45.771  1.00  2.13
ATOM   1963  NH2 ARG A 120      55.614  56.723  48.036  1.00  2.13
ATOM   1964 1HH2 ARG A 120      55.824  56.614  49.023  1.00  2.13
ATOM   1965 2HH2 ARG A 120      56.033  56.081  47.389  1.00  2.13
ATOM   1966  C   ARG A 120     55.256  57.175  54.418  1.00  2.13
ATOM   1967  O   ARG A 120     55.793  58.254  54.667  1.00  2.13
ATOM   1968  N   ILE A 121     55.669  56.043  54.969  1.00  1.99
ATOM   1969  H   ILE A 121     55.242  55.180  54.654  1.00  1.99
ATOM   1970  CA  ILE A 121     56.808  55.935  55.882  1.00  1.99
ATOM   1971  HA  ILE A 121     56.793  56.780  56.574  1.00  1.99
ATOM   1972  CB  ILE A 121     56.789  54.608  56.696  1.00  1.99
ATOM   1973  HB  ILE A 121     57.250  53.816  56.109  1.00  1.99
ATOM   1974  CG2 ILE A 121     57.658  54.778  57.948  1.00  1.99
ATOM   1975 1HG2 ILE A 121     57.848  53.813  58.413  1.00  1.99
ATOM   1976 2HG2 ILE A 121     58.615  55.230  57.700  1.00  1.99
ATOM   1977 3HG2 ILE A 121     57.158  55.424  58.659  1.00  1.99
ATOM   1978  CG1 ILE A 121     55.388  54.079  57.010  1.00  1.99
ATOM   1979 1HG1 ILE A 121     54.960  53.738  56.076  1.00  1.99
ATOM   1980 2HG1 ILE A 121     54.780  54.900  57.364  1.00  1.99
ATOM   1981  CD1 ILE A 121     55.335  52.906  57.999  1.00  1.99
ATOM   1982  HD1 ILE A 121     54.314  52.537  58.065  1.00  1.99
ATOM   1983  HD2 ILE A 121     55.980  52.097  57.659  1.00  1.99
ATOM   1984  HD3 ILE A 121     55.652  53.224  58.991  1.00  1.99
ATOM   1985  C   ILE A 121     58.136  55.884  55.148  1.00  1.99
ATOM   1986  O   ILE A 121     58.381  54.985  54.321  1.00  1.99
ATOM   1987  N   PHE A 122     59.040  56.779  55.534  1.00  2.05
ATOM   1988  H   PHE A 122     58.831  57.406  56.308  1.00  2.05
ATOM   1989  CA  PHE A 122     60.438  56.605  55.155  1.00  2.05
ATOM   1990  HA  PHE A 122     60.597  55.702  54.565  1.00  2.05
ATOM   1991  CB  PHE A 122     60.884  57.809  54.300  1.00  2.05
ATOM   1992  HB1 PHE A 122     61.966  57.739  54.210  1.00  2.05
ATOM   1993  HB2 PHE A 122     60.670  58.716  54.865  1.00  2.05
ATOM   1994  CG  PHE A 122     60.323  58.004  52.875  1.00  2.05
ATOM   1995  CD1 PHE A 122     59.023  57.629  52.462  1.00  2.05
ATOM   1996  HD1 PHE A 122     58.343  57.143  53.122  1.00  2.05
ATOM   1997  CE1 PHE A 122     58.539  57.922  51.177  1.00  2.05
ATOM   1998  HE1 PHE A 122     57.547  57.602  50.896  1.00  2.05
ATOM   1999  CZ  PHE A 122     59.313  58.686  50.295  1.00  2.05
ATOM   2000  HZ  PHE A 122     58.913  58.988  49.338  1.00  2.05
ATOM   2001  CE2 PHE A 122     60.589  59.102  50.697  1.00  2.05
ATOM   2002  HE2 PHE A 122     61.159  59.758  50.054  1.00  2.05
ATOM   2003  CD2 PHE A 122     61.108  58.710  51.946  1.00  2.05
ATOM   2004  HD2 PHE A 122     62.101  59.017  52.229  1.00  2.05
ATOM   2005  C   PHE A 122     61.181  56.422  56.473  1.00  2.05
ATOM   2006  O   PHE A 122     60.838  57.086  57.454  1.00  2.05
ATOM   2007  N   VAL A 123     62.123  55.478  56.550  1.00  2.09
ATOM   2008  H   VAL A 123     62.374  54.950  55.721  1.00  2.09
ATOM   2009  CA  VAL A 123     62.778  55.188  57.848  1.00  2.09
ATOM   2010  HA  VAL A 123     62.520  55.970  58.560  1.00  2.09
ATOM   2011  CB  VAL A 123     62.344  53.849  58.501  1.00  2.09
```

```
ATOM   2012  HB  VAL A 123      62.713  53.855  59.528  1.00  2.09
ATOM   2013  CG1 VAL A 123      60.823  53.737  58.565  1.00  2.09
ATOM   2014 1HG1 VAL A 123      60.526  52.855  59.127  1.00  2.09
ATOM   2015 2HG1 VAL A 123      60.414  54.619  59.053  1.00  2.09
ATOM   2016 3HG1 VAL A 123      60.416  53.654  57.557  1.00  2.09
ATOM   2017  CG2 VAL A 123      62.867  52.563  57.841  1.00  2.09
ATOM   2018 1HG2 VAL A 123      62.420  51.691  58.312  1.00  2.09
ATOM   2019 2HG2 VAL A 123      62.621  52.557  56.782  1.00  2.09
ATOM   2020 3HG2 VAL A 123      63.948  52.491  57.961  1.00  2.09
ATOM   2021  C   VAL A 123     64.276  55.235  57.718  1.00  2.09
ATOM   2022  O   VAL A 123     64.841  54.596  56.829  1.00  2.09
ATOM   2023  N   SER A 124     64.935  55.956  58.627  1.00  2.18
ATOM   2024  H   SER A 124     64.430  56.562  59.269  1.00  2.18
ATOM   2025  CA  SER A 124     66.398  55.922  58.626  1.00  2.18
ATOM   2026  HA  SER A 124     66.725  56.005  57.590  1.00  2.18
ATOM   2027  CB  SER A 124     66.978  57.136  59.345  1.00  2.18
ATOM   2028  HB1 SER A 124     66.719  57.117  60.404  1.00  2.18
ATOM   2029  HB2 SER A 124     66.591  58.048  58.888  1.00  2.18
ATOM   2030  OG  SER A 124     68.377  57.075  59.169  1.00  2.18
ATOM   2031  HG  SER A 124     68.800  57.905  59.434  1.00  2.18
ATOM   2032  C   SER A 124     67.010  54.621  59.150  1.00  2.18
ATOM   2033  O   SER A 124     68.031  54.175  58.634  1.00  2.18
ATOM   2034  N   ARG A 125     66.401  53.967  60.146  1.00  2.05
ATOM   2035  H   ARG A 125     65.599  54.396  60.593  1.00  2.05
ATOM   2036  CA  ARG A 125     66.870  52.666  60.652  1.00  2.05
ATOM   2037  HA  ARG A 125     67.500  52.182  59.903  1.00  2.05
ATOM   2038  CB  ARG A 125     67.665  52.830  61.964  1.00  2.05
ATOM   2039  HB1 ARG A 125     67.976  51.835  62.284  1.00  2.05
ATOM   2040  HB2 ARG A 125     67.006  53.227  62.734  1.00  2.05
ATOM   2041  CG  ARG A 125     68.935  53.688  61.887  1.00  2.05
ATOM   2042  HG1 ARG A 125     69.428  53.469  60.943  1.00  2.05
ATOM   2043  HG2 ARG A 125     69.609  53.379  62.687  1.00  2.05
ATOM   2044  CD  ARG A 125     68.684  55.203  62.030  1.00  2.05
ATOM   2045  HD1 ARG A 125     68.005  55.542  61.262  1.00  2.05
ATOM   2046  HD2 ARG A 125     69.610  55.747  61.881  1.00  2.05
ATOM   2047  NE  ARG A 125     68.110  55.546  63.337  1.00  2.05
ATOM   2048  HE  ARG A 125     67.098  55.574  63.383  1.00  2.05
ATOM   2049  CZ  ARG A 125     68.770  55.709  64.461  1.00  2.05
ATOM   2050  NH1 ARG A 125     68.104  55.762  65.576  1.00  2.05
ATOM   2051 1HH1 ARG A 125     67.092  55.800  65.528  1.00  2.05
ATOM   2052 2HH1 ARG A 125     68.566  55.826  66.457  1.00  2.05
ATOM   2053  NH2 ARG A 125     70.073  55.803  64.496  1.00  2.05
ATOM   2054 1HH2 ARG A 125     70.594  55.835  63.626  1.00  2.05
ATOM   2055 2HH2 ARG A 125     70.562  55.913  65.357  1.00  2.05
ATOM   2056  C   ARG A 125     65.686  51.761  60.940  1.00  2.05
ATOM   2057  O   ARG A 125     64.613  52.212  61.343  1.00  2.05
ATOM   2058  N   LEU A 126     65.912  50.462  60.791  1.00  2.39
ATOM   2059  H   LEU A 126     66.836  50.164  60.533  1.00  2.39
ATOM   2060  CA  LEU A 126     64.902  49.438  60.999  1.00  2.39
ATOM   2061  HA  LEU A 126     63.914  49.886  60.911  1.00  2.39
ATOM   2062  CB  LEU A 126     65.028  48.391  59.883  1.00  2.39
ATOM   2063  HB1 LEU A 126     64.408  47.539  60.145  1.00  2.39
ATOM   2064  HB2 LEU A 126     66.068  48.066  59.827  1.00  2.39
```

```
ATOM   2065  CG  LEU A 126      64.569  48.912  58.506  1.00  2.39
ATOM   2066  HG  LEU A 126      64.888  49.946  58.371  1.00  2.39
ATOM   2067  CD1 LEU A 126      65.187  48.081  57.382  1.00  2.39
ATOM   2068 1HD1 LEU A 126      64.899  48.506  56.422  1.00  2.39
ATOM   2069 2HD1 LEU A 126      66.272  48.103  57.447  1.00  2.39
ATOM   2070 3HD1 LEU A 126      64.844  47.049  57.436  1.00  2.39
ATOM   2071  CD2 LEU A 126      63.047  48.832  58.361  1.00  2.39
ATOM   2072 1HD2 LEU A 126      62.751  49.256  57.401  1.00  2.39
ATOM   2073 2HD2 LEU A 126      62.713  47.795  58.416  1.00  2.39
ATOM   2074 3HD2 LEU A 126      62.568  49.399  59.157  1.00  2.39
ATOM   2075  C   LEU A 126     65.034  48.878  62.407  1.00  2.39
ATOM   2076  O   LEU A 126     65.991  48.180  62.739  1.00  2.39
ATOM   2077  N   TYR A 127     64.081  49.276  63.243  1.00  2.35
ATOM   2078  H   TYR A 127     63.329  49.821  62.842  1.00  2.35
ATOM   2079  CA  TYR A 127      63.884  48.793  64.601  1.00  2.35
ATOM   2080  HA  TYR A 127      64.821  48.889  65.147  1.00  2.35
ATOM   2081  CB  TYR A 127      62.806  49.668  65.268  1.00  2.35
ATOM   2082  HB1 TYR A 127      61.837  49.382  64.857  1.00  2.35
ATOM   2083  HB2 TYR A 127      62.986  50.700  64.967  1.00  2.35
ATOM   2084  CG  TYR A 127      62.672  49.684  66.786  1.00  2.35
ATOM   2085  CD1 TYR A 127      63.504  48.928  67.640  1.00  2.35
ATOM   2086  HD1 TYR A 127      64.274  48.291  67.241  1.00  2.35
ATOM   2087  CE1 TYR A 127      63.345  49.004  69.038  1.00  2.35
ATOM   2088  HE1 TYR A 127      63.979  48.424  69.692  1.00  2.35
ATOM   2089  CZ  TYR A 127      62.368  49.859  69.593  1.00  2.35
ATOM   2090  OH  TYR A 127      62.230  49.945  70.942  1.00  2.35
ATOM   2091  HH  TYR A 127      61.560  50.584  71.195  1.00  2.35
ATOM   2092  CE2 TYR A 127      61.528  50.614  68.744  1.00  2.35
ATOM   2093  HE2 TYR A 127      60.783  51.273  69.163  1.00  2.35
ATOM   2094  CD2 TYR A 127      61.684  50.520  67.348  1.00  2.35
ATOM   2095  HD2 TYR A 127      61.049  51.101  66.692  1.00  2.35
ATOM   2096  C   TYR A 127     63.451  47.325  64.522  1.00  2.35
ATOM   2097  O   TYR A 127     62.454  46.978  63.884  1.00  2.35
ATOM   2098  N   PHE A 128     64.236  46.480  65.176  1.00  2.67
ATOM   2099  H   PHE A 128     65.054  46.855  65.632  1.00  2.67
ATOM   2100  CA  PHE A 128      63.852  45.139  65.578  1.00  2.67
ATOM   2101  HA  PHE A 128      62.810  44.955  65.320  1.00  2.67
ATOM   2102  CB  PHE A 128      64.742  44.081  64.922  1.00  2.67
ATOM   2103  HB1 PHE A 128      64.393  43.088  65.209  1.00  2.67
ATOM   2104  HB2 PHE A 128      65.769  44.198  65.271  1.00  2.67
ATOM   2105  CG  PHE A 128      64.690  44.226  63.429  1.00  2.67
ATOM   2106  CD1 PHE A 128      65.654  45.014  62.780  1.00  2.67
ATOM   2107  HD1 PHE A 128      66.479  45.438  63.338  1.00  2.67
ATOM   2108  CE1 PHE A 128      65.473  45.366  61.436  1.00  2.67
ATOM   2109  HE1 PHE A 128      66.182  46.030  60.961  1.00  2.67
ATOM   2110  CZ  PHE A 128      64.315  44.954  60.755  1.00  2.67
ATOM   2111  HZ  PHE A 128      64.132  45.301  59.748  1.00  2.67
ATOM   2112  CE2 PHE A 128      63.357  44.157  61.406  1.00  2.67
ATOM   2113  HE2 PHE A 128      62.429  43.900  60.917  1.00  2.67
ATOM   2114  CD2 PHE A 128      63.552  43.776  62.739  1.00  2.67
ATOM   2115  HD2 PHE A 128      62.778  43.227  63.260  1.00  2.67
ATOM   2116  C   PHE A 128     64.000  45.112  67.093  1.00  2.67
ATOM   2117  O   PHE A 128     64.904  45.756  67.631  1.00  2.67
```

```
ATOM   2118  N   CYS A 129      63.109  44.420  67.788  1.00  3.13
ATOM   2119  H   CYS A 129      62.438  43.829  67.323  1.00  3.13
ATOM   2120  CA  CYS A 129      63.194  44.383  69.232  1.00  3.13
ATOM   2121  HA  CYS A 129      63.423  45.385  69.597  1.00  3.13
ATOM   2122  CB  CYS A 129      61.838  43.947  69.798  1.00  3.13
ATOM   2123  HB1 CYS A 129      61.738  42.861  69.725  1.00  3.13
ATOM   2124  HB2 CYS A 129      61.038  44.403  69.221  1.00  3.13
ATOM   2125  SG  CYS A 129      61.702  44.481  71.530  1.00  3.13
ATOM   2126  HG  CYS A 129      61.903  45.791  71.341  1.00  3.13
ATOM   2127  C   CYS A 129      64.298  43.428  69.676  1.00  3.13
ATOM   2128  O   CYS A 129      64.373  42.307  69.184  1.00  3.13
ATOM   2129  N   ASP A 130      65.077  43.821  70.683  1.00  4.50
ATOM   2130  H   ASP A 130      65.041  44.773  71.018  1.00  4.50
ATOM   2131  CA  ASP A 130      65.972  42.888  71.376  1.00  4.50
ATOM   2132  HA  ASP A 130      66.679  42.469  70.657  1.00  4.50
ATOM   2133  CB  ASP A 130      66.760  43.645  72.452  1.00  4.50
ATOM   2134  HB1 ASP A 130      67.337  42.935  73.046  1.00  4.50
ATOM   2135  HB2 ASP A 130      66.066  44.160  73.119  1.00  4.50
ATOM   2136  CG  ASP A 130      67.728  44.643  71.832  1.00  4.50
ATOM   2137  OD1 ASP A 130      68.700  44.214  71.169  1.00  4.50
ATOM   2138  OD2 ASP A 130      67.531  45.870  71.992  1.00  4.50
ATOM   2139  C   ASP A 130      65.239  41.700  72.029  1.00  4.50
ATOM   2140  O   ASP A 130      65.869  40.709  72.386  1.00  4.50
ATOM   2141  N   LEU A 131      63.913  41.801  72.195  1.00  6.56
ATOM   2142  H   LEU A 131      63.472  42.665  71.929  1.00  6.56
ATOM   2143  CA  LEU A 131      63.074  40.699  72.654  1.00  6.56
ATOM   2144  HA  LEU A 131      63.598  40.222  73.483  1.00  6.56
ATOM   2145  CB  LEU A 131      61.740  41.266  73.178  1.00  6.56
ATOM   2146  HB1 LEU A 131      61.180  41.658  72.329  1.00  6.56
ATOM   2147  HB2 LEU A 131      61.956  42.093  73.856  1.00  6.56
ATOM   2148  CG  LEU A 131      60.849  40.250  73.923  1.00  6.56
ATOM   2149  HG  LEU A 131      60.602  39.422  73.260  1.00  6.56
ATOM   2150  CD1 LEU A 131      61.494  39.697  75.198  1.00  6.56
ATOM   2151 1HD1 LEU A 131      60.788  39.051  75.719  1.00  6.56
ATOM   2152 2HD1 LEU A 131      62.366  39.098  74.933  1.00  6.56
ATOM   2153 3HD1 LEU A 131      61.800  40.514  75.852  1.00  6.56
ATOM   2154  CD2 LEU A 131      59.538  40.932  74.324  1.00  6.56
ATOM   2155 1HD2 LEU A 131      58.882  40.205  74.803  1.00  6.56
ATOM   2156 2HD2 LEU A 131      59.736  41.752  75.016  1.00  6.56
ATOM   2157 3HD2 LEU A 131      59.039  41.317  73.436  1.00  6.56
ATOM   2158  C   LEU A 131      62.860  39.608  71.594  1.00  6.56
ATOM   2159  O   LEU A 131      62.725  38.447  71.967  1.00  6.56
ATOM   2160  N   GLU A 132      62.815  39.979  70.307  1.00  4.85
ATOM   2161  H   GLU A 132      63.054  40.938  70.095  1.00  4.85
ATOM   2162  CA  GLU A 132      62.746  39.080  69.145  1.00  4.85
ATOM   2163  HA  GLU A 132      63.696  38.543  69.089  1.00  4.85
ATOM   2164  CB  GLU A 132      61.594  38.037  69.246  1.00  4.85
ATOM   2165  HB1 GLU A 132      60.809  38.280  68.533  1.00  4.85
ATOM   2166  HB2 GLU A 132      61.108  38.067  70.219  1.00  4.85
ATOM   2167  CG  GLU A 132      62.056  36.595  68.974  1.00  4.85
ATOM   2168  HG1 GLU A 132      61.239  35.914  69.225  1.00  4.85
ATOM   2169  HG2 GLU A 132      62.905  36.354  69.618  1.00  4.85
ATOM   2170  CD  GLU A 132      62.431  36.410  67.501  1.00  4.85
```

```
ATOM   2171  OE1 GLU A 132      61.500  36.273  66.678  1.00  4.85
ATOM   2172  OE2 GLU A 132      63.619  36.593  67.156  1.00  4.85
ATOM   2173  C   GLU A 132      62.561  39.867  67.832  1.00  4.85
ATOM   2174  O   GLU A 132      61.934  40.934  67.792  1.00  4.85
ATOM   2175  N   GLY A 133      63.008  39.275  66.723  1.00  3.82
ATOM   2176  H   GLY A 133      63.481  38.379  66.833  1.00  3.82
ATOM   2177  CA  GLY A 133      62.678  39.682  65.363  1.00  3.82
ATOM   2178  HA1 GLY A 133      63.201  39.018  64.676  1.00  3.82
ATOM   2179  HA2 GLY A 133      63.033  40.700  65.204  1.00  3.82
ATOM   2180  C   GLY A 133      61.178  39.636  65.014  1.00  3.82
ATOM   2181  O   GLY A 133      60.641  40.587  64.414  1.00  3.82
ATOM   2182  N   SER A 134      60.518  38.543  65.425  1.00  3.24
ATOM   2183  H   SER A 134      61.062  37.821  65.896  1.00  3.24
ATOM   2184  CA  SER A 134      59.126  38.192  65.132  1.00  3.24
ATOM   2185  HA  SER A 134      59.114  37.860  64.094  1.00  3.24
ATOM   2186  CB  SER A 134      58.641  36.985  65.943  1.00  3.24
ATOM   2187  HB1 SER A 134      58.765  37.165  67.010  1.00  3.24
ATOM   2188  HB2 SER A 134      59.232  36.112  65.665  1.00  3.24
ATOM   2189  OG  SER A 134      57.279  36.728  65.648  1.00  3.24
ATOM   2190  HG  SER A 134      57.062  35.846  65.963  1.00  3.24
ATOM   2191  C   SER A 134      58.161  39.396  65.175  1.00  3.24
ATOM   2192  O   SER A 134      57.623  39.773  64.123  1.00  3.24
ATOM   2193  N   PRO A 135      57.977  40.087  66.321  1.00  2.97
ATOM   2194  CD  PRO A 135      58.636  39.883  67.602  1.00  2.97
ATOM   2195  HD1 PRO A 135      59.715  39.848  67.491  1.00  2.97
ATOM   2196  HD2 PRO A 135      58.269  38.968  68.067  1.00  2.97
ATOM   2197  CG  PRO A 135      58.250  41.079  68.466  1.00  2.97
ATOM   2198  HG1 PRO A 135      58.933  41.908  68.269  1.00  2.97
ATOM   2199  HG2 PRO A 135      58.245  40.827  69.526  1.00  2.97
ATOM   2200  CB  PRO A 135      56.854  41.419  67.949  1.00  2.97
ATOM   2201  HB1 PRO A 135      56.597  42.463  68.133  1.00  2.97
ATOM   2202  HB2 PRO A 135      56.124  40.764  68.428  1.00  2.97
ATOM   2203  CA  PRO A 135      56.924  41.086  66.454  1.00  2.97
ATOM   2204  HA  PRO A 135      55.974  40.642  66.153  1.00  2.97
ATOM   2205  C   PRO A 135      57.124  42.347  65.597  1.00  2.97
ATOM   2206  O   PRO A 135      56.152  42.915  65.092  1.00  2.97
ATOM   2207  N   HIS A 136      58.363  42.816  65.412  1.00  2.61
ATOM   2208  H   HIS A 136      59.147  42.283  65.768  1.00  2.61
ATOM   2209  CA  HIS A 136      58.601  44.023  64.602  1.00  2.61
ATOM   2210  HA  HIS A 136      57.809  44.750  64.795  1.00  2.61
ATOM   2211  CB  HIS A 136      59.932  44.676  65.004  1.00  2.61
ATOM   2212  HB1 HIS A 136      60.231  45.359  64.207  1.00  2.61
ATOM   2213  HB2 HIS A 136      60.704  43.910  65.087  1.00  2.61
ATOM   2214  CG  HIS A 136      59.894  45.484  66.288  1.00  2.61
ATOM   2215  ND1 HIS A 136      60.678  46.609  66.529  1.00  2.61
ATOM   2216  CE1 HIS A 136      60.405  47.023  67.770  1.00  2.61
ATOM   2217  HE1 HIS A 136      60.886  47.864  68.253  1.00  2.61
ATOM   2218  NE2 HIS A 136      59.477  46.226  68.325  1.00  2.61
ATOM   2219  HE2 HIS A 136      59.186  46.275  69.294  1.00  2.61
ATOM   2220  CD2 HIS A 136      59.144  45.248  67.408  1.00  2.61
ATOM   2221  HD2 HIS A 136      58.459  44.425  67.555  1.00  2.61
ATOM   2222  C   HIS A 136      58.522  43.752  63.101  1.00  2.61
ATOM   2223  O   HIS A 136      57.954  44.570  62.365  1.00  2.61
```

```
ATOM   2224  N   VAL A 137      59.010  42.585  62.660  1.00  2.73
ATOM   2225  H   VAL A 137      59.408  41.931  63.330  1.00  2.73
ATOM   2226  CA  VAL A 137      58.846  42.143  61.260  1.00  2.73
ATOM   2227  HA  VAL A 137      59.258  42.897  60.590  1.00  2.73
ATOM   2228  CB  VAL A 137      59.630  40.823  61.071  1.00  2.73
ATOM   2229  HB  VAL A 137      59.380  40.142  61.886  1.00  2.73
ATOM   2230  CG1 VAL A 137      59.337  40.083  59.759  1.00  2.73
ATOM   2231 1HG1 VAL A 137      58.313  39.713  59.772  1.00  2.73
ATOM   2232 2HG1 VAL A 137      59.487  40.737  58.906  1.00  2.73
ATOM   2233 3HG1 VAL A 137      59.993  39.219  59.667  1.00  2.73
ATOM   2234  CG2 VAL A 137      61.138  41.110  61.132  1.00  2.73
ATOM   2235 1HG2 VAL A 137      61.405  41.845  60.379  1.00  2.73
ATOM   2236 2HG2 VAL A 137      61.403  41.499  62.113  1.00  2.73
ATOM   2237 3HG2 VAL A 137      61.702  40.191  60.970  1.00  2.73
ATOM   2238  C   VAL A 137      57.370  41.986  60.897  1.00  2.73
ATOM   2239  O   VAL A 137      56.932  42.522  59.873  1.00  2.73
ATOM   2240  N   GLU A 138      56.574  41.329  61.747  1.00  2.82
ATOM   2241  H   GLU A 138      56.969  40.879  62.571  1.00  2.82
ATOM   2242  CA  GLU A 138      55.122  41.245  61.525  1.00  2.82
ATOM   2243  HA  GLU A 138      54.965  40.868  60.517  1.00  2.82
ATOM   2244  CB  GLU A 138      54.481  40.224  62.484  1.00  2.82
ATOM   2245  HB1 GLU A 138      54.660  40.547  63.511  1.00  2.82
ATOM   2246  HB2 GLU A 138      54.981  39.264  62.348  1.00  2.82
ATOM   2247  CG  GLU A 138      52.960  40.016  62.295  1.00  2.82
ATOM   2248  HG1 GLU A 138      52.441  40.927  62.603  1.00  2.82
ATOM   2249  HG2 GLU A 138      52.641  39.224  62.976  1.00  2.82
ATOM   2250  CD  GLU A 138      52.529  39.636  60.864  1.00  2.82
ATOM   2251  OE1 GLU A 138      53.309  38.969  60.147  1.00  2.82
ATOM   2252  OE2 GLU A 138      51.412  39.998  60.434  1.00  2.82
ATOM   2253  C   GLU A 138      54.413  42.601  61.597  1.00  2.82
ATOM   2254  O   GLU A 138      53.465  42.801  60.856  1.00  2.82
ATOM   2255  N   GLY A 139      54.882  43.572  62.387  1.00  2.70
ATOM   2256  H   GLY A 139      55.609  43.357  63.059  1.00  2.70
ATOM   2257  CA  GLY A 139      54.320  44.930  62.321  1.00  2.70
ATOM   2258  HA1 GLY A 139      54.779  45.531  63.105  1.00  2.70
ATOM   2259  HA2 GLY A 139      53.247  44.882  62.509  1.00  2.70
ATOM   2260  C   GLY A 139      54.544  45.640  60.986  1.00  2.70
ATOM   2261  O   GLY A 139      53.619  46.235  60.427  1.00  2.70
ATOM   2262  N   LEU A 140      55.761  45.552  60.439  1.00  2.65
ATOM   2263  H   LEU A 140      56.477  45.027  60.932  1.00  2.65
ATOM   2264  CA  LEU A 140      56.051  46.103  59.102  1.00  2.65
ATOM   2265  HA  LEU A 140      55.795  47.161  59.093  1.00  2.65
ATOM   2266  CB  LEU A 140      57.559  45.943  58.822  1.00  2.65
ATOM   2267  HB1 LEU A 140      57.749  46.084  57.758  1.00  2.65
ATOM   2268  HB2 LEU A 140      57.851  44.923  59.074  1.00  2.65
ATOM   2269  CG  LEU A 140      58.439  46.932  59.610  1.00  2.65
ATOM   2270  HG  LEU A 140      58.102  46.989  60.645  1.00  2.65
ATOM   2271  CD1 LEU A 140      59.898  46.472  59.609  1.00  2.65
ATOM   2272 1HD1 LEU A 140      60.497  47.159  60.209  1.00  2.65
ATOM   2273 2HD1 LEU A 140      59.966  45.481  60.054  1.00  2.65
ATOM   2274 3HD1 LEU A 140      60.284  46.451  58.589  1.00  2.65
ATOM   2275  CD2 LEU A 140      58.392  48.328  58.988  1.00  2.65
ATOM   2276 1HD2 LEU A 140      59.047  48.993  59.542  1.00  2.65
```

```
ATOM   2277 2HD2 LEU A 140      58.724  48.288  57.951  1.00  2.65
ATOM   2278 3HD2 LEU A 140      57.379  48.724  59.032  1.00  2.65
ATOM   2279  C   LEU A 140     55.209  45.442  58.006  1.00  2.65
ATOM   2280  O   LEU A 140     54.662  46.141  57.143  1.00  2.65
ATOM   2281  N   ARG A 141     55.075  44.111  58.070  1.00  2.80
ATOM   2282  H   ARG A 141     55.568  43.631  58.819  1.00  2.80
ATOM   2283  CA  ARG A 141     54.167  43.337  57.209  1.00  2.80
ATOM   2284  HA  ARG A 141     54.419  43.493  56.161  1.00  2.80
ATOM   2285  CB  ARG A 141     54.314  41.837  57.535  1.00  2.80
ATOM   2286  HB1 ARG A 141     53.439  41.294  57.171  1.00  2.80
ATOM   2287  HB2 ARG A 141     54.349  41.716  58.615  1.00  2.80
ATOM   2288  CG  ARG A 141     55.562  41.212  56.889  1.00  2.80
ATOM   2289  HG1 ARG A 141     56.319  41.985  56.750  1.00  2.80
ATOM   2290  HG2 ARG A 141     55.295  40.830  55.902  1.00  2.80
ATOM   2291  CD  ARG A 141     56.212  40.089  57.712  1.00  2.80
ATOM   2292  HD1 ARG A 141     56.672  40.537  58.585  1.00  2.80
ATOM   2293  HD2 ARG A 141     57.008  39.649  57.111  1.00  2.80
ATOM   2294  NE  ARG A 141     55.275  39.050  58.179  1.00  2.80
ATOM   2295  HE  ARG A 141     54.476  39.340  58.738  1.00  2.80
ATOM   2296  CZ  ARG A 141     55.384  37.745  58.067  1.00  2.80
ATOM   2297  NH1 ARG A 141     54.468  36.987  58.578  1.00  2.80
ATOM   2298 1HH1 ARG A 141     53.732  37.460  59.107  1.00  2.80
ATOM   2299 2HH1 ARG A 141     54.487  35.996  58.463  1.00  2.80
ATOM   2300  NH2 ARG A 141     56.398  37.199  57.465  1.00  2.80
ATOM   2301 1HH2 ARG A 141     57.073  37.812  57.057  1.00  2.80
ATOM   2302 2HH2 ARG A 141     56.405  36.223  57.192  1.00  2.80
ATOM   2303  C   ARG A 141     52.709  43.758  57.370  1.00  2.80
ATOM   2304  O   ARG A 141     52.054  43.896  56.345  1.00  2.80
ATOM   2305  N   ASP A 142     52.171  44.026  58.564  1.00  2.84
ATOM   2306  H   ASP A 142     52.678  43.804  59.417  1.00  2.84
ATOM   2307  CA  ASP A 142     50.770  44.444  58.627  1.00  2.84
ATOM   2308  HA  ASP A 142     50.295  43.885  57.825  1.00  2.84
ATOM   2309  CB  ASP A 142     49.952  44.025  59.856  1.00  2.84
ATOM   2310  HB1 ASP A 142     50.095  44.757  60.654  1.00  2.84
ATOM   2311  HB2 ASP A 142     50.300  43.056  60.214  1.00  2.84
ATOM   2312  CG  ASP A 142     48.444  43.922  59.485  1.00  2.84
ATOM   2313  OD1 ASP A 142     48.081  43.433  58.382  1.00  2.84
ATOM   2314  OD2 ASP A 142     47.577  44.297  60.298  1.00  2.84
ATOM   2315  C   ASP A 142     50.517  45.909  58.235  1.00  2.84
ATOM   2316  O   ASP A 142     49.449  46.213  57.715  1.00  2.84
ATOM   2317  N   LEU A 143     51.487  46.816  58.385  1.00  2.64
ATOM   2318  H   LEU A 143     52.322  46.527  58.885  1.00  2.64
ATOM   2319  CA  LEU A 143     51.440  48.165  57.783  1.00  2.64
ATOM   2320  HA  LEU A 143     50.572  48.717  58.137  1.00  2.64
ATOM   2321  CB  LEU A 143     52.731  48.917  58.169  1.00  2.64
ATOM   2322  HB1 LEU A 143     52.926  49.704  57.437  1.00  2.64
ATOM   2323  HB2 LEU A 143     53.567  48.221  58.114  1.00  2.64
ATOM   2324  CG  LEU A 143     52.706  49.551  59.566  1.00  2.64
ATOM   2325  HG  LEU A 143     52.232  48.873  60.272  1.00  2.64
ATOM   2326  CD1 LEU A 143     54.137  49.834  60.028  1.00  2.64
ATOM   2327 1HD1 LEU A 143     54.123  50.315  61.004  1.00  2.64
ATOM   2328 2HD1 LEU A 143     54.676  48.893  60.118  1.00  2.64
ATOM   2329 3HD1 LEU A 143     54.646  50.474  59.311  1.00  2.64
```

```
ATOM   2330  CD2 LEU A 143      51.946  50.875  59.549  1.00  2.64
ATOM   2331 1HD2 LEU A 143      52.075  51.371  60.505  1.00  2.64
ATOM   2332 2HD2 LEU A 143      52.321  51.520  58.758  1.00  2.64
ATOM   2333 3HD2 LEU A 143      50.884  50.677  59.392  1.00  2.64
ATOM   2334  C   LEU A 143     51.340  48.115  56.243  1.00  2.64
ATOM   2335  O   LEU A 143     50.478  48.719  55.569  1.00  2.64
ATOM   2336  N   ARG A 144     52.245  47.301  55.701  1.00  2.84
ATOM   2337  H   ARG A 144     52.924  46.880  56.330  1.00  2.84
ATOM   2338  CA  ARG A 144     52.312  46.962  54.293  1.00  2.84
ATOM   2339  HA  ARG A 144     52.533  47.880  53.747  1.00  2.84
ATOM   2340  CB  ARG A 144     53.481  45.966  54.144  1.00  2.84
ATOM   2341  HB1 ARG A 144     53.364  45.138  54.829  1.00  2.84
ATOM   2342  HB2 ARG A 144     54.400  46.477  54.435  1.00  2.84
ATOM   2343  CG  ARG A 144     53.674  45.366  52.750  1.00  2.84
ATOM   2344  HG1 ARG A 144     52.901  44.626  52.569  1.00  2.84
ATOM   2345  HG2 ARG A 144     54.637  44.858  52.701  1.00  2.84
ATOM   2346  CD  ARG A 144     53.619  46.455  51.689  1.00  2.84
ATOM   2347  HD1 ARG A 144     52.593  46.804  51.585  1.00  2.84
ATOM   2348  HD2 ARG A 144     53.928  46.060  50.725  1.00  2.84
ATOM   2349  NE  ARG A 144     54.512  47.566  52.030  1.00  2.84
ATOM   2350  HE  ARG A 144     55.329  47.356  52.572  1.00  2.84
ATOM   2351  CZ  ARG A 144     54.475  48.703  51.400  1.00  2.84
ATOM   2352  NH1 ARG A 144     55.603  49.342  51.254  1.00  2.84
ATOM   2353 1HH1 ARG A 144     56.451  48.951  51.640  1.00  2.84
ATOM   2354 2HH1 ARG A 144     55.692  50.040  50.526  1.00  2.84
ATOM   2355  NH2 ARG A 144     53.358  49.115  50.858  1.00  2.84
ATOM   2356 1HH2 ARG A 144     52.511  48.643  51.128  1.00  2.84
ATOM   2357 2HH2 ARG A 144     53.289  49.860  50.180  1.00  2.84
ATOM   2358  C   ARG A 144     50.975  46.455  53.759  1.00  2.84
ATOM   2359  O   ARG A 144     50.408  47.101  52.882  1.00  2.84
ATOM   2360  N   ARG A 145     50.441  45.378  54.343  1.00  3.04
ATOM   2361  H   ARG A 145     51.020  44.892  55.023  1.00  3.04
ATOM   2362  CA  ARG A 145     49.106  44.821  54.061  1.00  3.04
ATOM   2363  HA  ARG A 145     49.023  44.597  52.995  1.00  3.04
ATOM   2364  CB  ARG A 145     48.918  43.504  54.838  1.00  3.04
ATOM   2365  HB1 ARG A 145     47.894  43.156  54.696  1.00  3.04
ATOM   2366  HB2 ARG A 145     49.056  43.727  55.895  1.00  3.04
ATOM   2367  CG  ARG A 145     49.853  42.349  54.429  1.00  3.04
ATOM   2368  HG1 ARG A 145     50.890  42.626  54.594  1.00  3.04
ATOM   2369  HG2 ARG A 145     49.733  42.156  53.363  1.00  3.04
ATOM   2370  CD  ARG A 145     49.559  41.042  55.194  1.00  3.04
ATOM   2371  HD1 ARG A 145     50.407  40.367  55.062  1.00  3.04
ATOM   2372  HD2 ARG A 145     48.691  40.569  54.731  1.00  3.04
ATOM   2373  NE  ARG A 145     49.265  41.259  56.632  1.00  3.04
ATOM   2374  HE  ARG A 145     48.617  42.001  56.866  1.00  3.04
ATOM   2375  CZ  ARG A 145     49.761  40.641  57.686  1.00  3.04
ATOM   2376  NH1 ARG A 145     49.317  40.955  58.859  1.00  3.04
ATOM   2377 1HH1 ARG A 145     48.650  41.709  58.947  1.00  3.04
ATOM   2378 2HH1 ARG A 145     49.856  40.623  59.658  1.00  3.04
ATOM   2379  NH2 ARG A 145     50.696  39.743  57.656  1.00  3.04
ATOM   2380 1HH2 ARG A 145     51.038  39.372  56.787  1.00  3.04
ATOM   2381 2HH2 ARG A 145     51.102  39.492  58.557  1.00  3.04
ATOM   2382  C   ARG A 145     47.938  45.752  54.402  1.00  3.04
```

```
ATOM   2383  O   ARG A 145     46.794  45.445  54.064  1.00  3.04
ATOM   2384  N   ALA A 146     48.159  46.851  55.115  1.00  2.99
ATOM   2385  H   ALA A 146     49.077  47.022  55.499  1.00  2.99
ATOM   2386  CA  ALA A 146     47.140  47.865  55.337  1.00  2.99
ATOM   2387  HA  ALA A 146     46.161  47.393  55.365  1.00  2.99
ATOM   2388  CB  ALA A 146     47.361  48.539  56.695  1.00  2.99
ATOM   2389  HB1 ALA A 146     46.516  49.188  56.928  1.00  2.99
ATOM   2390  HB2 ALA A 146     47.454  47.776  57.464  1.00  2.99
ATOM   2391  HB3 ALA A 146     48.268  49.140  56.685  1.00  2.99
ATOM   2392  C   ALA A 146     47.072  48.890  54.215  1.00  2.99
ATOM   2393  O   ALA A 146     46.038  49.534  54.038  1.00  2.99
ATOM   2394  N   GLY A 147     48.163  49.018  53.463  1.00  2.95
ATOM   2395  H   GLY A 147     48.949  48.404  53.640  1.00  2.95
ATOM   2396  CA  GLY A 147     48.240  49.916  52.317  1.00  2.95
ATOM   2397  HA1 GLY A 147     47.255  50.247  51.990  1.00  2.95
ATOM   2398  HA2 GLY A 147     48.722  49.381  51.529  1.00  2.95
ATOM   2399  C   GLY A 147     49.129  51.112  52.568  1.00  2.95
ATOM   2400  O   GLY A 147     49.031  52.124  51.877  1.00  2.95
ATOM   2401  N   VAL A 148     49.995  50.991  53.574  1.00  2.82
ATOM   2402  H   VAL A 148     50.036  50.116  54.085  1.00  2.82
ATOM   2403  CA  VAL A 148     50.995  52.003  53.871  1.00  2.82
ATOM   2404  HA  VAL A 148     50.619  52.985  53.581  1.00  2.82
ATOM   2405  CB  VAL A 148     51.289  52.028  55.389  1.00  2.82
ATOM   2406  HB  VAL A 148     51.767  51.093  55.681  1.00  2.82
ATOM   2407  CG1 VAL A 148     52.230  53.178  55.741  1.00  2.82
ATOM   2408 1HG1 VAL A 148     52.472  53.151  56.802  1.00  2.82
ATOM   2409 2HG1 VAL A 148     53.148  53.084  55.169  1.00  2.82
ATOM   2410 3HG1 VAL A 148     51.767  54.133  55.515  1.00  2.82
ATOM   2411  CG2 VAL A 148     50.036  52.222  56.255  1.00  2.82
ATOM   2412 1HG2 VAL A 148     50.310  52.266  57.308  1.00  2.82
ATOM   2413 2HG2 VAL A 148     49.527  53.144  55.981  1.00  2.82
ATOM   2414 3HG2 VAL A 148     49.362  51.377  56.125  1.00  2.82
ATOM   2415  C   VAL A 148     52.262  51.704  53.114  1.00  2.82
ATOM   2416  O   VAL A 148     52.773  50.587  53.182  1.00  2.82
ATOM   2417  N   GLN A 149     52.772  52.702  52.391  1.00  2.57
ATOM   2418  H   GLN A 149     52.358  53.625  52.467  1.00  2.57
ATOM   2419  CA  GLN A 149     54.018  52.519  51.634  1.00  2.57
ATOM   2420  HA  GLN A 149     54.051  51.520  51.216  1.00  2.57
ATOM   2421  CB  GLN A 149     53.956  53.531  50.479  1.00  2.57
ATOM   2422  HB1 GLN A 149     54.002  54.540  50.881  1.00  2.57
ATOM   2423  HB2 GLN A 149     52.985  53.417  49.992  1.00  2.57
ATOM   2424  CG  GLN A 149     55.013  53.340  49.388  1.00  2.57
ATOM   2425  HG1 GLN A 149     54.663  53.813  48.471  1.00  2.57
ATOM   2426  HG2 GLN A 149     55.121  52.279  49.181  1.00  2.57
ATOM   2427  CD  GLN A 149     56.358  53.955  49.737  1.00  2.57
ATOM   2428  OE1 GLN A 149     56.451  55.104  50.148  1.00  2.57
ATOM   2429  NE2 GLN A 149     57.449  53.259  49.529  1.00  2.57
ATOM   2430 1HE2 GLN A 149     57.376  52.365  49.057  1.00  2.57
ATOM   2431 2HE2 GLN A 149     58.328  53.675  49.772  1.00  2.57
ATOM   2432  C   GLN A 149     55.180  52.720  52.609  1.00  2.57
ATOM   2433  O   GLN A 149     55.145  53.617  53.439  1.00  2.57
ATOM   2434  N   VAL A 150     56.182  51.846  52.551  1.00  2.20
ATOM   2435  H   VAL A 150     56.165  51.147  51.827  1.00  2.20
```

```
ATOM   2436  CA  VAL A 150      57.245  51.721  53.554  1.00  2.20
ATOM   2437  HA  VAL A 150      57.313  52.650  54.119  1.00  2.20
ATOM   2438  CB  VAL A 150      56.998  50.559  54.549  1.00  2.20
ATOM   2439  HB  VAL A 150      57.165  49.608  54.042  1.00  2.20
ATOM   2440  CG1 VAL A 150      57.975  50.647  55.728  1.00  2.20
ATOM   2441 1HG1 VAL A 150      57.773  49.848  56.438  1.00  2.20
ATOM   2442 2HG1 VAL A 150      59.001  50.538  55.377  1.00  2.20
ATOM   2443 3HG1 VAL A 150      57.873  51.606  56.232  1.00  2.20
ATOM   2444  CG2 VAL A 150      55.573  50.517  55.110  1.00  2.20
ATOM   2445 1HG2 VAL A 150      55.508  49.830  55.954  1.00  2.20
ATOM   2446 2HG2 VAL A 150      55.267  51.511  55.428  1.00  2.20
ATOM   2447 3HG2 VAL A 150      54.896  50.165  54.338  1.00  2.20
ATOM   2448  C   VAL A 150     58.548  51.484  52.828  1.00  2.20
ATOM   2449  O   VAL A 150     58.669  50.528  52.058  1.00  2.20
ATOM   2450  N   LYS A 151     59.498  52.377  53.078  1.00  2.19
ATOM   2451  H   LYS A 151     59.268  53.144  53.704  1.00  2.19
ATOM   2452  CA  LYS A 151      60.795  52.451  52.404  1.00  2.19
ATOM   2453  HA  LYS A 151      61.025  51.518  51.889  1.00  2.19
ATOM   2454  CB  LYS A 151      60.640  53.605  51.392  1.00  2.19
ATOM   2455  HB1 LYS A 151      60.250  54.468  51.938  1.00  2.19
ATOM   2456  HB2 LYS A 151      59.891  53.313  50.658  1.00  2.19
ATOM   2457  CG  LYS A 151      61.891  54.064  50.631  1.00  2.19
ATOM   2458  HG1 LYS A 151      62.254  53.267  49.981  1.00  2.19
ATOM   2459  HG2 LYS A 151      62.670  54.332  51.340  1.00  2.19
ATOM   2460  CD  LYS A 151      61.564  55.316  49.806  1.00  2.19
ATOM   2461  HD1 LYS A 151      60.951  55.987  50.411  1.00  2.19
ATOM   2462  HD2 LYS A 151      61.012  55.042  48.906  1.00  2.19
ATOM   2463  CE  LYS A 151      62.857  56.048  49.447  1.00  2.19
ATOM   2464  HE1 LYS A 151      63.418  55.494  48.690  1.00  2.19
ATOM   2465  HE2 LYS A 151      63.473  56.108  50.349  1.00  2.19
ATOM   2466  NZ  LYS A 151      62.578  57.427  48.998  1.00  2.19
ATOM   2467  HZ1 LYS A 151      62.292  57.493  48.038  1.00  2.19
ATOM   2468  HZ2 LYS A 151      61.893  57.857  49.605  1.00  2.19
ATOM   2469  HZ3 LYS A 151      63.395  58.021  49.165  1.00  2.19
ATOM   2470  C   LYS A 151     61.878  52.746  53.456  1.00  2.19
ATOM   2471  O   LYS A 151     61.579  53.188  54.571  1.00  2.19
ATOM   2472  N   VAL A 152     63.145  52.518  53.097  1.00  2.18
ATOM   2473  H   VAL A 152     63.328  52.117  52.192  1.00  2.18
ATOM   2474  CA  VAL A 152      64.291  53.025  53.880  1.00  2.18
ATOM   2475  HA  VAL A 152      64.069  52.799  54.923  1.00  2.18
ATOM   2476  CB  VAL A 152      65.596  52.254  53.590  1.00  2.18
ATOM   2477  HB  VAL A 152      65.995  52.537  52.616  1.00  2.18
ATOM   2478  CG1 VAL A 152      66.671  52.441  54.676  1.00  2.18
ATOM   2479 1HG1 VAL A 152      67.528  51.815  54.444  1.00  2.18
ATOM   2480 2HG1 VAL A 152      67.057  53.455  54.722  1.00  2.18
ATOM   2481 3HG1 VAL A 152      66.272  52.166  55.654  1.00  2.18
ATOM   2482  CG2 VAL A 152      65.280  50.748  53.604  1.00  2.18
ATOM   2483 1HG2 VAL A 152      66.208  50.189  53.622  1.00  2.18
ATOM   2484 2HG2 VAL A 152      64.705  50.482  54.493  1.00  2.18
ATOM   2485 3HG2 VAL A 152      64.731  50.465  52.707  1.00  2.18
ATOM   2486  C   VAL A 152     64.338  54.571  53.845  1.00  2.18
ATOM   2487  O   VAL A 152     63.314  55.245  53.713  1.00  2.18
ATOM   2488  N   MET A 153     65.512  55.165  53.950  1.00  2.28
```

```
ATOM   2489  H   MET A 153      66.291  54.603  54.266  1.00  2.28
ATOM   2490  CA  MET A 153      65.820  56.508  53.483  1.00  2.28
ATOM   2491  HA  MET A 153      65.039  56.877  52.817  1.00  2.28
ATOM   2492  CB  MET A 153      65.989  57.475  54.664  1.00  2.28
ATOM   2493  HB1 MET A 153      66.528  58.359  54.322  1.00  2.28
ATOM   2494  HB2 MET A 153      66.574  56.995  55.448  1.00  2.28
ATOM   2495  CG  MET A 153      64.640  57.917  55.235  1.00  2.28
ATOM   2496  HG1 MET A 153      64.069  57.026  55.488  1.00  2.28
ATOM   2497  HG2 MET A 153      64.097  58.459  54.461  1.00  2.28
ATOM   2498  SD  MET A 153      64.727  58.953  56.722  1.00  2.28
ATOM   2499  CE  MET A 153      65.428  60.482  56.041  1.00  2.28
ATOM   2500  HE1 MET A 153      66.397  60.277  55.592  1.00  2.28
ATOM   2501  HE2 MET A 153      64.766  60.890  55.279  1.00  2.28
ATOM   2502  HE3 MET A 153      65.546  61.214  56.842  1.00  2.28
ATOM   2503  C   MET A 153      67.133  56.371  52.706  1.00  2.28
ATOM   2504  O   MET A 153      68.081  55.720  53.145  1.00  2.28
ATOM   2505  N   SER A 154      67.156  56.924  51.509  1.00  2.66
ATOM   2506  H   SER A 154      66.325  57.419  51.211  1.00  2.66
ATOM   2507  CA  SER A 154      68.334  57.005  50.654  1.00  2.66
ATOM   2508  HA  SER A 154      69.063  56.244  50.932  1.00  2.66
ATOM   2509  CB  SER A 154      67.878  56.757  49.209  1.00  2.66
ATOM   2510  HB1 SER A 154      67.370  55.794  49.150  1.00  2.66
ATOM   2511  HB2 SER A 154      68.747  56.728  48.550  1.00  2.66
ATOM   2512  OG  SER A 154      66.992  57.778  48.775  1.00  2.66
ATOM   2513  HG  SER A 154      66.223  57.778  49.373  1.00  2.66
ATOM   2514  C   SER A 154      68.989  58.388  50.755  1.00  2.66
ATOM   2515  O   SER A 154      68.408  59.308  51.331  1.00  2.66
ATOM   2516  N   TYR A 155      70.148  58.597  50.116  1.00  2.61
ATOM   2517  H   TYR A 155      70.618  57.805  49.702  1.00  2.61
ATOM   2518  CA  TYR A 155      70.802  59.918  50.058  1.00  2.61
ATOM   2519  HA  TYR A 155      71.210  60.164  51.033  1.00  2.61
ATOM   2520  CB  TYR A 155      71.998  59.868  49.082  1.00  2.61
ATOM   2521  HB1 TYR A 155      71.629  59.915  48.057  1.00  2.61
ATOM   2522  HB2 TYR A 155      72.489  58.900  49.196  1.00  2.61
ATOM   2523  CG  TYR A 155      73.059  60.954  49.269  1.00  2.61
ATOM   2524  CD1 TYR A 155      74.367  60.588  49.650  1.00  2.61
ATOM   2525  HD1 TYR A 155      74.622  59.548  49.791  1.00  2.61
ATOM   2526  CE1 TYR A 155      75.358  61.573  49.835  1.00  2.61
ATOM   2527  HE1 TYR A 155      76.362  61.296  50.118  1.00  2.61
ATOM   2528  CZ  TYR A 155      75.050  62.936  49.658  1.00  2.61
ATOM   2529  OH  TYR A 155      76.005  63.873  49.895  1.00  2.61
ATOM   2530  HH  TYR A 155      75.689  64.767  49.747  1.00  2.61
ATOM   2531  CE2 TYR A 155      73.750  63.309  49.252  1.00  2.61
ATOM   2532  HE2 TYR A 155      73.510  64.351  49.101  1.00  2.61
ATOM   2533  CD2 TYR A 155      72.769  62.318  49.039  1.00  2.61
ATOM   2534  HD2 TYR A 155      71.789  62.623  48.705  1.00  2.61
ATOM   2535  C   TYR A 155      69.843  61.063  49.671  1.00  2.61
ATOM   2536  O   TYR A 155      69.887  62.131  50.277  1.00  2.61
ATOM   2537  N   LYS A 156      68.947  60.846  48.698  1.00  2.62
ATOM   2538  H   LYS A 156      68.905  59.930  48.276  1.00  2.62
ATOM   2539  CA  LYS A 156      67.976  61.869  48.282  1.00  2.62
ATOM   2540  HA  LYS A 156      68.515  62.799  48.091  1.00  2.62
ATOM   2541  CB  LYS A 156      67.267  61.450  46.983  1.00  2.62
```

```
ATOM   2542  HB1 LYS A 156     66.431  62.131  46.812  1.00  2.62
ATOM   2543  HB2 LYS A 156     66.869  60.438  47.086  1.00  2.62
ATOM   2544  CG  LYS A 156     68.212  61.530  45.772  1.00  2.62
ATOM   2545  HG1 LYS A 156     68.999  60.782  45.871  1.00  2.62
ATOM   2546  HG2 LYS A 156     68.671  62.520  45.749  1.00  2.62
ATOM   2547  CD  LYS A 156     67.460  61.307  44.452  1.00  2.62
ATOM   2548  HD1 LYS A 156     66.635  62.019  44.386  1.00  2.62
ATOM   2549  HD2 LYS A 156     67.060  60.291  44.432  1.00  2.62
ATOM   2550  CE  LYS A 156     68.413  61.520  43.267  1.00  2.62
ATOM   2551  HE1 LYS A 156     69.224  60.788  43.336  1.00  2.62
ATOM   2552  HE2 LYS A 156     68.858  62.517  43.357  1.00  2.62
ATOM   2553  NZ  LYS A 156     67.715  61.398  41.962  1.00  2.62
ATOM   2554  HZ1 LYS A 156     68.362  61.567  41.200  1.00  2.62
ATOM   2555  HZ2 LYS A 156     67.324  60.472  41.848  1.00  2.62
ATOM   2556  HZ3 LYS A 156     66.963  62.073  41.895  1.00  2.62
ATOM   2557  C   LYS A 156     66.947  62.235  49.355  1.00  2.62
ATOM   2558  O   LYS A 156     66.561  63.398  49.450  1.00  2.62
ATOM   2559  N   ASP A 157     66.552  61.264  50.172  1.00  2.45
ATOM   2560  H   ASP A 157     67.011  60.367  50.095  1.00  2.45
ATOM   2561  CA  ASP A 157     65.634  61.477  51.295  1.00  2.45
ATOM   2562  HA  ASP A 157     64.792  62.093  50.971  1.00  2.45
ATOM   2563  CB  ASP A 157     65.092  60.130  51.796  1.00  2.45
ATOM   2564  HB1 ASP A 157     64.367  60.309  52.590  1.00  2.45
ATOM   2565  HB2 ASP A 157     65.910  59.554  52.230  1.00  2.45
ATOM   2566  CG  ASP A 157     64.433  59.303  50.698  1.00  2.45
ATOM   2567  OD1 ASP A 157     63.541  59.793  49.981  1.00  2.45
ATOM   2568  OD2 ASP A 157     64.820  58.127  50.518  1.00  2.45
ATOM   2569  C   ASP A 157     66.335  62.179  52.461  1.00  2.45
ATOM   2570  O   ASP A 157     65.758  63.080  53.064  1.00  2.45
ATOM   2571  N   TYR A 158     67.594  61.826  52.754  1.00  2.36
ATOM   2572  H   TYR A 158     68.012  61.036  52.272  1.00  2.36
ATOM   2573  CA  TYR A 158     68.388  62.560  53.742  1.00  2.36
ATOM   2574  HA  TYR A 158     67.832  62.573  54.679  1.00  2.36
ATOM   2575  CB  TYR A 158     69.731  61.844  54.000  1.00  2.36
ATOM   2576  HB1 TYR A 158     70.285  62.426  54.737  1.00  2.36
ATOM   2577  HB2 TYR A 158     70.313  61.855  53.078  1.00  2.36
ATOM   2578  CG  TYR A 158     69.648  60.410  54.520  1.00  2.36
ATOM   2579  CD1 TYR A 158     70.229  59.353  53.791  1.00  2.36
ATOM   2580  HD1 TYR A 158     70.755  59.560  52.881  1.00  2.36
ATOM   2581  CE1 TYR A 158     70.129  58.021  54.235  1.00  2.36
ATOM   2582  HE1 TYR A 158     70.543  57.208  53.658  1.00  2.36
ATOM   2583  CZ  TYR A 158     69.454  57.731  55.437  1.00  2.36
ATOM   2584  OH  TYR A 158     69.358  56.439  55.858  1.00  2.36
ATOM   2585  HH  TYR A 158     68.825  56.351  56.655  1.00  2.36
ATOM   2586  CE2 TYR A 158     68.914  58.791  56.198  1.00  2.36
ATOM   2587  HE2 TYR A 158     68.415  58.583  57.128  1.00  2.36
ATOM   2588  CD2 TYR A 158     69.020  60.124  55.747  1.00  2.36
ATOM   2589  HD2 TYR A 158     68.608  60.925  56.346  1.00  2.36
ATOM   2590  C   TYR A 158     68.615  64.027  53.342  1.00  2.36
ATOM   2591  O   TYR A 158     68.477  64.920  54.183  1.00  2.36
ATOM   2592  N   PHE A 159     68.885  64.285  52.058  1.00  2.55
ATOM   2593  H   PHE A 159     69.014  63.496  51.433  1.00  2.55
ATOM   2594  CA  PHE A 159     68.974  65.637  51.501  1.00  2.55
```

```
ATOM   2595  HA  PHE A 159      69.717  66.196  52.071  1.00  2.55
ATOM   2596  CB  PHE A 159      69.446  65.577  50.040  1.00  2.55
ATOM   2597  HB1 PHE A 159      68.820  64.877  49.486  1.00  2.55
ATOM   2598  HB2 PHE A 159      70.470  65.201  50.012  1.00  2.55
ATOM   2599  CG  PHE A 159      69.394  66.931  49.357  1.00  2.55
ATOM   2600  CD1 PHE A 159      70.360  67.909  49.663  1.00  2.55
ATOM   2601  HD1 PHE A 159      71.172  67.673  50.334  1.00  2.55
ATOM   2602  CE1 PHE A 159      70.250  69.203  49.125  1.00  2.55
ATOM   2603  HE1 PHE A 159      70.982  69.959  49.372  1.00  2.55
ATOM   2604  CZ  PHE A 159      69.170  69.526  48.285  1.00  2.55
ATOM   2605  HZ  PHE A 159      69.070  70.529  47.892  1.00  2.55
ATOM   2606  CE2 PHE A 159      68.208  68.550  47.969  1.00  2.55
ATOM   2607  HE2 PHE A 159      67.369  68.810  47.338  1.00  2.55
ATOM   2608  CD2 PHE A 159      68.320  67.254  48.503  1.00  2.55
ATOM   2609  HD2 PHE A 159      67.558  66.518  48.285  1.00  2.55
ATOM   2610  C   PHE A 159     67.659  66.413  51.614  1.00  2.55
ATOM   2611  O   PHE A 159     67.665  67.556  52.071  1.00  2.55
ATOM   2612  N   TYR A 160     66.533  65.788  51.251  1.00  2.50
ATOM   2613  H   TYR A 160     66.600  64.869  50.829  1.00  2.50
ATOM   2614  CA  TYR A 160     65.206  66.383  51.401  1.00  2.50
ATOM   2615  HA  TYR A 160     65.161  67.290  50.796  1.00  2.50
ATOM   2616  CB  TYR A 160     64.137  65.410  50.888  1.00  2.50
ATOM   2617  HB1 TYR A 160      64.229  64.456  51.406  1.00  2.50
ATOM   2618  HB2 TYR A 160      64.301  65.227  49.825  1.00  2.50
ATOM   2619  CG  TYR A 160     62.734  65.942  51.089  1.00  2.50
ATOM   2620  CD1 TYR A 160      62.194  66.849  50.158  1.00  2.50
ATOM   2621  HD1 TYR A 160      62.759  67.125  49.279  1.00  2.50
ATOM   2622  CE1 TYR A 160      60.928  67.420  50.386  1.00  2.50
ATOM   2623  HE1 TYR A 160      60.515  68.133  49.691  1.00  2.50
ATOM   2624  CZ  TYR A 160     60.216  67.107  51.561  1.00  2.50
ATOM   2625  OH  TYR A 160      59.010  67.680  51.797  1.00  2.50
ATOM   2626  HH  TYR A 160      58.710  67.490  52.687  1.00  2.50
ATOM   2627  CE2 TYR A 160      60.760  66.202  52.496  1.00  2.50
ATOM   2628  HE2 TYR A 160      60.219  65.966  53.397  1.00  2.50
ATOM   2629  CD2 TYR A 160      62.012  65.610  52.253  1.00  2.50
ATOM   2630  HD2 TYR A 160      62.428  64.917  52.975  1.00  2.50
ATOM   2631  C   TYR A 160     64.927  66.791  52.847  1.00  2.50
ATOM   2632  O   TYR A 160     64.496  67.911  53.103  1.00  2.50
ATOM   2633  N   CYS A 161     65.210  65.904  53.798  1.00  2.35
ATOM   2634  H   CYS A 161     65.569  64.996  53.526  1.00  2.35
ATOM   2635  CA  CYS A 161     64.942  66.166  55.204  1.00  2.35
ATOM   2636  HA  CYS A 161     63.912  66.503  55.283  1.00  2.35
ATOM   2637  CB  CYS A 161     65.118  64.871  56.006  1.00  2.35
ATOM   2638  HB1 CYS A 161      64.974  65.085  57.067  1.00  2.35
ATOM   2639  HB2 CYS A 161      66.117  64.457  55.851  1.00  2.35
ATOM   2640  SG  CYS A 161     63.881  63.666  55.488  1.00  2.35
ATOM   2641  HG  CYS A 161     64.427  63.404  54.290  1.00  2.35
ATOM   2642  C   CYS A 161     65.828  67.264  55.786  1.00  2.35
ATOM   2643  O   CYS A 161     65.349  68.059  56.593  1.00  2.35
ATOM   2644  N   TRP A 162     67.095  67.333  55.356  1.00  2.41
ATOM   2645  H   TRP A 162     67.443  66.613  54.734  1.00  2.41
ATOM   2646  CA  TRP A 162      67.954  68.478  55.641  1.00  2.41
ATOM   2647  HA  TRP A 162      68.078  68.566  56.718  1.00  2.41
```

```
ATOM   2648  CB  TRP A 162      69.347  68.291  55.033  1.00  2.41
ATOM   2649  HB1 TRP A 162      69.260  68.164  53.956  1.00  2.41
ATOM   2650  HB2 TRP A 162      69.771  67.368  55.431  1.00  2.41
ATOM   2651  CG  TRP A 162      70.322  69.405  55.297  1.00  2.41
ATOM   2652  CD1 TRP A 162      70.376  70.595  54.651  1.00  2.41
ATOM   2653  HD1 TRP A 162      69.684  70.916  53.883  1.00  2.41
ATOM   2654  NE1 TRP A 162      71.438  71.342  55.127  1.00  2.41
ATOM   2655  HE1 TRP A 162      71.617  72.309  54.854  1.00  2.41
ATOM   2656  CE2 TRP A 162      72.148  70.649  56.085  1.00  2.41
ATOM   2657  CZ2 TRP A 162      73.317  70.923  56.805  1.00  2.41
ATOM   2658  HZ2 TRP A 162      73.851  71.840  56.636  1.00  2.41
ATOM   2659  CH2 TRP A 162      73.774  69.999  57.759  1.00  2.41
ATOM   2660  HH2 TRP A 162      74.670  70.208  58.327  1.00  2.41
ATOM   2661  CZ3 TRP A 162      73.044  68.822  57.998  1.00  2.41
ATOM   2662  HZ3 TRP A 162      73.374  68.127  58.759  1.00  2.41
ATOM   2663  CE3 TRP A 162      71.875  68.554  57.261  1.00  2.41
ATOM   2664  HE3 TRP A 162      71.306  67.662  57.458  1.00  2.41
ATOM   2665  CD2 TRP A 162      71.416  69.438  56.263  1.00  2.41
ATOM   2666  C   TRP A 162     67.293  69.767  55.155  1.00  2.41
ATOM   2667  O   TRP A 162     66.955  70.636  55.950  1.00  2.41
ATOM   2668  N   GLN A 163     67.029  69.856  53.854  1.00  2.56
ATOM   2669  H   GLN A 163     67.268  69.067  53.260  1.00  2.56
ATOM   2670  CA  GLN A 163     66.507  71.056  53.213  1.00  2.56
ATOM   2671  HA  GLN A 163     67.209  71.867  53.405  1.00  2.56
ATOM   2672  CB  GLN A 163     66.474  70.764  51.697  1.00  2.56
ATOM   2673  HB1 GLN A 163      65.780  69.942  51.513  1.00  2.56
ATOM   2674  HB2 GLN A 163      67.466  70.433  51.386  1.00  2.56
ATOM   2675  CG  GLN A 163     66.068  71.940  50.796  1.00  2.56
ATOM   2676  HG1 GLN A 163      65.043  72.241  51.013  1.00  2.56
ATOM   2677  HG2 GLN A 163      66.100  71.611  49.757  1.00  2.56
ATOM   2678  CD  GLN A 163     67.000  73.136  50.948  1.00  2.56
ATOM   2679  OE1 GLN A 163     68.016  73.256  50.285  1.00  2.56
ATOM   2680  NE2 GLN A 163     66.722  74.046  51.855  1.00  2.56
ATOM   2681 1HE2 GLN A 163      65.921  73.916  52.468  1.00  2.56
ATOM   2682 2HE2 GLN A 163      67.365  74.809  51.943  1.00  2.56
ATOM   2683  C   GLN A 163     65.131  71.518  53.729  1.00  2.56
ATOM   2684  O   GLN A 163     64.801  72.696  53.578  1.00  2.56
ATOM   2685  N   THR A 164     64.343  70.603  54.307  1.00  2.57
ATOM   2686  H   THR A 164     64.700  69.656  54.317  1.00  2.57
ATOM   2687  CA  THR A 164     62.888  70.759  54.454  1.00  2.57
ATOM   2688  HA  THR A 164     62.650  71.800  54.239  1.00  2.57
ATOM   2689  CB  THR A 164     62.117  69.922  53.407  1.00  2.57
ATOM   2690  HB  THR A 164     62.030  68.893  53.760  1.00  2.57
ATOM   2691  CG2 THR A 164      60.719  70.461  53.102  1.00  2.57
ATOM   2692 1HG2 THR A 164      60.333  69.995  52.197  1.00  2.57
ATOM   2693 2HG2 THR A 164      60.042  70.230  53.924  1.00  2.57
ATOM   2694 3HG2 THR A 164      60.760  71.538  52.952  1.00  2.57
ATOM   2695  OG1 THR A 164     62.793  69.901  52.166  1.00  2.57
ATOM   2696  HG1 THR A 164     63.411  69.160  52.221  1.00  2.57
ATOM   2697  C   THR A 164     62.347  70.528  55.869  1.00  2.57
ATOM   2698  O   THR A 164     61.171  70.786  56.125  1.00  2.57
ATOM   2699  N   PHE A 165     63.195  70.130  56.824  1.00  2.37
ATOM   2700  H   PHE A 165     64.136  69.862  56.569  1.00  2.37
```

```
ATOM   2701  CA  PHE A 165      62.817  70.093  58.242  1.00  2.37
ATOM   2702  HA  PHE A 165      61.954  70.742  58.392  1.00  2.37
ATOM   2703  CB  PHE A 165      62.409  68.681  58.679  1.00  2.37
ATOM   2704  HB1 PHE A 165      61.949  68.742  59.666  1.00  2.37
ATOM   2705  HB2 PHE A 165      63.316  68.082  58.787  1.00  2.37
ATOM   2706  CG  PHE A 165      61.452  67.952  57.765  1.00  2.37
ATOM   2707  CD1 PHE A 165      61.937  66.873  57.011  1.00  2.37
ATOM   2708  HD1 PHE A 165      62.965  66.576  57.121  1.00  2.37
ATOM   2709  CE1 PHE A 165      61.083  66.156  56.164  1.00  2.37
ATOM   2710  HE1 PHE A 165      61.462  65.325  55.587  1.00  2.37
ATOM   2711  CZ  PHE A 165      59.722  66.487  56.122  1.00  2.37
ATOM   2712  HZ  PHE A 165      59.066  65.899  55.510  1.00  2.37
ATOM   2713  CE2 PHE A 165      59.217  67.548  56.893  1.00  2.37
ATOM   2714  HE2 PHE A 165      58.162  67.787  56.864  1.00  2.37
ATOM   2715  CD2 PHE A 165      60.087  68.294  57.709  1.00  2.37
ATOM   2716  HD2 PHE A 165      59.708  69.123  58.292  1.00  2.37
ATOM   2717  C   PHE A 165     63.887  70.596  59.206  1.00  2.37
ATOM   2718  O   PHE A 165     63.571  70.746  60.386  1.00  2.37
ATOM   2719  N   VAL A 166     65.122  70.861  58.760  1.00  2.48
ATOM   2720  H   VAL A 166     65.334  70.778  57.774  1.00  2.48
ATOM   2721  CA  VAL A 166      66.168  71.405  59.642  1.00  2.48
ATOM   2722  HA  VAL A 166      65.865  71.205  60.668  1.00  2.48
ATOM   2723  CB  VAL A 166      67.554  70.737  59.518  1.00  2.48
ATOM   2724  HB  VAL A 166      68.066  70.924  60.463  1.00  2.48
ATOM   2725  CG1 VAL A 166      67.436  69.222  59.421  1.00  2.48
ATOM   2726 1HG1 VAL A 166      68.437  68.795  59.430  1.00  2.48
ATOM   2727 2HG1 VAL A 166      66.871  68.853  60.275  1.00  2.48
ATOM   2728 3HG1 VAL A 166      66.922  68.940  58.506  1.00  2.48
ATOM   2729  CG2 VAL A 166      68.499  71.274  58.432  1.00  2.48
ATOM   2730 1HG2 VAL A 166      69.314  71.817  58.906  1.00  2.48
ATOM   2731 2HG2 VAL A 166      68.933  70.467  57.845  1.00  2.48
ATOM   2732 3HG2 VAL A 166      67.983  71.958  57.761  1.00  2.48
ATOM   2733  C   VAL A 166     66.304  72.917  59.542  1.00  2.48
ATOM   2734  O   VAL A 166     65.865  73.541  58.579  1.00  2.48
ATOM   2735  N   ALA A 167     66.934  73.496  60.558  1.00  2.62
ATOM   2736  H   ALA A 167     67.203  72.915  61.343  1.00  2.62
ATOM   2737  CA  ALA A 167      67.202  74.924  60.650  1.00  2.62
ATOM   2738  HA  ALA A 167      66.492  75.453  60.013  1.00  2.62
ATOM   2739  CB  ALA A 167      66.926  75.346  62.104  1.00  2.62
ATOM   2740  HB1 ALA A 167      65.924  75.035  62.403  1.00  2.62
ATOM   2741  HB2 ALA A 167      67.655  74.884  62.770  1.00  2.62
ATOM   2742  HB3 ALA A 167      66.997  76.430  62.193  1.00  2.62
ATOM   2743  C   ALA A 167     68.610  75.326  60.173  1.00  2.62
ATOM   2744  O   ALA A 167     69.165  76.327  60.633  1.00  2.62
ATOM   2745  N   HIS A 168     69.180  74.524  59.267  1.00  2.88
ATOM   2746  CA  HIS A 168     70.498  74.514  58.639  1.00  2.88
ATOM   2747  CB  HIS A 168     71.477  73.567  59.354  1.00  2.88
ATOM   2748  CG  HIS A 168     72.642  74.185  60.102  1.00  2.88
ATOM   2749  ND1 HIS A 168      73.991  73.969  59.806  1.00  2.88
ATOM   2750  CE1 HIS A 168     74.682  74.631  60.740  1.00  2.88
ATOM   2751  NE2 HIS A 168     73.843  75.180  61.636  1.00  2.88
ATOM   2752  CD2 HIS A 168     72.546  74.920  61.251  1.00  2.88
ATOM   2753  C   HIS A 168     71.138  75.762  58.061  1.00  2.88
```

```
ATOM   2754  O   HIS A 168      72.055  75.607  57.278  1.00  2.88
ATOM   2755  N   ARG A 169      70.700  76.980  58.390  1.00  3.07
ATOM   2756  H   ARG A 169      69.945  76.976  59.068  1.00  3.07
ATOM   2757  CA  ARG A 169      71.389  78.266  58.132  1.00  3.07
ATOM   2758  HA  ARG A 169      70.620  79.038  58.122  1.00  3.07
ATOM   2759  CB  ARG A 169      72.330  78.579  59.318  1.00  3.07
ATOM   2760  HB1 ARG A 169      72.812  79.541  59.136  1.00  3.07
ATOM   2761  HB2 ARG A 169      73.105  77.812  59.368  1.00  3.07
ATOM   2762  CG  ARG A 169      71.610  78.670  60.674  1.00  3.07
ATOM   2763  HG1 ARG A 169      71.223  77.692  60.954  1.00  3.07
ATOM   2764  HG2 ARG A 169      70.776  79.369  60.589  1.00  3.07
ATOM   2765  CD  ARG A 169      72.546  79.156  61.787  1.00  3.07
ATOM   2766  HD1 ARG A 169      71.946  79.353  62.678  1.00  3.07
ATOM   2767  HD2 ARG A 169      73.016  80.088  61.469  1.00  3.07
ATOM   2768  NE  ARG A 169      73.565  78.141  62.113  1.00  3.07
ATOM   2769  HE  ARG A 169      73.331  77.182  61.868  1.00  3.07
ATOM   2770  CZ  ARG A 169      74.742  78.327  62.678  1.00  3.07
ATOM   2771  NH1 ARG A 169      75.505  77.300  62.909  1.00  3.07
ATOM   2772 1HH1 ARG A 169      75.163  76.391  62.618  1.00  3.07
ATOM   2773 2HH1 ARG A 169      76.424  77.411  63.294  1.00  3.07
ATOM   2774  NH2 ARG A 169      75.181  79.506  63.030  1.00  3.07
ATOM   2775 1HH2 ARG A 169      74.595  80.305  62.870  1.00  3.07
ATOM   2776 2HH2 ARG A 169      76.087  79.611  63.442  1.00  3.07
ATOM   2777  C   ARG A 169      72.114  78.430  56.775  1.00  3.07
ATOM   2778  O   ARG A 169      73.245  78.907  56.729  1.00  3.07
ATOM   2779  N   LEU A 170      71.486  78.028  55.666  1.00  3.15
ATOM   2780  H   LEU A 170      70.614  77.546  55.814  1.00  3.15
ATOM   2781  CA  LEU A 170      72.087  77.956  54.315  1.00  3.15
ATOM   2782  HA  LEU A 170      71.343  77.466  53.686  1.00  3.15
ATOM   2783  CB  LEU A 170      72.306  79.365  53.714  1.00  3.15
ATOM   2784  HB1 LEU A 170      72.664  79.246  52.691  1.00  3.15
ATOM   2785  HB2 LEU A 170      73.092  79.869  54.277  1.00  3.15
ATOM   2786  CG  LEU A 170      71.065  80.280  53.679  1.00  3.15
ATOM   2787  HG  LEU A 170      70.745  80.497  54.699  1.00  3.15
ATOM   2788  CD1 LEU A 170      71.429  81.601  53.000  1.00  3.15
ATOM   2789 1HD1 LEU A 170      70.567  82.267  53.004  1.00  3.15
ATOM   2790 2HD1 LEU A 170      72.241  82.080  53.546  1.00  3.15
ATOM   2791 3HD1 LEU A 170      71.741  81.425  51.970  1.00  3.15
ATOM   2792  CD2 LEU A 170      69.889  79.671  52.914  1.00  3.15
ATOM   2793 1HD2 LEU A 170      69.075  80.394  52.855  1.00  3.15
ATOM   2794 2HD2 LEU A 170      70.196  79.400  51.904  1.00  3.15
ATOM   2795 3HD2 LEU A 170      69.517  78.790  53.436  1.00  3.15
ATOM   2796  C   LEU A 170      73.323  77.035  54.184  1.00  3.15
ATOM   2797  O   LEU A 170      73.976  77.011  53.136  1.00  3.15
ATOM   2798  N   SER A 171      73.630  76.246  55.213  1.00  2.98
ATOM   2799  H   SER A 171      73.058  76.350  56.043  1.00  2.98
ATOM   2800  CA  SER A 171      74.435  75.024  55.174  1.00  2.98
ATOM   2801  HA  SER A 171      75.403  75.288  54.751  1.00  2.98
ATOM   2802  CB  SER A 171      74.692  74.506  56.589  1.00  2.98
ATOM   2803  HB1 SER A 171      73.812  73.987  56.967  1.00  2.98
ATOM   2804  HB2 SER A 171      74.923  75.343  57.251  1.00  2.98
ATOM   2805  OG  SER A 171      75.793  73.623  56.563  1.00  2.98
ATOM   2806  HG  SER A 171      75.921  73.313  57.468  1.00  2.98
```

```
ATOM   2807  C   SER A 171     73.815  73.953  54.281  1.00  2.98
ATOM   2808  O   SER A 171     72.622  73.959  53.974  1.00  2.98
ATOM   2809  N   ARG A 172     74.653  73.018  53.837  1.00  2.83
ATOM   2810  H   ARG A 172     75.572  72.985  54.265  1.00  2.83
ATOM   2811  CA  ARG A 172     74.320  72.011  52.827  1.00  2.83
ATOM   2812  HA  ARG A 172     73.240  71.991  52.675  1.00  2.83
ATOM   2813  CB  ARG A 172     74.997  72.373  51.494  1.00  2.83
ATOM   2814  HB1 ARG A 172     75.073  71.476  50.876  1.00  2.83
ATOM   2815  HB2 ARG A 172     76.013  72.722  51.691  1.00  2.83
ATOM   2816  CG  ARG A 172     74.197  73.430  50.704  1.00  2.83
ATOM   2817  HG1 ARG A 172     73.427  73.885  51.328  1.00  2.83
ATOM   2818  HG2 ARG A 172     73.690  72.927  49.880  1.00  2.83
ATOM   2819  CD  ARG A 172     75.078  74.545  50.124  1.00  2.83
ATOM   2820  HD1 ARG A 172     74.516  75.046  49.333  1.00  2.83
ATOM   2821  HD2 ARG A 172     75.969  74.102  49.677  1.00  2.83
ATOM   2822  NE  ARG A 172     75.420  75.547  51.154  1.00  2.83
ATOM   2823  HE  ARG A 172     74.655  75.873  51.737  1.00  2.83
ATOM   2824  CZ  ARG A 172     76.585  76.115  51.401  1.00  2.83
ATOM   2825  NH1 ARG A 172     76.664  77.025  52.327  1.00  2.83
ATOM   2826 1HH1 ARG A 172     75.802  77.300  52.788  1.00  2.83
ATOM   2827 2HH1 ARG A 172     77.526  77.481  52.551  1.00  2.83
ATOM   2828  NH2 ARG A 172     77.672  75.811  50.750  1.00  2.83
ATOM   2829 1HH2 ARG A 172     77.617  75.114  50.033  1.00  2.83
ATOM   2830 2HH2 ARG A 172     78.537  76.269  50.965  1.00  2.83
ATOM   2831  C   ARG A 172     74.726  70.636  53.337  1.00  2.83
ATOM   2832  O   ARG A 172     75.852  70.454  53.793  1.00  2.83
ATOM   2833  N   PHE A 173     73.776  69.702  53.268  1.00  2.76
ATOM   2834  H   PHE A 173     72.884  69.974  52.890  1.00  2.76
ATOM   2835  CA  PHE A 173     73.852  68.364  53.851  1.00  2.76
ATOM   2836  HA  PHE A 173     73.718  68.449  54.931  1.00  2.76
ATOM   2837  CB  PHE A 173     72.702  67.500  53.294  1.00  2.76
ATOM   2838  HB1 PHE A 173     72.711  67.575  52.206  1.00  2.76
ATOM   2839  HB2 PHE A 173     71.751  67.900  53.634  1.00  2.76
ATOM   2840  CG  PHE A 173     72.780  66.029  53.676  1.00  2.76
ATOM   2841  CD1 PHE A 173     72.545  65.626  55.004  1.00  2.76
ATOM   2842  HD1 PHE A 173     72.271  66.358  55.747  1.00  2.76
ATOM   2843  CE1 PHE A 173     72.682  64.275  55.368  1.00  2.76
ATOM   2844  HE1 PHE A 173     72.502  63.964  56.386  1.00  2.76
ATOM   2845  CZ  PHE A 173     73.056  63.322  54.408  1.00  2.76
ATOM   2846  HZ  PHE A 173     73.160  62.289  54.700  1.00  2.76
ATOM   2847  CE2 PHE A 173     73.287  63.712  53.078  1.00  2.76
ATOM   2848  HE2 PHE A 173     73.581  62.981  52.335  1.00  2.76
ATOM   2849  CD2 PHE A 173     73.146  65.064  52.715  1.00  2.76
ATOM   2850  HD2 PHE A 173     73.345  65.370  51.698  1.00  2.76
ATOM   2851  C   PHE A 173     75.194  67.658  53.590  1.00  2.76
ATOM   2852  O   PHE A 173     75.464  67.143  52.501  1.00  2.76
ATOM   2853  N   LYS A 174     76.017  67.567  54.634  1.00  3.31
ATOM   2854  H   LYS A 174     75.772  68.080  55.468  1.00  3.31
ATOM   2855  CA  LYS A 174     77.196  66.706  54.657  1.00  3.31
ATOM   2856  HA  LYS A 174     77.640  66.683  53.660  1.00  3.31
ATOM   2857  CB  LYS A 174     78.217  67.314  55.636  1.00  3.31
ATOM   2858  HB1 LYS A 174     77.797  67.284  56.643  1.00  3.31
ATOM   2859  HB2 LYS A 174     78.402  68.359  55.376  1.00  3.31
```

```
ATOM   2860  CG  LYS A 174      79.550  66.552  55.599  1.00  3.31
ATOM   2861  HG1 LYS A 174      80.133  66.877  54.736  1.00  3.31
ATOM   2862  HG2 LYS A 174      79.346  65.495  55.473  1.00  3.31
ATOM   2863  CD  LYS A 174      80.372  66.746  56.882  1.00  3.31
ATOM   2864  HD1 LYS A 174      79.719  66.719  57.756  1.00  3.31
ATOM   2865  HD2 LYS A 174      80.860  67.721  56.846  1.00  3.31
ATOM   2866  CE  LYS A 174      81.432  65.644  57.019  1.00  3.31
ATOM   2867  HE1 LYS A 174      82.111  65.911  57.834  1.00  3.31
ATOM   2868  HE2 LYS A 174      82.014  65.603  56.093  1.00  3.31
ATOM   2869  NZ  LYS A 174      80.809  64.325  57.297  1.00  3.31
ATOM   2870  HZ1 LYS A 174      81.493  63.584  57.292  1.00  3.31
ATOM   2871  HZ2 LYS A 174      80.352  64.320  58.205  1.00  3.31
ATOM   2872  HZ3 LYS A 174      80.100  64.095  56.601  1.00  3.31
ATOM   2873  C   LYS A 174     76.810  65.278  55.054  1.00  3.31
ATOM   2874  O   LYS A 174     76.504  65.020  56.218  1.00  3.31
ATOM   2875  N   ALA A 175     76.917  64.351  54.103  1.00  2.88
ATOM   2876  H   ALA A 175     77.104  64.659  53.162  1.00  2.88
ATOM   2877  CA  ALA A 175      76.841  62.912  54.347  1.00  2.88
ATOM   2878  HA  ALA A 175      75.850  62.693  54.746  1.00  2.88
ATOM   2879  CB  ALA A 175      76.989  62.192  53.003  1.00  2.88
ATOM   2880  HB1 ALA A 175      76.911  61.112  53.140  1.00  2.88
ATOM   2881  HB2 ALA A 175      76.198  62.520  52.332  1.00  2.88
ATOM   2882  HB3 ALA A 175      77.957  62.424  52.556  1.00  2.88
ATOM   2883  C   ALA A 175     77.874  62.415  55.368  1.00  2.88
ATOM   2884  O   ALA A 175     78.908  63.047  55.627  1.00  2.88
ATOM   2885  N   TRP A 176     77.610  61.231  55.913  1.00  2.77
ATOM   2886  H   TRP A 176     76.781  60.731  55.624  1.00  2.77
ATOM   2887  CA  TRP A 176      78.432  60.621  56.944  1.00  2.77
ATOM   2888  HA  TRP A 176      79.377  61.155  57.030  1.00  2.77
ATOM   2889  CB  TRP A 176      77.727  60.727  58.301  1.00  2.77
ATOM   2890  HB1 TRP A 176      78.255  60.101  59.021  1.00  2.77
ATOM   2891  HB2 TRP A 176      76.706  60.351  58.218  1.00  2.77
ATOM   2892  CG  TRP A 176      77.712  62.128  58.829  1.00  2.77
ATOM   2893  CD1 TRP A 176      76.698  63.014  58.713  1.00  2.77
ATOM   2894  HD1 TRP A 176      75.743  62.807  58.242  1.00  2.77
ATOM   2895  NE1 TRP A 176      77.089  64.239  59.220  1.00  2.77
ATOM   2896  HE1 TRP A 176      76.493  65.058  59.191  1.00  2.77
ATOM   2897  CE2 TRP A 176      78.392  64.211  59.662  1.00  2.77
ATOM   2898  CZ2 TRP A 176      79.250  65.185  60.190  1.00  2.77
ATOM   2899  HZ2 TRP A 176      78.892  66.193  60.351  1.00  2.77
ATOM   2900  CH2 TRP A 176      80.558  64.823  60.554  1.00  2.77
ATOM   2901  HH2 TRP A 176      81.218  65.550  61.011  1.00  2.77
ATOM   2902  CZ3 TRP A 176      80.995  63.499  60.366  1.00  2.77
ATOM   2903  HZ3 TRP A 176      81.992  63.214  60.683  1.00  2.77
ATOM   2904  CE3 TRP A 176      80.130  62.530  59.819  1.00  2.77
ATOM   2905  HE3 TRP A 176      80.468  61.507  59.706  1.00  2.77
ATOM   2906  CD2 TRP A 176      78.806  62.861  59.454  1.00  2.77
ATOM   2907  C   TRP A 176     78.756  59.187  56.556  1.00  2.77
ATOM   2908  O   TRP A 176     77.869  58.336  56.464  1.00  2.77
ATOM   2909  N   GLU A 177     80.061  58.940  56.392  1.00  2.88
ATOM   2910  H   GLU A 177     80.701  59.712  56.354  1.00  2.88
ATOM   2911  CA  GLU A 177      80.633  57.595  56.460  1.00  2.88
ATOM   2912  HA  GLU A 177      81.706  57.622  56.271  1.00  2.88
```

```
ATOM   2913  CB  GLU A 177      80.388  57.170  57.942  1.00  2.88
ATOM   2914  HB1 GLU A 177      79.311  57.092  58.088  1.00  2.88
ATOM   2915  HB2 GLU A 177      80.724  57.994  58.576  1.00  2.88
ATOM   2916  CG  GLU A 177      80.981  55.898  58.560  1.00  2.88
ATOM   2917  HG1 GLU A 177      82.054  56.032  58.712  1.00  2.88
ATOM   2918  HG2 GLU A 177      80.830  55.032  57.918  1.00  2.88
ATOM   2919  CD  GLU A 177      80.258  55.678  59.901  1.00  2.88
ATOM   2920  OE1 GLU A 177      79.218  54.976  59.939  1.00  2.88
ATOM   2921  OE2 GLU A 177      80.625  56.327  60.901  1.00  2.88
ATOM   2922  C   GLU A 177     79.981  56.686  55.401  1.00  2.88
ATOM   2923  O   GLU A 177     79.755  57.107  54.267  1.00  2.88
ATOM   2924  N   GLY A 178     79.609  55.461  55.763  1.00  3.09
ATOM   2925  H   GLY A 178     79.779  55.169  56.715  1.00  3.09
ATOM   2926  CA  GLY A 178      78.921  54.534  54.880  1.00  3.09
ATOM   2927  HA1 GLY A 178      79.057  53.521  55.259  1.00  3.09
ATOM   2928  HA2 GLY A 178      79.374  54.582  53.889  1.00  3.09
ATOM   2929  C   GLY A 178     77.433  54.788  54.722  1.00  3.09
ATOM   2930  O   GLY A 178     76.704  53.808  54.657  1.00  3.09
ATOM   2931  N   LEU A 179     76.970  56.043  54.660  1.00  2.83
ATOM   2932  H   LEU A 179     77.636  56.800  54.786  1.00  2.83
ATOM   2933  CA  LEU A 179      75.550  56.384  54.472  1.00  2.83
ATOM   2934  HA  LEU A 179      75.049  56.218  55.426  1.00  2.83
ATOM   2935  CB  LEU A 179      75.449  57.881  54.099  1.00  2.83
ATOM   2936  HB1 LEU A 179      75.685  57.993  53.040  1.00  2.83
ATOM   2937  HB2 LEU A 179      76.195  58.454  54.645  1.00  2.83
ATOM   2938  CG  LEU A 179      74.070  58.504  54.381  1.00  2.83
ATOM   2939  HG  LEU A 179      73.285  57.814  54.070  1.00  2.83
ATOM   2940  CD1 LEU A 179      73.914  58.825  55.871  1.00  2.83
ATOM   2941 1HD1 LEU A 179       72.924  59.245  56.052  1.00  2.83
ATOM   2942 2HD1 LEU A 179       74.006  57.914  56.463  1.00  2.83
ATOM   2943 3HD1 LEU A 179       74.676  59.535  56.191  1.00  2.83
ATOM   2944  CD2 LEU A 179      73.915  59.808  53.596  1.00  2.83
ATOM   2945 1HD2 LEU A 179       72.931  60.234  53.791  1.00  2.83
ATOM   2946 2HD2 LEU A 179       74.682  60.519  53.894  1.00  2.83
ATOM   2947 3HD2 LEU A 179       74.000  59.614  52.526  1.00  2.83
ATOM   2948  C   LEU A 179     74.856  55.509  53.416  1.00  2.83
ATOM   2949  O   LEU A 179     73.837  54.876  53.698  1.00  2.83
ATOM   2950  N   HIS A 180     75.460  55.420  52.225  1.00  2.86
ATOM   2951  H   HIS A 180     76.309  55.951  52.094  1.00  2.86
ATOM   2952  CA  HIS A 180      74.969  54.589  51.120  1.00  2.86
ATOM   2953  HA  HIS A 180      73.919  54.827  50.935  1.00  2.86
ATOM   2954  CB  HIS A 180      75.752  54.928  49.844  1.00  2.86
ATOM   2955  HB1 HIS A 180       76.811  54.710  49.997  1.00  2.86
ATOM   2956  HB2 HIS A 180       75.650  55.993  49.634  1.00  2.86
ATOM   2957  CG  HIS A 180      75.265  54.166  48.636  1.00  2.86
ATOM   2958  ND1 HIS A 180       73.981  54.247  48.096  1.00  2.86
ATOM   2959  CE1 HIS A 180       73.962  53.386  47.068  1.00  2.86
ATOM   2960  HE1 HIS A 180       73.105  53.213  46.431  1.00  2.86
ATOM   2961  NE2 HIS A 180       75.153  52.778  46.939  1.00  2.86
ATOM   2962  HE2 HIS A 180       75.385  52.094  46.232  1.00  2.86
ATOM   2963  CD2 HIS A 180       75.991  53.259  47.921  1.00  2.86
ATOM   2964  HD2 HIS A 180       77.018  52.978  48.102  1.00  2.86
ATOM   2965  C   HIS A 180     75.036  53.090  51.423  1.00  2.86
```

```
ATOM   2966  O   HIS A 180     74.081  52.368  51.157  1.00  2.86
ATOM   2967  N   THR A 181     76.127  52.594  52.007  1.00  2.85
ATOM   2968  H   THR A 181     76.888  53.215  52.247  1.00  2.85
ATOM   2969  CA  THR A 181     76.307  51.157  52.276  1.00  2.85
ATOM   2970  HA  THR A 181     76.132  50.619  51.345  1.00  2.85
ATOM   2971  CB  THR A 181     77.741  50.865  52.747  1.00  2.85
ATOM   2972  HB  THR A 181     77.812  51.076  53.815  1.00  2.85
ATOM   2973  CG2 THR A 181     78.149  49.414  52.493  1.00  2.85
ATOM   2974 1HG2 THR A 181     79.160  49.250  52.867  1.00  2.85
ATOM   2975 2HG2 THR A 181     77.462  48.740  53.001  1.00  2.85
ATOM   2976 3HG2 THR A 181     78.123  49.202  51.423  1.00  2.85
ATOM   2977  OG1 THR A 181     78.684  51.677  52.088  1.00  2.85
ATOM   2978  HG1 THR A 181     78.914  51.263  51.250  1.00  2.85
ATOM   2979  C   THR A 181     75.316  50.610  53.311  1.00  2.85
ATOM   2980  O   THR A 181     74.773  49.505  53.175  1.00  2.85
ATOM   2981  N   ASN A 182     75.079  51.410  54.349  1.00  2.56
ATOM   2982  H   ASN A 182     75.566  52.303  54.362  1.00  2.56
ATOM   2983  CA  ASN A 182     74.083  51.195  55.382  1.00  2.56
ATOM   2984  HA  ASN A 182     74.211  50.213  55.826  1.00  2.56
ATOM   2985  CB  ASN A 182     74.239  52.279  56.464  1.00  2.56
ATOM   2986  HB1 ASN A 182     73.410  52.207  57.167  1.00  2.56
ATOM   2987  HB2 ASN A 182     74.178  53.253  55.980  1.00  2.56
ATOM   2988  CG  ASN A 182     75.530  52.217  57.264  1.00  2.56
ATOM   2989  OD1 ASN A 182     76.101  51.151  57.489  1.00  2.56
ATOM   2990  ND2 ASN A 182     76.018  53.340  57.741  1.00  2.56
ATOM   2991 1HD2 ASN A 182     75.498  54.200  57.587  1.00  2.56
ATOM   2992 2HD2 ASN A 182     76.890  53.333  58.243  1.00  2.56
ATOM   2993  C   ASN A 182     72.684  51.229  54.768  1.00  2.56
ATOM   2994  O   ASN A 182     71.938  50.279  54.977  1.00  2.56
ATOM   2995  N   TYR A 183     72.358  52.233  53.944  1.00  2.53
ATOM   2996  H   TYR A 183     73.004  53.006  53.830  1.00  2.53
ATOM   2997  CA  TYR A 183     71.102  52.270  53.180  1.00  2.53
ATOM   2998  HA  TYR A 183     70.277  52.347  53.890  1.00  2.53
ATOM   2999  CB  TYR A 183     71.058  53.537  52.297  1.00  2.53
ATOM   3000  HB1 TYR A 183     72.074  53.805  52.015  1.00  2.53
ATOM   3001  HB2 TYR A 183     70.675  54.362  52.898  1.00  2.53
ATOM   3002  CG  TYR A 183     70.237  53.444  51.014  1.00  2.53
ATOM   3003  CD1 TYR A 183     68.830  53.372  51.065  1.00  2.53
ATOM   3004  HD1 TYR A 183     68.330  53.390  52.023  1.00  2.53
ATOM   3005  CE1 TYR A 183     68.079  53.274  49.875  1.00  2.53
ATOM   3006  HE1 TYR A 183     67.003  53.195  49.909  1.00  2.53
ATOM   3007  CZ  TYR A 183     68.732  53.261  48.623  1.00  2.53
ATOM   3008  OH  TYR A 183     68.006  53.223  47.473  1.00  2.53
ATOM   3009  HH  TYR A 183     68.563  53.187  46.693  1.00  2.53
ATOM   3010  CE2 TYR A 183     70.141  53.322  48.571  1.00  2.53
ATOM   3011  HE2 TYR A 183     70.652  53.304  47.621  1.00  2.53
ATOM   3012  CD2 TYR A 183     70.887  53.409  49.763  1.00  2.53
ATOM   3013  HD2 TYR A 183     71.968  53.447  49.719  1.00  2.53
ATOM   3014  C   TYR A 183     70.827  51.006  52.354  1.00  2.53
ATOM   3015  O   TYR A 183     69.719  50.470  52.392  1.00  2.53
ATOM   3016  N   VAL A 184     71.828  50.482  51.640  1.00  2.87
ATOM   3017  H   VAL A 184     72.691  51.014  51.575  1.00  2.87
ATOM   3018  CA  VAL A 184     71.680  49.249  50.850  1.00  2.87
```

```
ATOM   3019  HA  VAL A 184      70.795  49.349  50.220  1.00  2.87
ATOM   3020  CB  VAL A 184      72.896  49.046  49.920  1.00  2.87
ATOM   3021  HB  VAL A 184      73.814  49.129  50.504  1.00  2.87
ATOM   3022  CG1 VAL A 184      72.890  47.679  49.219  1.00  2.87
ATOM   3023 1HG1 VAL A 184      73.718  47.628  48.511  1.00  2.87
ATOM   3024 2HG1 VAL A 184      73.017  46.880  49.948  1.00  2.87
ATOM   3025 3HG1 VAL A 184      71.948  47.543  48.685  1.00  2.87
ATOM   3026  CG2 VAL A 184      72.912  50.111  48.816  1.00  2.87
ATOM   3027 1HG2 VAL A 184      73.807  50.003  48.204  1.00  2.87
ATOM   3028 2HG2 VAL A 184      72.029  50.016  48.183  1.00  2.87
ATOM   3029 3HG2 VAL A 184      72.920  51.111  49.249  1.00  2.87
ATOM   3030  C   VAL A 184     71.450  48.029  51.743  1.00  2.87
ATOM   3031  O   VAL A 184     70.564  47.218  51.453  1.00  2.87
ATOM   3032  N   ARG A 185     72.211  47.881  52.839  1.00  2.80
ATOM   3033  H   ARG A 185     72.917  48.588  53.031  1.00  2.80
ATOM   3034  CA  ARG A 185      72.012  46.759  53.783  1.00  2.80
ATOM   3035  HA  ARG A 185      72.048  45.826  53.225  1.00  2.80
ATOM   3036  CB  ARG A 185      73.146  46.744  54.829  1.00  2.80
ATOM   3037  HB1 ARG A 185      72.841  46.111  55.665  1.00  2.80
ATOM   3038  HB2 ARG A 185      73.320  47.754  55.204  1.00  2.80
ATOM   3039  CG  ARG A 185      74.445  46.165  54.235  1.00  2.80
ATOM   3040  HG1 ARG A 185      74.792  46.788  53.409  1.00  2.80
ATOM   3041  HG2 ARG A 185      74.213  45.180  53.829  1.00  2.80
ATOM   3042  CD  ARG A 185      75.577  45.982  55.266  1.00  2.80
ATOM   3043  HD1 ARG A 185      76.127  45.079  54.991  1.00  2.80
ATOM   3044  HD2 ARG A 185      75.158  45.806  56.258  1.00  2.80
ATOM   3045  NE  ARG A 185      76.544  47.102  55.283  1.00  2.80
ATOM   3046  HE  ARG A 185      77.321  47.027  54.647  1.00  2.80
ATOM   3047  CZ  ARG A 185      76.524  48.181  56.045  1.00  2.80
ATOM   3048  NH1 ARG A 185      77.437  49.100  55.942  1.00  2.80
ATOM   3049 1HH1 ARG A 185      78.210  49.006  55.312  1.00  2.80
ATOM   3050 2HH1 ARG A 185      77.336  49.941  56.500  1.00  2.80
ATOM   3051  NH2 ARG A 185      75.612  48.407  56.943  1.00  2.80
ATOM   3052 1HH2 ARG A 185      74.854  47.763  57.044  1.00  2.80
ATOM   3053 2HH2 ARG A 185      75.621  49.311  57.398  1.00  2.80
ATOM   3054  C   ARG A 185     70.626  46.769  54.435  1.00  2.80
ATOM   3055  O   ARG A 185     69.969  45.728  54.498  1.00  2.80
ATOM   3056  N   LEU A 186     70.159  47.943  54.850  1.00  2.69
ATOM   3057  H   LEU A 186     70.760  48.755  54.739  1.00  2.69
ATOM   3058  CA  LEU A 186      68.816  48.160  55.373  1.00  2.69
ATOM   3059  HA  LEU A 186      68.653  47.522  56.241  1.00  2.69
ATOM   3060  CB  LEU A 186      68.694  49.638  55.798  1.00  2.69
ATOM   3061  HB1 LEU A 186      67.638  49.887  55.905  1.00  2.69
ATOM   3062  HB2 LEU A 186      69.093  50.254  54.993  1.00  2.69
ATOM   3063  CG  LEU A 186      69.411  50.018  57.108  1.00  2.69
ATOM   3064  HG  LEU A 186      70.386  49.533  57.154  1.00  2.69
ATOM   3065  CD1 LEU A 186      69.616  51.531  57.176  1.00  2.69
ATOM   3066 1HD1 LEU A 186      70.098  51.799  58.115  1.00  2.69
ATOM   3067 2HD1 LEU A 186      70.249  51.862  56.355  1.00  2.69
ATOM   3068 3HD1 LEU A 186      68.658  52.047  57.113  1.00  2.69
ATOM   3069  CD2 LEU A 186      68.585  49.597  58.328  1.00  2.69
ATOM   3070 1HD2 LEU A 186      69.094  49.916  59.237  1.00  2.69
ATOM   3071 2HD2 LEU A 186      67.602  50.067  58.282  1.00  2.69
```

```
ATOM   3072 3HD2 LEU A 186     68.481  48.513  58.344  1.00  2.69
ATOM   3073  C   LEU A 186     67.745  47.820  54.343  1.00  2.69
ATOM   3074  O   LEU A 186     66.838  47.071  54.665  1.00  2.69
ATOM   3075  N   SER A 187     67.883  48.283  53.099  1.00  2.78
ATOM   3076  H   SER A 187     68.660  48.907  52.909  1.00  2.78
ATOM   3077  CA  SER A 187     66.917  48.039  52.021  1.00  2.78
ATOM   3078  HA  SER A 187     65.930  48.356  52.356  1.00  2.78
ATOM   3079  CB  SER A 187     67.264  48.846  50.764  1.00  2.78
ATOM   3080  HB1 SER A 187     66.550  48.598  49.976  1.00  2.78
ATOM   3081  HB2 SER A 187     68.269  48.592  50.424  1.00  2.78
ATOM   3082  OG  SER A 187     67.185  50.234  51.022  1.00  2.78
ATOM   3083  HG  SER A 187     68.005  50.509  51.461  1.00  2.78
ATOM   3084  C   SER A 187     66.794  46.566  51.671  1.00  2.78
ATOM   3085  O   SER A 187     65.694  46.069  51.431  1.00  2.78
ATOM   3086  N   ARG A 188     67.908  45.825  51.720  1.00  2.98
ATOM   3087  H   ARG A 188     68.793  46.298  51.886  1.00  2.98
ATOM   3088  CA  ARG A 188     67.885  44.363  51.620  1.00  2.98
ATOM   3089  HA  ARG A 188     67.297  44.085  50.744  1.00  2.98
ATOM   3090  CB  ARG A 188     69.320  43.834  51.438  1.00  2.98
ATOM   3091  HB1 ARG A 188     69.308  42.746  51.523  1.00  2.98
ATOM   3092  HB2 ARG A 188     69.955  44.232  52.231  1.00  2.98
ATOM   3093  CG  ARG A 188     69.912  44.208  50.065  1.00  2.98
ATOM   3094  HG1 ARG A 188     69.870  45.287  49.920  1.00  2.98
ATOM   3095  HG2 ARG A 188     69.325  43.731  49.279  1.00  2.98
ATOM   3096  CD  ARG A 188     71.373  43.747  49.956  1.00  2.98
ATOM   3097  HD1 ARG A 188     71.395  42.662  50.073  1.00  2.98
ATOM   3098  HD2 ARG A 188     71.943  44.200  50.769  1.00  2.98
ATOM   3099  NE  ARG A 188     71.986  44.128  48.663  1.00  2.98
ATOM   3100  HE  ARG A 188     71.526  44.856  48.140  1.00  2.98
ATOM   3101  CZ  ARG A 188     73.090  43.617  48.138  1.00  2.98
ATOM   3102  NH1 ARG A 188     73.557  44.045  46.995  1.00  2.98
ATOM   3103 1HH1 ARG A 188     73.047  44.727  46.462  1.00  2.98
ATOM   3104 2HH1 ARG A 188     74.395  43.647  46.604  1.00  2.98
ATOM   3105  NH2 ARG A 188     73.766  42.678  48.742  1.00  2.98
ATOM   3106 1HH2 ARG A 188     73.436  42.316  49.618  1.00  2.98
ATOM   3107 2HH2 ARG A 188     74.590  42.290  48.315  1.00  2.98
ATOM   3108  C   ARG A 188     67.181  43.707  52.804  1.00  2.98
ATOM   3109  O   ARG A 188     66.435  42.752  52.579  1.00  2.98
ATOM   3110  N   LYS A 189     67.365  44.213  54.037  1.00  2.83
ATOM   3111  H   LYS A 189     68.025  44.973  54.162  1.00  2.83
ATOM   3112  CA  LYS A 189     66.573  43.732  55.181  1.00  2.83
ATOM   3113  HA  LYS A 189     66.666  42.646  55.145  1.00  2.83
ATOM   3114  CB  LYS A 189     67.107  44.174  56.556  1.00  2.83
ATOM   3115  HB1 LYS A 189     66.621  45.100  56.868  1.00  2.83
ATOM   3116  HB2 LYS A 189     68.186  44.334  56.509  1.00  2.83
ATOM   3117  CG  LYS A 189     66.811  43.033  57.555  1.00  2.83
ATOM   3118  HG1 LYS A 189     67.423  42.169  57.289  1.00  2.83
ATOM   3119  HG2 LYS A 189     65.766  42.742  57.451  1.00  2.83
ATOM   3120  CD  LYS A 189     67.060  43.370  59.029  1.00  2.83
ATOM   3121  HD1 LYS A 189     66.474  44.253  59.270  1.00  2.83
ATOM   3122  HD2 LYS A 189     68.118  43.584  59.190  1.00  2.83
ATOM   3123  CE  LYS A 189     66.616  42.172  59.895  1.00  2.83
ATOM   3124  HE1 LYS A 189     67.300  41.338  59.710  1.00  2.83
```

```
ATOM   3125  HE2 LYS A 189      65.620  41.856  59.569  1.00  2.83
ATOM   3126  NZ  LYS A 189      66.567  42.478  61.349  1.00  2.83
ATOM   3127  HZ1 LYS A 189      66.343  41.650  61.886  1.00  2.83
ATOM   3128  HZ2 LYS A 189      67.445  42.845  61.689  1.00  2.83
ATOM   3129  HZ3 LYS A 189      65.840  43.167  61.536  1.00  2.83
ATOM   3130  C   LYS A 189      65.064  43.967  54.993  1.00  2.83
ATOM   3131  O   LYS A 189      64.290  43.023  55.140  1.00  2.83
ATOM   3132  N   LEU A 190      64.667  45.172  54.580  1.00  2.76
ATOM   3133  H   LEU A 190      65.377  45.891  54.496  1.00  2.76
ATOM   3134  CA  LEU A 190      63.287  45.548  54.300  1.00  2.76
ATOM   3135  HA  LEU A 190      62.722  45.369  55.214  1.00  2.76
ATOM   3136  CB  LEU A 190      63.190  47.049  53.974  1.00  2.76
ATOM   3137  HB1 LEU A 190      63.745  47.245  53.056  1.00  2.76
ATOM   3138  HB2 LEU A 190      63.666  47.607  54.778  1.00  2.76
ATOM   3139  CG  LEU A 190      61.747  47.574  53.803  1.00  2.76
ATOM   3140  HG  LEU A 190      61.276  47.076  52.956  1.00  2.76
ATOM   3141  CD1 LEU A 190      60.866  47.386  55.042  1.00  2.76
ATOM   3142 1HD1 LEU A 190      59.897  47.858  54.884  1.00  2.76
ATOM   3143 2HD1 LEU A 190      60.702  46.325  55.225  1.00  2.76
ATOM   3144 3HD1 LEU A 190      61.344  47.833  55.912  1.00  2.76
ATOM   3145  CD2 LEU A 190      61.806  49.073  53.507  1.00  2.76
ATOM   3146 1HD2 LEU A 190      60.798  49.447  53.340  1.00  2.76
ATOM   3147 2HD2 LEU A 190      62.258  49.610  54.342  1.00  2.76
ATOM   3148 3HD2 LEU A 190      62.392  49.246  52.605  1.00  2.76
ATOM   3149  C   LEU A 190      62.645  44.676  53.211  1.00  2.76
ATOM   3150  O   LEU A 190      61.529  44.198  53.407  1.00  2.76
ATOM   3151  N   ASN A 191      63.350  44.378  52.116  1.00  2.97
ATOM   3152  H   ASN A 191      64.221  44.867  51.944  1.00  2.97
ATOM   3153  CA  ASN A 191      62.862  43.402  51.143  1.00  2.97
ATOM   3154  HA  ASN A 191      61.840  43.691  50.889  1.00  2.97
ATOM   3155  CB  ASN A 191      63.686  43.412  49.832  1.00  2.97
ATOM   3156  HB1 ASN A 191      63.723  42.397  49.435  1.00  2.97
ATOM   3157  HB2 ASN A 191      64.706  43.742  50.024  1.00  2.97
ATOM   3158  CG  ASN A 191      63.067  44.282  48.737  1.00  2.97
ATOM   3159  OD1 ASN A 191      62.102  44.998  48.927  1.00  2.97
ATOM   3160  ND2 ASN A 191      63.575  44.258  47.524  1.00  2.97
ATOM   3161 1HD2 ASN A 191      64.351  43.665  47.295  1.00  2.97
ATOM   3162 2HD2 ASN A 191      63.175  44.913  46.875  1.00  2.97
ATOM   3163  C   ASN A 191      62.689  41.986  51.734  1.00  2.97
ATOM   3164  O   ASN A 191      61.629  41.369  51.603  1.00  2.97
ATOM   3165  N   ARG A 192      63.696  41.504  52.474  1.00  3.06
ATOM   3166  H   ARG A 192      64.515  42.098  52.573  1.00  3.06
ATOM   3167  CA  ARG A 192      63.665  40.248  53.255  1.00  3.06
ATOM   3168  HA  ARG A 192      63.277  39.448  52.624  1.00  3.06
ATOM   3169  CB  ARG A 192      65.109  39.909  53.702  1.00  3.06
ATOM   3170  HB1 ARG A 192      65.099  39.104  54.438  1.00  3.06
ATOM   3171  HB2 ARG A 192      65.528  40.783  54.200  1.00  3.06
ATOM   3172  CG  ARG A 192      66.055  39.493  52.562  1.00  3.06
ATOM   3173  HG1 ARG A 192      67.081  39.716  52.860  1.00  3.06
ATOM   3174  HG2 ARG A 192      65.836  40.084  51.673  1.00  3.06
ATOM   3175  CD  ARG A 192      65.951  37.998  52.210  1.00  3.06
ATOM   3176  HD1 ARG A 192      66.408  37.837  51.232  1.00  3.06
ATOM   3177  HD2 ARG A 192      64.895  37.731  52.119  1.00  3.06
```

```
ATOM   3178  NE  ARG A 192      66.602  37.113  53.210  1.00  3.06
ATOM   3179  HE  ARG A 192      65.988  36.666  53.872  1.00  3.06
ATOM   3180  CZ  ARG A 192      67.888  36.795  53.280  1.00  3.06
ATOM   3181  NH1 ARG A 192      68.313  35.876  54.102  1.00  3.06
ATOM   3182 1HH1 ARG A 192      67.655  35.365  54.666  1.00  3.06
ATOM   3183 2HH1 ARG A 192      69.287  35.633  54.138  1.00  3.06
ATOM   3184  NH2 ARG A 192      68.777  37.380  52.530  1.00  3.06
ATOM   3185 1HH2 ARG A 192      68.459  38.079  51.884  1.00  3.06
ATOM   3186 2HH2 ARG A 192      69.742  37.107  52.569  1.00  3.06
ATOM   3187  C   ARG A 192     62.737  40.292  54.485  1.00  3.06
ATOM   3188  O   ARG A 192     62.827  39.413  55.340  1.00  3.06
ATOM   3189  N   ILE A 193     61.861  41.292  54.588  1.00  2.91
ATOM   3190  H   ILE A 193     61.927  42.024  53.898  1.00  2.91
ATOM   3191  CA  ILE A 193     60.866  41.450  55.656  1.00  2.91
ATOM   3192  HA  ILE A 193     60.860  40.565  56.293  1.00  2.91
ATOM   3193  CB  ILE A 193     61.241  42.679  56.514  1.00  2.91
ATOM   3194  HB  ILE A 193     61.675  43.433  55.860  1.00  2.91
ATOM   3195  CG2 ILE A 193     60.044  43.354  57.216  1.00  2.91
ATOM   3196 1HG2 ILE A 193      59.500  42.644  57.834  1.00  2.91
ATOM   3197 2HG2 ILE A 193      60.376  44.182  57.836  1.00  2.91
ATOM   3198 3HG2 ILE A 193      59.367  43.775  56.476  1.00  2.91
ATOM   3199  CG1 ILE A 193     62.325  42.228  57.508  1.00  2.91
ATOM   3200 1HG1 ILE A 193      63.083  41.640  56.992  1.00  2.91
ATOM   3201 2HG1 ILE A 193      61.877  41.584  58.262  1.00  2.91
ATOM   3202  CD1 ILE A 193     63.040  43.398  58.174  1.00  2.91
ATOM   3203  HD1 ILE A 193     63.806  43.003  58.838  1.00  2.91
ATOM   3204  HD2 ILE A 193     63.506  44.035  57.425  1.00  2.91
ATOM   3205  HD3 ILE A 193     62.336  43.995  58.747  1.00  2.91
ATOM   3206  C   ILE A 193     59.459  41.568  55.083  1.00  2.91
ATOM   3207  O   ILE A 193     58.536  40.941  55.596  1.00  2.91
ATOM   3208  N   LEU A 194     59.301  42.306  53.984  1.00  3.02
ATOM   3209  H   LEU A 194     60.083  42.853  53.643  1.00  3.02
ATOM   3210  CA  LEU A 194     58.023  42.429  53.295  1.00  3.02
ATOM   3211  HA  LEU A 194     57.230  42.328  54.037  1.00  3.02
ATOM   3212  CB  LEU A 194     57.893  43.838  52.704  1.00  3.02
ATOM   3213  HB1 LEU A 194      56.931  43.905  52.195  1.00  3.02
ATOM   3214  HB2 LEU A 194      58.684  43.989  51.967  1.00  3.02
ATOM   3215  CG  LEU A 194     57.976  44.954  53.771  1.00  3.02
ATOM   3216  HG  LEU A 194     58.980  44.978  54.189  1.00  3.02
ATOM   3217  CD1 LEU A 194     57.718  46.305  53.106  1.00  3.02
ATOM   3218 1HD1 LEU A 194      58.470  46.470  52.332  1.00  3.02
ATOM   3219 2HD1 LEU A 194      56.732  46.300  52.655  1.00  3.02
ATOM   3220 3HD1 LEU A 194      57.797  47.103  53.844  1.00  3.02
ATOM   3221  CD2 LEU A 194     56.985  44.803  54.933  1.00  3.02
ATOM   3222 1HD2 LEU A 194      56.981  45.709  55.539  1.00  3.02
ATOM   3223 2HD2 LEU A 194      55.985  44.615  54.554  1.00  3.02
ATOM   3224 3HD2 LEU A 194      57.278  43.968  55.568  1.00  3.02
ATOM   3225  C   LEU A 194     57.731  41.297  52.310  1.00  3.02
ATOM   3226  O   LEU A 194     56.569  41.096  51.983  1.00  3.02
ATOM   3227  N   GLN A 195     58.732  40.517  51.889  1.00  3.45
ATOM   3228  H   GLN A 195     59.687  40.803  52.071  1.00  3.45
ATOM   3229  CA  GLN A 195     58.505  39.280  51.133  1.00  3.45
ATOM   3230  HA  GLN A 195     57.703  39.476  50.439  1.00  3.45
```

```
ATOM   3231  CB  GLN A 195      59.766  39.004  50.290  1.00  3.45
ATOM   3232  HB1 GLN A 195      60.650  38.981  50.925  1.00  3.45
ATOM   3233  HB2 GLN A 195      59.905  39.850  49.616  1.00  3.45
ATOM   3234  CG  GLN A 195      59.699  37.725  49.436  1.00  3.45
ATOM   3235  HG1 GLN A 195      60.323  37.862  48.554  1.00  3.45
ATOM   3236  HG2 GLN A 195      58.675  37.566  49.095  1.00  3.45
ATOM   3237  CD  GLN A 195      60.204  36.476  50.156  1.00  3.45
ATOM   3238  OE1 GLN A 195      61.034  36.521  51.050  1.00  3.45
ATOM   3239  NE2 GLN A 195      59.742  35.302  49.788  1.00  3.45
ATOM   3240 1HE2 GLN A 195      59.006  35.223  49.093  1.00  3.45
ATOM   3241 2HE2 GLN A 195      60.112  34.501  50.267  1.00  3.45
ATOM   3242  C   GLN A 195     58.032  38.049  51.957  1.00  3.45
ATOM   3243  O   GLN A 195     57.034  37.417  51.588  1.00  3.45
ATOM   3244  N   PRO A 196     58.695  37.663  53.068  1.00  4.43
ATOM   3245  CD  PRO A 196      59.873  38.284  53.662  1.00  4.43
ATOM   3246  HD1 PRO A 196      59.682  39.313  53.943  1.00  4.43
ATOM   3247  HD2 PRO A 196      60.713  38.234  52.971  1.00  4.43
ATOM   3248  CG  PRO A 196      60.203  37.461  54.904  1.00  4.43
ATOM   3249  HG1 PRO A 196      59.596  37.795  55.747  1.00  4.43
ATOM   3250  HG2 PRO A 196      61.262  37.485  55.150  1.00  4.43
ATOM   3251  CB  PRO A 196      59.778  36.069  54.467  1.00  4.43
ATOM   3252  HB1 PRO A 196      59.648  35.397  55.315  1.00  4.43
ATOM   3253  HB2 PRO A 196      60.529  35.660  53.788  1.00  4.43
ATOM   3254  CA  PRO A 196      58.488  36.349  53.684  1.00  4.43
ATOM   3255  HA  PRO A 196      58.409  35.605  52.892  1.00  4.43
ATOM   3256  C   PRO A 196     57.219  36.197  54.543  1.00  4.43
ATOM   3257  O   PRO A 196     57.197  35.422  55.500  1.00  4.43
ATOM   3258  N   CYS A 197     56.161  36.948  54.251  1.00  6.71
ATOM   3259  H   CYS A 197     56.247  37.537  53.431  1.00  6.71
ATOM   3260  CA  CYS A 197      54.805  36.497  54.565  1.00  6.71
ATOM   3261  HA  CYS A 197      54.785  35.900  55.477  1.00  6.71
ATOM   3262  CB  CYS A 197      53.933  37.760  54.760  1.00  6.71
ATOM   3263  HB1 CYS A 197      53.599  38.160  53.800  1.00  6.71
ATOM   3264  HB2 CYS A 197      54.519  38.531  55.259  1.00  6.71
ATOM   3265  SG  CYS A 197      52.485  37.405  55.791  1.00  6.71
ATOM   3266  HG  CYS A 197      51.926  36.483  54.969  1.00  6.71
ATOM   3267  C   CYS A 197     54.371  35.593  53.421  1.00  6.71
ATOM   3268  O   CYS A 197     54.410  34.368  53.509  1.00  6.71
ATOM   3269  N   GLU A 198     54.182  36.219  52.269  1.00  7.80
ATOM   3270  H   GLU A 198     54.363  37.209  52.219  1.00  7.80
ATOM   3271  CA  GLU A 198      53.713  35.599  51.047  1.00  7.80
ATOM   3272  HA  GLU A 198      54.092  34.578  51.000  1.00  7.80
ATOM   3273  CB  GLU A 198      52.172  35.543  50.960  1.00  7.80
ATOM   3274  HB1 GLU A 198      51.909  35.553  49.901  1.00  7.80
ATOM   3275  HB2 GLU A 198      51.732  36.429  51.409  1.00  7.80
ATOM   3276  CG  GLU A 198      51.533  34.270  51.560  1.00  7.80
ATOM   3277  HG1 GLU A 198      52.146  33.407  51.286  1.00  7.80
ATOM   3278  HG2 GLU A 198      50.564  34.136  51.073  1.00  7.80
ATOM   3279  CD  GLU A 198      51.290  34.279  53.085  1.00  7.80
ATOM   3280  OE1 GLU A 198      50.927  33.204  53.620  1.00  7.80
ATOM   3281  OE2 GLU A 198      51.422  35.364  53.699  1.00  7.80
ATOM   3282  C   GLU A 198     54.348  36.372  49.916  1.00  7.80
ATOM   3283  O   GLU A 198     54.667  37.535  50.099  1.00  7.80
```

```
ATOM   3284  N   THR A 199      54.576  35.764  48.754  1.00  8.80
ATOM   3285  H   THR A 199      54.275  34.808  48.618  1.00  8.80
ATOM   3286  CA  THR A 199      55.328  36.453  47.687  1.00  8.80
ATOM   3287  HA  THR A 199      55.849  37.297  48.112  1.00  8.80
ATOM   3288  CB  THR A 199      56.442  35.537  47.150  1.00  8.80
ATOM   3289  HB  THR A 199      55.997  34.656  46.687  1.00  8.80
ATOM   3290  CG2 THR A 199      57.418  36.189  46.170  1.00  8.80
ATOM   3291 1HG2 THR A 199      57.791  37.127  46.582  1.00  8.80
ATOM   3292 2HG2 THR A 199      58.254  35.515  45.983  1.00  8.80
ATOM   3293 3HG2 THR A 199      56.922  36.382  45.219  1.00  8.80
ATOM   3294  OG1 THR A 199      57.245  35.133  48.239  1.00  8.80
ATOM   3295  HG1 THR A 199      56.647  35.007  48.984  1.00  8.80
ATOM   3296  C   THR A 199      54.420  36.965  46.571  1.00  8.80
ATOM   3297  O   THR A 199      54.850  37.768  45.752  1.00  8.80
ATOM   3298  N   GLU A 200      53.145  36.585  46.592  1.00 10.18
ATOM   3299  H   GLU A 200      52.858  35.878  47.247  1.00 10.18
ATOM   3300  CA  GLU A 200      52.088  37.242  45.820  1.00 10.18
ATOM   3301  HA  GLU A 200      52.529  37.917  45.089  1.00 10.18
ATOM   3302  CB  GLU A 200      51.286  36.207  45.018  1.00 10.18
ATOM   3303  HB1 GLU A 200      50.363  36.662  44.658  1.00 10.18
ATOM   3304  HB2 GLU A 200      51.032  35.367  45.666  1.00 10.18
ATOM   3305  CG  GLU A 200      52.097  35.726  43.803  1.00 10.18
ATOM   3306  HG1 GLU A 200      53.094  35.415  44.128  1.00 10.18
ATOM   3307  HG2 GLU A 200      52.223  36.557  43.104  1.00 10.18
ATOM   3308  CD  GLU A 200      51.413  34.543  43.110  1.00 10.18
ATOM   3309  OE1 GLU A 200      50.818  34.752  42.031  1.00 10.18
ATOM   3310  OE2 GLU A 200      51.501  33.442  43.699  1.00 10.18
ATOM   3311  C   GLU A 200      51.201  38.130  46.686  1.00 10.18
ATOM   3312  O   GLU A 200      50.691  39.106  46.143  1.00 10.18
ATOM   3313  N   ASP A 201      51.183  37.956  48.023  1.00 10.69
ATOM   3314  H   ASP A 201      51.571  37.129  48.438  1.00 10.69
ATOM   3315  CA  ASP A 201      51.077  39.156  48.861  1.00 10.69
ATOM   3316  HA  ASP A 201      50.156  39.680  48.593  1.00 10.69
ATOM   3317  CB  ASP A 201      51.023  38.919  50.380  1.00 10.69
ATOM   3318  HB1 ASP A 201      51.976  38.503  50.705  1.00 10.69
ATOM   3319  HB2 ASP A 201      50.231  38.202  50.602  1.00 10.69
ATOM   3320  CG  ASP A 201      50.748  40.230  51.160  1.00 10.69
ATOM   3321  OD1 ASP A 201      49.637  40.786  51.022  1.00 10.69
ATOM   3322  OD2 ASP A 201      51.656  40.687  51.893  1.00 10.69
ATOM   3323  C   ASP A 201      52.247  40.065  48.493  1.00 10.69
ATOM   3324  O   ASP A 201      52.004  41.056  47.863  1.00 10.69
ATOM   3325  N   LEU A 202      53.523  39.713  48.648  1.00 10.16
ATOM   3326  H   LEU A 202      53.721  38.890  49.206  1.00 10.16
ATOM   3327  CA  LEU A 202      54.658  40.581  48.296  1.00 10.16
ATOM   3328  HA  LEU A 202      54.750  41.338  49.073  1.00 10.16
ATOM   3329  CB  LEU A 202      55.985  39.815  48.283  1.00 10.16
ATOM   3330  HB1 LEU A 202      55.882  39.020  47.568  1.00 10.16
ATOM   3331  HB2 LEU A 202      56.090  39.372  49.266  1.00 10.16
ATOM   3332  CG  LEU A 202      57.317  40.512  47.914  1.00 10.16
ATOM   3333  HG  LEU A 202      58.094  39.837  48.257  1.00 10.16
ATOM   3334  CD1 LEU A 202      57.588  40.662  46.413  1.00 10.16
ATOM   3335 1HD1 LEU A 202      58.644  40.880  46.259  1.00 10.16
ATOM   3336 2HD1 LEU A 202      57.341  39.734  45.899  1.00 10.16
```

```
ATOM   3337 3HD1 LEU A 202      57.009  41.478  45.990  1.00 10.16
ATOM   3338  CD2 LEU A 202      57.558  41.857  48.584  1.00 10.16
ATOM   3339 1HD2 LEU A 202      58.599  42.153  48.461  1.00 10.16
ATOM   3340 2HD2 LEU A 202      56.922  42.617  48.149  1.00 10.16
ATOM   3341 3HD2 LEU A 202      57.331  41.783  49.643  1.00 10.16
ATOM   3342  C   LEU A 202     54.527  41.289  46.968  1.00 10.16
ATOM   3343  O   LEU A 202     54.886  42.443  46.894  1.00 10.16
ATOM   3344  N   ARG A 203     54.040  40.652  45.912  1.00 11.14
ATOM   3345  H   ARG A 203     53.794  39.677  46.015  1.00 11.14
ATOM   3346  CA  ARG A 203      53.918  41.299  44.612  1.00 11.14
ATOM   3347  HA  ARG A 203      54.728  42.006  44.494  1.00 11.14
ATOM   3348  CB  ARG A 203      54.054  40.203  43.521  1.00 11.14
ATOM   3349  HB1 ARG A 203      53.189  40.210  42.853  1.00 11.14
ATOM   3350  HB2 ARG A 203      54.057  39.216  43.973  1.00 11.14
ATOM   3351  CG  ARG A 203      55.329  40.338  42.681  1.00 11.14
ATOM   3352  HG1 ARG A 203      55.552  39.375  42.219  1.00 11.14
ATOM   3353  HG2 ARG A 203      56.165  40.613  43.326  1.00 11.14
ATOM   3354  CD  ARG A 203      55.142  41.381  41.571  1.00 11.14
ATOM   3355  HD1 ARG A 203      54.706  42.283  42.000  1.00 11.14
ATOM   3356  HD2 ARG A 203      54.427  40.989  40.845  1.00 11.14
ATOM   3357  NE  ARG A 203     56.420  41.684  40.890  1.00 11.14
ATOM   3358  HE  ARG A 203     56.797  40.959  40.299  1.00 11.14
ATOM   3359  CZ  ARG A 203     57.126  42.793  41.021  1.00 11.14
ATOM   3360  NH1 ARG A 203      58.272  42.895  40.407  1.00 11.14
ATOM   3361 1HH1 ARG A 203      58.829  43.726  40.493  1.00 11.14
ATOM   3362 2HH1 ARG A 203      58.605  42.145  39.827  1.00 11.14
ATOM   3363  NH2 ARG A 203      56.714  43.795  41.753  1.00 11.14
ATOM   3364 1HH2 ARG A 203      57.290  44.608  41.902  1.00 11.14
ATOM   3365 2HH2 ARG A 203      55.817  43.733  42.205  1.00 11.14
ATOM   3366  C   ARG A 203     52.625  42.029  44.371  1.00 11.14
ATOM   3367  O   ARG A 203     52.642  42.829  43.437  1.00 11.14
ATOM   3368  N   ASP A 204     51.576  41.815  45.164  1.00 11.83
ATOM   3369  H   ASP A 204     51.599  41.068  45.854  1.00 11.83
ATOM   3370  CA  ASP A 204      50.354  42.609  45.065  1.00 11.83
ATOM   3371  HA  ASP A 204      50.560  43.304  44.260  1.00 11.83
ATOM   3372  CB  ASP A 204      49.126  41.807  44.606  1.00 11.83
ATOM   3373  HB1 ASP A 204      48.606  41.397  45.473  1.00 11.83
ATOM   3374  HB2 ASP A 204      49.451  40.980  43.972  1.00 11.83
ATOM   3375  CG  ASP A 204      48.177  42.693  43.780  1.00 11.83
ATOM   3376  OD1 ASP A 204      46.988  42.365  43.592  1.00 11.83
ATOM   3377  OD2 ASP A 204      48.643  43.687  43.173  1.00 11.83
ATOM   3378  C   ASP A 204     50.053  43.577  46.197  1.00 11.83
ATOM   3379  O   ASP A 204     49.316  44.525  46.026  1.00 11.83
ATOM   3380  N   VAL A 205     50.760  43.408  47.288  1.00 11.33
ATOM   3381  H   VAL A 205     51.174  42.486  47.370  1.00 11.33
ATOM   3382  CA  VAL A 205      51.211  44.351  48.293  1.00 11.33
ATOM   3383  HA  VAL A 205      50.367  45.014  48.483  1.00 11.33
ATOM   3384  CB  VAL A 205      51.505  43.569  49.597  1.00 11.33
ATOM   3385  HB  VAL A 205      51.021  42.600  49.521  1.00 11.33
ATOM   3386  CG1 VAL A 205      52.987  43.300  49.843  1.00 11.33
ATOM   3387 1HG1 VAL A 205      53.101  42.547  50.625  1.00 11.33
ATOM   3388 2HG1 VAL A 205      53.423  42.933  48.925  1.00 11.33
ATOM   3389 3HG1 VAL A 205      53.505  44.199  50.122  1.00 11.33
```

```
ATOM   3390  CG2 VAL A 205      50.862  44.202  50.821  1.00 11.33
ATOM   3391 1HG2 VAL A 205      51.117  43.618  51.705  1.00 11.33
ATOM   3392 2HG2 VAL A 205      51.169  45.238  50.919  1.00 11.33
ATOM   3393 3HG2 VAL A 205      49.782  44.127  50.697  1.00 11.33
ATOM   3394  C   VAL A 205      52.345  45.241  47.775  1.00 11.33
ATOM   3395  O   VAL A 205      52.417  46.397  48.212  1.00 11.33
ATOM   3396  N   PHE A 206      53.123  44.748  46.778  1.00 11.76
ATOM   3397  H   PHE A 206      53.056  43.755  46.617  1.00 11.76
ATOM   3398  CA  PHE A 206      54.023  45.522  45.904  1.00 11.76
ATOM   3399  HA  PHE A 206      54.394  46.308  46.502  1.00 11.76
ATOM   3400  CB  PHE A 206      55.246  44.800  45.312  1.00 11.76
ATOM   3401  HB1 PHE A 206      54.897  44.016  44.653  1.00 11.76
ATOM   3402  HB2 PHE A 206      55.779  44.380  46.156  1.00 11.76
ATOM   3403  CG  PHE A 206      56.310  45.602  44.561  1.00 11.76
ATOM   3404  CD1 PHE A 206      57.652  45.492  44.980  1.00 11.76
ATOM   3405  HD1 PHE A 206      57.902  44.891  45.844  1.00 11.76
ATOM   3406  CE1 PHE A 206      58.682  46.139  44.276  1.00 11.76
ATOM   3407  HE1 PHE A 206      59.704  46.037  44.611  1.00 11.76
ATOM   3408  CZ  PHE A 206      58.380  46.933  43.158  1.00 11.76
ATOM   3409  HZ  PHE A 206      59.164  47.467  42.637  1.00 11.76
ATOM   3410  CE2 PHE A 206      57.049  47.054  42.731  1.00 11.76
ATOM   3411  HE2 PHE A 206      56.805  47.690  41.889  1.00 11.76
ATOM   3412  CD2 PHE A 206      56.026  46.369  43.408  1.00 11.76
ATOM   3413  HD2 PHE A 206      55.024  46.455  43.022  1.00 11.76
ATOM   3414  C   PHE A 206      53.301  46.324  44.841  1.00 11.76
ATOM   3415  O   PHE A 206      53.499  47.529  44.743  1.00 11.76
ATOM   3416  N   ARG A 207      52.516  45.710  43.970  1.00 13.42
ATOM   3417  H   ARG A 207      52.346  44.713  44.052  1.00 13.42
ATOM   3418  CA  ARG A 207      51.888  46.469  42.898  1.00 13.42
ATOM   3419  HA  ARG A 207      52.601  47.170  42.463  1.00 13.42
ATOM   3420  CB  ARG A 207      51.433  45.466  41.816  1.00 13.42
ATOM   3421  HB1 ARG A 207      50.808  44.713  42.293  1.00 13.42
ATOM   3422  HB2 ARG A 207      52.309  44.949  41.421  1.00 13.42
ATOM   3423  CG  ARG A 207      50.664  46.094  40.635  1.00 13.42
ATOM   3424  HG1 ARG A 207      51.374  46.613  39.991  1.00 13.42
ATOM   3425  HG2 ARG A 207      49.934  46.821  40.991  1.00 13.42
ATOM   3426  CD  ARG A 207      49.910  45.039  39.810  1.00 13.42
ATOM   3427  HD1 ARG A 207      50.603  44.247  39.520  1.00 13.42
ATOM   3428  HD2 ARG A 207      49.537  45.517  38.904  1.00 13.42
ATOM   3429  NE  ARG A 207      48.796  44.453  40.582  1.00 13.42
ATOM   3430  HE  ARG A 207      49.008  44.121  41.524  1.00 13.42
ATOM   3431  CZ  ARG A 207      47.518  44.339  40.281  1.00 13.42
ATOM   3432  NH1 ARG A 207      46.697  43.820  41.142  1.00 13.42
ATOM   3433 1HH1 ARG A 207      47.074  43.496  42.047  1.00 13.42
ATOM   3434 2HH1 ARG A 207      45.723  43.694  40.978  1.00 13.42
ATOM   3435  NH2 ARG A 207      47.045  44.736  39.133  1.00 13.42
ATOM   3436 1HH2 ARG A 207      47.687  45.136  38.479  1.00 13.42
ATOM   3437 2HH2 ARG A 207      46.068  44.642  38.942  1.00 13.42
ATOM   3438  C   ARG A 207      50.725  47.306  43.413  1.00 13.42
ATOM   3439  O   ARG A 207      50.845  48.534  43.468  1.00 13.42
ATOM   3440  N   LEU A 208      49.649  46.652  43.873  1.00 14.00
ATOM   3441  H   LEU A 208      49.639  45.636  43.865  1.00 14.00
ATOM   3442  CA  LEU A 208      48.900  47.283  44.945  1.00 14.00
```

```
ATOM   3443  HA  LEU A 208      48.771  48.332  44.680  1.00 14.00
ATOM   3444  CB  LEU A 208      47.455  46.734  45.095  1.00 14.00
ATOM   3445  HB1 LEU A 208      46.911  47.387  45.778  1.00 14.00
ATOM   3446  HB2 LEU A 208      47.446  45.751  45.549  1.00 14.00
ATOM   3447  CG  LEU A 208      46.653  46.636  43.783  1.00 14.00
ATOM   3448  HG  LEU A 208      47.184  46.018  43.062  1.00 14.00
ATOM   3449  CD1 LEU A 208      45.292  45.993  44.053  1.00 14.00
ATOM   3450 1HD1 LEU A 208      44.712  45.939  43.134  1.00 14.00
ATOM   3451 2HD1 LEU A 208      45.451  44.977  44.421  1.00 14.00
ATOM   3452 3HD1 LEU A 208      44.745  46.560  44.805  1.00 14.00
ATOM   3453  CD2 LEU A 208      46.399  48.011  43.154  1.00 14.00
ATOM   3454 1HD2 LEU A 208      45.795  47.895  42.255  1.00 14.00
ATOM   3455 2HD2 LEU A 208      45.873  48.654  43.858  1.00 14.00
ATOM   3456 3HD2 LEU A 208      47.346  48.471  42.874  1.00 14.00
ATOM   3457  C   LEU A 208     49.764  47.299  46.183  1.00 14.00
ATOM   3458  O   LEU A 208     50.911  46.912  46.147  1.00 14.00
ATOM   3459  N   PHE A 209     49.298  48.033  47.158  1.00 14.08
ATOM   3460  H   PHE A 209     48.302  48.042  47.228  1.00 14.08
ATOM   3461  CA  PHE A 209      50.013  48.928  48.047  1.00 14.08
ATOM   3462  HA  PHE A 209      49.390  49.817  48.097  1.00 14.08
ATOM   3463  CB  PHE A 209      49.912  48.299  49.442  1.00 14.08
ATOM   3464  HB1 PHE A 209      50.278  49.014  50.175  1.00 14.08
ATOM   3465  HB2 PHE A 209      50.561  47.428  49.483  1.00 14.08
ATOM   3466  CG  PHE A 209      48.476  47.861  49.796  1.00 14.08
ATOM   3467  CD1 PHE A 209      47.343  48.619  49.409  1.00 14.08
ATOM   3468  HD1 PHE A 209      47.454  49.536  48.852  1.00 14.08
ATOM   3469  CE1 PHE A 209      46.043  48.201  49.748  1.00 14.08
ATOM   3470  HE1 PHE A 209      45.189  48.787  49.438  1.00 14.08
ATOM   3471  CZ  PHE A 209     45.855  47.025  50.490  1.00 14.08
ATOM   3472  HZ  PHE A 209     44.857  46.698  50.744  1.00 14.08
ATOM   3473  CE2 PHE A 209      46.969  46.282  50.905  1.00 14.08
ATOM   3474  HE2 PHE A 209      46.824  45.380  51.481  1.00 14.08
ATOM   3475  CD2 PHE A 209      48.265  46.707  50.568  1.00 14.08
ATOM   3476  HD2 PHE A 209      49.108  46.145  50.916  1.00 14.08
ATOM   3477  C   PHE A 209     51.328  49.545  47.573  1.00 14.08
ATOM   3478  O   PHE A 209     52.000  50.155  48.413  1.00 14.08
ATOM   3479  N   GLY A 210     51.612  49.518  46.248  1.00 14.71
ATOM   3480  H   GLY A 210     51.104  48.871  45.666  1.00 14.71
ATOM   3481  CA  GLY A 210      52.696  50.212  45.585  1.00 14.71
ATOM   3482  HA1 GLY A 210      52.454  51.271  45.494  1.00 14.71
ATOM   3483  HA2 GLY A 210      52.886  49.791  44.598  1.00 14.71
ATOM   3484  C   GLY A 210     53.938  50.070  46.475  1.00 14.71
ATOM   3485  O   GLY A 210     54.472  51.093  46.883  1.00 14.71
ATOM   3486  N   LEU A 211     54.278  48.837  46.903  1.00 14.96
ATOM   3487  H   LEU A 211     53.703  48.081  46.547  1.00 14.96
ATOM   3488  CA  LEU A 211      55.129  48.576  48.055  1.00 14.96
ATOM   3489  HA  LEU A 211      54.549  48.856  48.903  1.00 14.96
ATOM   3490  CB  LEU A 211      55.640  47.116  48.251  1.00 14.96
ATOM   3491  HB1 LEU A 211      56.077  46.795  47.309  1.00 14.96
ATOM   3492  HB2 LEU A 211      54.764  46.521  48.465  1.00 14.96
ATOM   3493  CG  LEU A 211      56.665  46.671  49.328  1.00 14.96
ATOM   3494  HG  LEU A 211      56.324  46.953  50.314  1.00 14.96
ATOM   3495  CD1 LEU A 211      56.729  45.147  49.263  1.00 14.96
```

```
ATOM   3496 1HD1 LEU A 211      57.427  44.755  49.998  1.00 14.96
ATOM   3497 2HD1 LEU A 211      55.749  44.709  49.453  1.00 14.96
ATOM   3498 3HD1 LEU A 211      57.062  44.839  48.274  1.00 14.96
ATOM   3499  CD2 LEU A 211     58.114  47.143  49.162  1.00 14.96
ATOM   3500 1HD2 LEU A 211      58.199  48.206  49.375  1.00 14.96
ATOM   3501 2HD2 LEU A 211      58.764  46.624  49.869  1.00 14.96
ATOM   3502 3HD2 LEU A 211      58.462  46.933  48.152  1.00 14.96
ATOM   3503  C   LEU A 211     56.357  49.457  48.018  1.00 14.96
ATOM   3504  O   LEU A 211     56.577  50.294  48.896  1.00 14.96
ATOM   3505  N   LEU A 212     57.126  49.273  46.951  1.00 15.82
ATOM   3506  H   LEU A 212     56.860  48.573  46.277  1.00 15.82
ATOM   3507  CA  LEU A 212     58.067  50.288  46.587  1.00 15.82
ATOM   3508  HA  LEU A 212     58.569  50.661  47.480  1.00 15.82
ATOM   3509  CB  LEU A 212     59.147  49.760  45.609  1.00 15.82
ATOM   3510  HB1 LEU A 212     59.707  50.628  45.255  1.00 15.82
ATOM   3511  HB2 LEU A 212     58.657  49.317  44.741  1.00 15.82
ATOM   3512  CG  LEU A 212     60.172  48.747  46.153  1.00 15.82
ATOM   3513  HG  LEU A 212     59.671  47.799  46.334  1.00 15.82
ATOM   3514  CD1 LEU A 212     61.286  48.530  45.120  1.00 15.82
ATOM   3515 1HD1 LEU A 212      61.956  47.741  45.459  1.00 15.82
ATOM   3516 2HD1 LEU A 212      60.862  48.242  44.160  1.00 15.82
ATOM   3517 3HD1 LEU A 212      61.862  49.447  44.987  1.00 15.82
ATOM   3518  CD2 LEU A 212     60.847  49.196  47.450  1.00 15.82
ATOM   3519 1HD2 LEU A 212      61.603  48.466  47.743  1.00 15.82
ATOM   3520 2HD2 LEU A 212      61.320  50.168  47.317  1.00 15.82
ATOM   3521 3HD2 LEU A 212      60.114  49.252  48.252  1.00 15.82
ATOM   3522  C   LEU A 212     57.354  51.488  45.959  1.00 15.82
ATOM   3523  O   LEU A 212     57.422  52.605  46.467  1.00 15.82
ATOM   3524  N   THR A 213     56.657  51.205  44.862  1.00 17.40
ATOM   3525  H   THR A 213     56.731  50.255  44.535  1.00 17.40
ATOM   3526  CA  THR A 213     55.602  51.982  44.206  1.00 17.40
ATOM   3527  HA  THR A 213     54.812  52.265  44.904  1.00 17.40
ATOM   3528  CB  THR A 213     56.139  53.281  43.569  1.00 17.40
ATOM   3529  HB  THR A 213     57.114  53.095  43.118  1.00 17.40
ATOM   3530  CG2 THR A 213     55.227  53.917  42.516  1.00 17.40
ATOM   3531 1HG2 THR A 213      55.631  54.888  42.228  1.00 17.40
ATOM   3532 2HG2 THR A 213      55.189  53.292  41.624  1.00 17.40
ATOM   3533 3HG2 THR A 213      54.225  54.050  42.922  1.00 17.40
ATOM   3534  OG1 THR A 213     56.267  54.260  44.578  1.00 17.40
ATOM   3535  HG1 THR A 213     56.757  53.821  45.288  1.00 17.40
ATOM   3536  C   THR A 213     54.976  51.094  43.124  1.00 17.40
ATOM   3537  O   THR A 213     55.301  49.938  42.855  1.00 17.40
TER
```

Appendix 9: RaxML scripts

Model test:

GTRGAMMA: (100 runs)

raxmlHPC-PTHREADS-SSE3 -s 60_teleosts_nt_aligned.fasta -p 76565454343434 -m GTRGAMMA -N 20 -T 30 -n ModelTest

GTRCAT: (100 runs)

raxmlHPC-PTHREADS-SSE3 -s 60_teleosts_nt_aligned.fasta -p 1767656545454552 -m GTRCAT -N 20 -T 30 -n ModelTest_4

Initial rearrangement setting optimization: (20 runs each)

raxmlHPC-PTHREADS-SSE3 -f d -i 10 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -t RAxML_parsTree -N 10 -T 60 -n FI_10

raxmlHPC-PTHREADS-SSE3 -f d -i 20 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -t RAxML_parsTree -N 10 -T 60 -n FI_20

raxmlHPC-PTHREADS-SSE3 -f d -m GTRCAT -s 60_teleosts_nt_aligned.fasta -t RAxML_parsTree -N 10 -T 60 -n AI6

Number of categories optimization: (20 runs each)

raxmlHPC-PTHREADS-SSE3 -f d -i 10 -c 10 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -t RAxML_parsTree -N 10 -T 60 -n C10

raxmlHPC-PTHREADS-SSE3 -f d -i 10 -c 40 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -t RAxML_parsTree -N 10 -T 60 -n C40

raxmlHPC-PTHREADS-SSE3 -f d -i 10 -c 45 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -t RAxML_parsTree -N 10 -T 60 -n C45

raxmlHPC-PTHREADS-SSE3 -f d -i 10 -c 50 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -t RAxML_parsTree -N 10 -T 60 -n C50

raxmlHPC-PTHREADS-SSE3 -f d -i 10 -c 55 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -t RAxML_parsTree -N 10 -T 60 -n C55

raxmlHPC-PTHREADS-SSE3 -f d -i 10 -c 60 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -t RAxML_parsTree -N 10 -T 60 -n C60

raxmlHPC-PTHREADS-SSE3 -f d -i 10 -c 75 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -t RAxML_parsTree -T 40 -n C75

Finding the best-known likelihood tree (BKT):

raxmlHPC-PTHREADS-SSE3 -f d -i 10 -c 55 -p 767655454323 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -N 10 -T 40 -n BT0

raxmlHPC-PTHREADS-SSE3 -f d -d -i 10 -c 55 -p 987700011127 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -N 10 -T 40 -n BT10

raxmlHPC-PTHREADS-SSE3 -f o -i 10 -c 55 -p 443326776565000 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -N 10 -T 40 -n BT20

raxmlHPC-PTHREADS-SSE3 -f o -d -i 10 -c 55 -p 44335000 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -N 10 -T 40 -n BT30

Bootstrapping:

raxmlHPC-PTHREADS-SSE3 -f d -i 10 -c 55 -p 8121123 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -N 100 -b 76543434 -T 40 -n BS0

raxmlHPC-PTHREADS-SSE3 -f o -i 10 -c 55 -p 8776429 -m GTRCAT -s 60_teleosts_nt_aligned.fasta -N 100 -b 81010101 -T 40 -n BS20

Ancestral sequence prediction:

Based on the calculated BKT:

raxmlHPC-PTHREADS-SSE3 -f A -t 60_teleost_BT_rooted_nt_newick.txt -s 60_teleosts_nt_aligned.fasta -m GTRCAT -i 10 -c 55 -n ASR_nt

Based on the species tree published previously:

raxmlHPC-PTHREADS-SSE3 -f A -t 73g_nucl_conc_fossils.combined_latinnames.nex -s 60_teleosts_nt_aligned.fas -m GTRCAT -i 10 -c 55 -n ASR_nt_species

Appendix 10: MrBayes input files

```
#NEXUS
begin taxa;
        dimensions ntax=75;
        taxlabels
        Acanthochaenus_luetkenii
        Anabas_testudineus
        Antennarius_striatus
        Arctogadus_glacialis
        Astyanax_mexicanus
        Bathygadus_melanobranchus
        Benthosema_glaciale
        Beryx_splendens
        Boreogadus_saida
        Borostomias_antarcticus
        Brosme_brosme
        Brotula_barbata
        Carapus_acus
        Chaenocephalus_aceratus
        Chatrabus_melanurus
        Chromis_chromis
        Cyttopsis_roseus
        Danio_rerio
        Gadiculus_argenteus
        Gadus_morhua
        Lampetra_tridentata
        Gasterosteus_aculeatus
        Guentherus_altivela
        Helostoma_temminckii
        Holocentrus_rufus
        Laemonema_laureysi
        Lampris_guttauts
        Lamprogrammus_exutus
        Lesueurigobius_cf_sanzoi
        Lota_lota
        Macrourus_berglax
        Malacocephalus_occidentalis
        Melanogrammus_aeglefinus
        Melanonus_zugmayeri
        Merlangius_merlangus
        Merluccius_merluccius
        Merluccius_polli
        Molva_molva
        Monocentris_japonica
        Mora_moro
        Muraenolepis_marmoratus
        Myoxocephalus_scorpius
        Myripristis_jacobus
        Neoniphon_sammara
        Oreochromis_niloticus
        Oryzias_latipes
        Osmerus_eperlanus
```

```
        Parablennius_parvicornis
        Parasudis_fraserbrunneri
        Perca_fluviatilis
        Percopsis_transmontana
        Phycis_blennoides
        Phycis_phycis
        Poecilia_formosa
        Pollachius_virens
        Polymixia_japonica
        Pseudochromis_fuscus
        Rondeletia_loricata
        Salmo_salar_1
        Salmo_salar_2
        Sebastes_norvegicus
        Selene_dorsalis
        Spondyliosoma_cantharus
        Stylephorus_chordatus
        Symphodus_melops
        Takifugu_rubripes
        Tetraodon_nigroviridis
        Theragra_chalcogramma
        Thunnus_albacares
        Trachyrincus_murrayi
        Trachyrincus_scabrus
        Trisopterus_minutus
        Typhlichthys_subterraneus
        Xiphophorus_maculatus
        Zeus_faber
;
end;
begin characters;
        dimensions nchar=711;
        format datatype=dna missing=? gap=-;
        matrix
        Acanthochaenus_luetkenii
ATGATTACAAAACTA---------GACCGTGTGCTTTTGGCCAAGGAAACGTTCATCTTCCATTAT
GAGAACATGCGCTGGGCAAAAGGTCGGCATGAGACATACCTCTGCTTTGTAGTGAAGAGGC
GGGTGGGGCCAGACTCCCTGTCCTTTGACTTTGGACACCTCCGCAAC----------CGCACT
GGCTGCCAT GTAGAGCTGCTGTTCCTGCGCCACCTG------GGAACCTTGTGCCCTGGACTGT
GGGGGTACGGAGGCGCTGGAGAG---AGGAGGCTCAGTTACTCCATCACCTGGTTCTGC TCC
TGGTCCCCCTGCGCTGACTGCGCCTTCAGAGTGGCCCAGTTAATCGGCCGGACG---------CCC
AACCTCCGCCTCAGGATCTTCGTCTCTCGCCTCTACTTCTGCGACC TGGAGGACAGCCGCG
AGAGAGGGGGCCTGAGGTTGCTGAAGAAAGCTGGCGTGCAGATCACTGTCATGAGCTACA
AAGACTTTTTCTATTGCTGGCAGACCTTTGTGGCTAATGGAGGGAGCAGCTTCAAGGCCTG
GGACGAGATGCACCAAAACTCTGTTCGCCTGGCCAGC---------CAA---CTCAACCACATCCTG
CAGCCATGTGATACAGAGGAC TTAAGAGATGCATTCAAGCTTCTTGGTCTG----------
TGA
        Anabas_testudineus
ATGATTACAAAGCTA---------GACAGTGTGCTTTTGCCCCGAAAGAAGTTTATCTACCATTAC
AAGAATGTGCGCTGGGCGAGGGGTCGTCATGAAACATACCTCTGTTTCGTAGTGAAGAGGC
GGGTGGGCCCAGACTCCTTGACCTTTGACTTTGGACACCTCCGCAAT----------CGCAAT
GGCTGCCATGTGGAGATGCTGTTCTTGCGCTATCTG------GGAGCCTTATGTCCTGGTATTTG
GGGGTACGGAGGTGCTGGAGAG---AAAAGGCTCAGTTACTCAATTACCTGGTTCTGTTCCTG
```

GTCTCCTTGTGCCAACTGCTCCCTTAGGCTGACCCAGTTCCTCAGTCAGACC---------CCCAAC
CTCCGCCTCAGGATCTTTGTGTCCCGCCTTTACTTCTGTGACATGGAGGACAGCCGCGAGCG
GGAGGGTCTGAGGATACTGAAAAATGCTGGCGTGCAGATCACAGTCATGACTTACAAAGA
CTTCTTCTATTGCTGGCAGACCTTTGTGGATCGTAAACAGAGCAGCTTCAAAGCGTGGGATG
AGCTGCACCAAAACTCTGTTCGCCTCACCAGA---------AAA---CTCTACCGCATCCTTCAGCCC
TGTGAAATAGAAGATTTAAGAGATGCCTTCAAGCTTCTTGGGCTG----------TGA

Antennarius_striatus

ATGATTACGAAGCTT---------GACAGCGTGCTTTTGCCCCGAAAAAAGTTCATCTACCATTAT
AAGAACATGCGCTGGGCGAGAGGCCGGTGTGAGACGTACCTCTGCTTTGTAGTGAAGAGAC
GAGAGGGGCCAGACACCTTAACTTTTGACTTTGGACACCTCCGTAAT----------CGCAAT
GGCTGTCATGTGGAGCTACTTTTCTTACGCTATCTG------GGGGCCTTGTGCCCTGGATTGTG
GGGCAGTGGGGGTACTGGGGAG---AAGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG
GTCTCCCTGTGCCAACTGTTCCATCAGACAGTGTGAATTCCTGAGCCGAACG---------CCCAA
CCTTCGCCTCAGGATCTTTGTCTCTCGTTTGTACTTCTGCGACCTGGAGGATAGCCGTGAAA
GGGAAGGCCTAAGAATGCTGAAGAAAGCCGGCGTGCAGATCTCAGTCATGAGTTACAAAG
ACTTCTTCTACTGCTGGCAGACCTTTGTGGCTAGTAAACAAAGTAGTTTCAAGGCTTGGGAA
GAGCTGCATCAAAATTCAGTACGCCTTGCCAGA---------AAA---CTGAACCGCATCCTCCAGC
CGTGTGAAGCTGAAGATTTAAGAGATGCCTTTAAGCTTCTTGGACTG----------TGA

Arctogadus_glacialis

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTGGCCCAAAATAAATTCATCTACAATTAC
AAGAACATGCGATGGGCAAAAGGCCGCAACGAGACCTATCTCTGCTTCGTAATGAAGAGA
AGGCTTGGACCTGATTCCCTCTCTTTCGACTTCGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGCAGAGCTGCTGTTCCTGAGCTACCTG------GGGGCGCTGTGCCCGGGCCTCT
GGGGCTGCGCAGACGACAGAAAC---CGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT
GGTCGCCCTGTGCCAACTGTGCGACCACGCTGACCCGGTTCCTGAGGCAGACA---------CCAA
ACCTGCGACTCAGGATCTTCGTGTCTCGCCTCTACTTTTGTGACCTGGAGGGCAGTCCGCAT
GTAGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTATGTGCGTCTGTCAAGA---------AAAAA?C?AAACCGCATCCTCC
AGCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Astyanax_mexicanus

ATGACGAGCAAGCTG---------GACAGCATTCTGCTCACCCAGAAGAAGTTTATCTATCACTAC
AAGAACGTGCGCTGGGCTCGTGGGAGGCATGAGACTTACCTCTGCTTCGTGGTGAAGAGGC
GAATCGGACCAAACTCGCTGTCCTTCGACTTCGGGCACCTGCGCAAC----------CGCTCC
GGCTGCCACGTGGAGCTCCTCTTCCTGCGCTACCTG------GGGGCACTGTGCCCGGGCCTGG
GGGGTCTGGGTGTGGACGGAGTG------AAGGTGGGCTACGCTGTGACCTGGTTCTGCTCATG
GTCGCCCTGCTCTAACTGCGCCCAGCGAATCGCCCACATCCTGTCCCAGACG---------CCCAG
CCTGCGACTCCGCATCTTCGTCTCCCGCCTGTACTTCTGCGACAACGAGGACAGCCTGGAGC
GGGAGGGGCTGCGGCACCTGCTGAGGGCAGGGGTGCAGATTACAGTCATGACGTATAAAG
ATTTTTTCTACTGTTGGCAGACGTTTGTGGCTCGCAGGGAGAGTCGCTTTAAAGCCTGGGAC
GGTCTTCACCAAAACTCTGTCAGACTGTCCCGC---------AAA---CTCAAACGCATCCTCCAGCC
CTGTCAGACTGAAGATCTGAGGGACGTCTTCGCTCTGCTGGGTCTC----------TGA

Bathygadus_melanobranchus

ATGATTAGTAAGCTC---------GACAGTGTGCTTTTGGCCCAGAAAAAATTCATGTACAATTAC
AAGAACGTGCGCTGGGCAAAAGGCCGCCACGAGACCTACCTCTGCTTCGTAGTGAGGAGA
AGGCTCGGACCAAATTCCCTGTCTTTTGACTTCGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGTAGAGCTGCTGTTTCTGAGCCACCTG------GGGGCGCTCTGCCCAGGCCTCT
GGGGGTGCGTAGGTGATGACAAC---AGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCT
GGTCTCCCTGCGCCAACTGTGCGGCCACACTGGCCCGGTTCCTGAGGCAGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCGAAT
ATAGAGGGCTTGAGAGAGCTGAGGAGGGCAGGAGTCCAGGTCATCGTTATGAGCTACAAA

GACTACTTCTACTGCTGGCAGACATTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGAGTTATTGGGCTGTTAAGC----------
TGA

Benthosema_glaciale

ATGATTACTAAACTA---------GACAGTGTGCTTTTGGGTCAGAAGAAGTTCCTCTTCCACTAT
AAGAACGTGCGCTGGGCGTGGGGTCGAAATGAAACGTACCTCTGCTTTGTGGTGAAGAGGA
GAGTAGGACCAAACTCCCTCTCCTTTGACTTTGGACATCTCCGCAAC----------CGCTCC
AGCTGCCACGCGGAGCTGCTGTTCCTTCGCCACCTG---GGGGGCGCCCTGTGCCCTGGTCTG
TGGGGCTACGGAGGTGACGGGGGAGAGGGGAGGTTCAACTACTCGGTCACCTGGTTCTGCT
CGTGGTCTCCGTGCGCCGACTGTTCTCTGAGACTGGCCCAGTTCCTCAGCCGGACC---------C
CCAACCTGCGCCTCCGCATCTTCGTCTCTCGCCTCTACTTCTGTGACGCGGAGGACAGCCGG
GAGAGGGAGGGTCTGAGGACGCTGAAAAGGGCAGGTGTACAGATCACCGTCATGAACTAC
AAAGACTACTACTATTGTTGGCAGACCTTTGTGGCTCACAGACAGAGCAGCTTCAAGGCCT
GGGCTGATCTGCACCAGAACTCTGTCCGTCTGGCCAGG---------AAA---CTCCACCGCATCCTC
CAGCCTTGTGAGACAGAGGATTTTAGAGACGCATTCAAGCTTCTTGGGTTG----------
TGA

Beryx_splendens

ATGATTACAAAACTA---------GACAGTGTGCTTTTGGCCAAGAAAAAGTTCATCTACCATTAC
AAGAACATGCGCTGGGCAAAGGGCCGGCATGAGACATACCTCTGCTTTGTGGTGAAGAGG
CGAGTGGGGCCAGACTCCCTGTCCTTCGACTTCGGACACCTCCGCAAC----------CGCGC
TGGCTGCCATGTAGAGCTGCTGTTCCTGCGCCACCTG------GGAGCCCTGTGCCCTGGACTGT
GGGGGCATGGAGGCAGCGGAGAG---AGGAAGCTGAGTTACTCCATCACCTGGTTCTGCTCC
TGGTCTCCCTGCGCTGACTGCTCCTTCAGACTGGCCCAGTTCCTCAACCGGACG---------CCC
AACCTCCGCCTCAGGATCTTCGTCTCCCGCCTCTACTTCTGCGACCAGGAGGACAGCCGCGA
GAGAGACGGCCTGAGGCTGCTGAAAAAGGCCGGCGTGAACATCACTGTCATGAGCTACAA
AGACTTCTTCTACTGCTGGCAGACCTTTGTGGCTAACAGAACGAGCAGATTCAAGGCCTGG
GATTTGCTGCACCAAAACTCTGTTCGCCTGGCCAGG---------AAA---CTCAACCGCATCCTCCA
GCCTTATGAGATAGAAGATTTAAGAGATGCCTTCAGACTTCTTGGTTTT----------TGA

Boreogadus_saida

ATGATTAGGAAGCTA---------GACAGTGTGCTCTTGGCCCAGAATAAATTCATCTACAATTAC
AAGAACATGCGATGGGCAAAAGGCCGCAACGAGACCTATCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCTGATTCCCTCTCTTTCGACTTCGGACACCTACACAAT----------CGCAC
TGGCTGCCACGCAGAGCTGCTGTTCCTGAGCTACCTG------GGGGCGCTGTGCCCGGGCCTCT
GGGGCTGCGCAGACGACAGAAAC---CGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT
GGTCGCCCTGTGCCAACTGTGCGACCACGCTGACCCGGTTCCTGAGGCAGACA---------CCAA
ACCTGCGACTCAGGATCTTCGTGTCTCGCCTCTACTTTTGTGACCTGGAGGGCAGTCCGCAT
GTAGAGGGCTTGAGGGACCCGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTATGTGC?TCTGTCAAGA---------AAA---CTAAACCGCATCCTCCAG
CCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Borostomias_antarcticus

ATGATCAGTAAACTA---------GACAGTGTTCTCCTGGCCCAGAAGAAGTTCCTCTTCCACTAC
AAGAACGTGCGCTGGGCCCGAGGACGACACGAGACGTACCTGTGCTTCGTGGTGAAGAGG
AGGGTGGGACCCGACTCGCTTACCTTCGACTTCGGACACCTGCGCAAT----------CGCA
CCGGCTGCCACGTTGAGCTGCTGTTCCTGCGCCATCTA------GGGGTGCTGTGTCCGGGCCTG
TCGGCGTCTGGAGGTGCTGGAGGGGGCAGGGGGCTGAACTACTCCATCACCTGGTTCTGCT
CATGGTCCCCCTGCTTCGACTGCTCGGCCCGGCTGGCCCAGTTCCTGAGACGGACC---------C
CCAACCTCAGGCTCCGCCTCTTCGTCTCCCGCCTCTACTTCTGTGACCCGGAGGACCGCCAC
GAGAGAGAGGGGCTCCGGGCGCTGAAGAGAGCCGGAGTCCACATCACCGTCATGAGCTAT
AAAGATTATTTTTACTGCTGGCAGACGTTTGTAGCTCACAGACAGAGGGCCTTCAAAGCCT
GGGAAGATCTTCAGCAGAACTCCGTCCGCCTGGCCAGG---------AAG---CTCAACAGCATCCT

GCTGCCCTGTGAGACGGAGGATCTGAGAGACCCGTTCAGGCTGCTTGGACTG----------
TGA

Brosme_brosme

ATGATGAGTAAGCTA---------GACAGTGTGCTCTTAGCCCAGAAGAAATTCATATACAATTAC
AAGAACCTGCGATGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGA
AGGCTCGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAAT----------CGTAC
TGGCTGCCACGTAGAGCTGCTGTTTCTGAGCTACCTG------GGGGCGCTGTGCCCAGGCCTCT
GGGGGTGCGGTGGCGACAGAAAC---CAAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCT
GGTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCACACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCAT
ATAGAGGGCTTGAGGGACCTGAGGAGAGCAGGGGTCCAGGTCAAAGTTATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGACTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTCTGTCCGTCTGTCAAGA---------GCT---CTAAACCGCATCCTCCAG
CCATGTGAAACAGAAGATTTAAGAGATCCTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Brotula_barbata

ATGATTGCGAAACTA---------GACAGTGTACTTTTACCACGGAAAAAGTTCATCTACCATTTC
AAGAACATGCGCTGGGCTAAGGGTCGGCATGAGACGTACCTGTGCTTTGTGGTGAAGAGGC
GAGTAGGGCCGGACTCGCTGTCCTTTGACTTTGGACACCTCCGCAAT----------CGCAAT
GGCTGCCACGTAGAGCTACTGTTCTTACGCTACCTA------GGAGCTTTATGCCCTGGACTGTG
GGGCTGTGGGAATTCTGGACAG------AGGTTGTGTTACTCCATCACTTTGTTCTGCTCTTGGT
CCCCCTGTGCCAACTGTTCCGAGAGACTGGCCAAGTTCCTCGGCCGGACA---------CCCAACC
TTCGCCTCAGGATCTTTGTCTCTCGCCTCTACTTCTGCGACATGGAAGACAGCCGTGAAAGA
GAGGGTCTGAGGATGCTGAAAAATGCTGGCGTAAACATCACAGTCATGAGCTACAAAGAC
TATTTCTATTGCTGGCAAACCTTTGTGGCTCGTGGCGCAAGCAACTTCAAAGCCTGGGATGG
GCTGCAAGAGAATTCAATTCGCCTTGCCAGG---------AAA---CTCACCCACATCCTACAGCCA
GGTGAGACGGAAGATTTAAGGGACGCATTCAAACTTCTGGGTATG----------TGA

Carapus_acus

ATGACTGCCAAGCTA---------GACAGGGTCCTTTTGCCACGGAAAAAGTTCCTCTTCCATTAC
AAGAACGTGCGCTGGGCGAAGGGCCGCCACGAGACGTACCTCTGCTTCGTGGTGAAGAGG
CGAGTGGGTCCAGACTCCATGTCCTTTGACTTTGGACACCTCCGCAAT----------CGCAG
TGGCTGCCACGTAGAGCTCTTGTTCCTGCGCTACCTG------GGAGCTCTGTGTCCTGGACTGT
GGGGGTATGAAGGTTCTGGACAG---AGGAGACTCAGCTACTCCATCACCTGGTTCTGCTCTT
GGTCCCCGTGCGCCAACTGCTCGGAGCGACTCGCCCAGTTCCTCAATCGGACC---------CCCA
ACCTCCGCCTCAGGATCTTCGTCTCTCGCCTCTACTTCTGCGACCTGGAGGACAGCCGTGAG
AGGGAGGGCCTGAGGACGCTGGAGAAAGCTGGCGTGCACATCACCATCATGAGCTACAAA
GACTATTTCTACTGCTGGCAAACCTTTGTGGCTTGTGGAACTTCAAAATTCAAAGCCTGGGA
TGAGCTCCACCAAAACACCACTCGTCTCAAGAGA---------AAA---CTGAATCGGATCCTCCAG
CCATGTGAGACAGAAGATTTAAGGGACGCATTCAAACTTCTAGGGTTGCTG----------TGA

Chaenocephalus_aceratus

ATGATCACAAAGCTT---------GACAGCATGCTTTTGCCTCGAAAAAAGTTCATCTACCATTAC
AAGAACATGCGCTGGGCAAGGGGCCGGTGTGAGACATACCTCTGCTTTGTAGTGAAGAGGC
GGGTGGGACCAGACTCCTTAACCTTTGACTTCGGACACCTTCGCAAT----------CGCAAT
GGCTGCCATGTAGAGATGCTGTTCCTGCGCTACCTG------GACGCCCTGTGCCCTGGTCTGTT
GGGATGTGAAGGTACTGGAGAG---AAGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG
GTCCCCCTGTGCAAACTGCTCCATCAGGCTGTCCCAGTTCCTCAGCCAGACG---------CCCAA
TCTTCGCCTCAGGATCTTCGTCTCTCGTCTTTACTTCTGTGACATGGAGAATAGCCCTGCAA
GAGACGGCCTAATAATGCTGAAAAAAGCTGGCGTGCAGACTTCAGTCATGAGTTACAAAG
ACTTTTTCTATTGCTGGCATAACTTTGTGGATTGTAAACAGAGTAAATTCAAGCCATGGGAA
GATCTGCACCAAAACTCTGTTCGCCTTGCCAGA---------AAA---CTCAAACGCATCCTTCAGCT
GTGTGAAACTGAAGATTTGAGAGATGCCTTCAAGCTTCTTGGACTG----------TAA

Chatrabus_melanurus

ATGATTACAAAACTA---------GACAGTGTGCTTTTGCCACGGAAGAAGTTCATCTACCATTAC

AAGAACATGCGCTGGGCAAAGGGCCGGCACGAGACATACCTCTGCTTTGTGGTGAAGAGA
CGAATGGGGCCAGACTCCCTGTCCTTTGATTTCGGACACCTCCGCAAT----------CGCAA
CGGCTGCCATGTAGAGCTGCTGTTCCTGCGTTACCTG------GGAGCCTTGTGCCCTGGTCTGT
GGGGGTATGGAATTGCTGGAGAG---AGGAAGCTTAGTTACTCCGTCACCTGGTTCTGCTCCT
GGTCCCCCTGTGTCAACTGCTCCCTCAGACTGACACAGTTCCTCATGCAGACG---------CCTA
ATCTTCGCCTCAGGATCTTCGTCTCTCGCCTTTACTTCTGTGATATGGAAGACAGCCGTGAG
AGAGAAGGTCTGAGGATGCTGAAAAAAGCCGGCGTGCACATCACAGTGATGAGTTACAAA
GACTTCTTCTACTGCTGGCAGACCTTTGTGGCTTGTAAAGAGAGCAAATTCAAGGCATGGG
AGGCGCTGCACCAAAACTCTGTTCGTCTGGCTAGA---------AAG---CTCAACCGCATCCTCCA
GCCCTGTGAGACAGAAGACTTCAGAGATGCCTTCAAGCTTCTTGGACTG----------TGA

Chromis_chromis

ATGATCACAAAACTC---------GACAGTGTGCTTTTGCCCCAGAAGAAGTTCATCTACCATTAT
AAGAACATGCGCTGGGCGAGAGGCCGCTGTGAGACGTACCTCTGCTTCGTGATTAAGAAAA
GAGCCGGTCCAGATTCTATATCCTTCGACTTCGGACATCTACGGAAC----------CGCAAC
GGCTGCCATGTAGAGCTGCTGTTCCTGCGCTACCTG------GGCGCCTTGTGTCCTGGTCTCTG
GGGTTATGGACAG----------AACCGGATCAGCTACTCCATCACCTGGTTCTGCTCCTGGTCTC
CCTGCGCTAACTGCTCCCTCAGACTGGCCCAGTTCCTGAACCAGACG---------CCCAACCTTC
GTCTCCGGATCTTCGTCTCTCGGCTCTACTTCTGCGACATGGAGGACAGCCGGGAGAGGGA
AGGTCTGAGGATCCTGAAGAAGGCCGGCGTTAACATCACCGTCATGAGCTACAAAGACTAC
TTCTACTGCTGGCAGACCTTCGTGGCTCGGAGGCTGAGTAAGTTCAAACCGTGGGACGGGC
TGCAACAGAACTACGTCCGTCTGTCCAGA---------AAA---CTGAACCGCATCCTGCAGCCCTG
TGAGACTGAAGACTTTCGAGACGCCTTCAGGCTCCTTGGACTC----------TGA

Cyttopsis_roseus

ATGATTACTAAACTA---------GACAGTGTGCTTCTGGCTCGGAAGACATTCATTTACCACTAT
AAGAACATGCGCTGGGCAAAAGGCCGGCATGAGACATACCTCTGCTTCGTCGTCAAGAGA
AGAGTTGGACCCGATTCCTTGTCCTTTGACTTTGGACACCTTCGCAAT----------CGGAC
TGGCTGCCATGTAGAGCTCCTGTTTCTACGTCACCTG------GGGGCCCTGTGCCCTGGACTGT
GGGGACAAGGAGGCGCTGATGAA---AGAAGGCTCAGTTACTCGGTCACCTGGTTCTGCTCC
TGGTCTCCCTGCGCCAACTGCTCCCTCAGACTGGTCCAATTCCTCGGGCAGACG---------CCC
AACCTCCGTCTCAGGATCTTCGTCTCCCGTCTCTACTACTGTGACCTTGAAGACAGCCGCGA
GAGAGAGGGCTTAAGAACCCTGAAAAGAGCCGGAGTCCAAATCACAGTCATGAGCTACAA
AGACTATTTCTATTGCTGGCAGACGTTCGTGGCTCGCCGACAGACCCGCTTCAAGGCGTGG
GATGAGCTGCACCAAAACTCAGTTCGTCTGGCCAGG---------AAA---CTAAACCGCATCCTCC
AGCCTTGTGAAACGGAAGATTTAAGAGATGCTTTCAAACTTCTCGGGTTCTTG----------
TAA

Danio_rerio

ATGATCTGCAAGCTG---------GACAGTGTGCTCATGACCCAGAAGAAATTCATCTTCCACTAT
AAGAATGTGCGCTGGGCTCGAGGGAGACACGAAACCTACCTTTGTTTTGTAGTAAAGCGAC
GCATCGGCCCTGATTCCCTCTCTTTTGACTTTGGACACCTGCGCAAT----------CGCTCC
GGATGCCATGTAGAGCTTCTCTTTCTGCGTCACTTG------GGTGCGTTGTGTCCGGGCCTGAG
CGCTTCCAGTGTGGACGGTGCA------AGATTGTGTTACTCAGTGACCTGGTTCTGCTCCTGGT
CGCCCTGCTCTAAATGCGCTCAACAGCTCGCCCACTTCCTGTCACAGACG---------CCCAATC
TGAGGCTGAGGATCTTTGTGTCACGCCTGTACTTCTGTGATGAAGAGGACAGCGTGGAGAG
AGAAGGTCTGCGACACCTGAAGAGGGCAGGAGTTCAGATCTCGGTCATGACTTATAAAGAC
TTTTTCTACTGCTGGCAAACGTTTGTTGCGAGGAGGGAGCGGAGTTTTAAAGCCTGGGATG
GACTTCATGAAAACTCTGTCCGGCTTGTTCGG---------AAA---CTCAATCGGATTCTGCAGCCT
TGCGAGACTGAGGATCTGAGGGATGTTTTTGCTCTTCTTGGGTTA----------TGA

Gadiculus_argenteus

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTGGCCCAGAAGAAATTCATATACAATTAC
AATAACATGCGATGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTCGTGAAGAGA
AGGCTTGGACCTGACTCCCTCTCCTTCGACTTCGGACACCTACGCAAT----------CGCAC
CGGCTGCCACGCAGAGGTGCTGTTCCTGAGCTACCTC------GGGGCACTGTGTCCGGGCCTCT
GGGGCTGCGCAGGCGACAGAAGC---CTAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT

GGTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTCGCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCAT
GTGGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTTATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTCTGTGCGTCTGTCAAGG---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Gadus_morhua

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTGGCCCAGAAAAAATTCATCTACAATTAC
AAGAACATGCGATGGGCAAAAGGCCGCAACGAGACCTATCTCTGCTTCGTAGTAAAGAGA
AGGCTTGGACCTGATTCCCTCTCTTTTGACTTCGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGCAGAGCTGCTGTTTTTGAGCTACCTG------GGGGCGCTGTGCCCGGGCCTCT
GGGGCTGCGCAGACGACAGAAAC---CGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT
GGTCGCCCTGTGCCAACTGTGCGACCACGCTGACCCGGTTCCTGAGGCAGACA---------CCCA
ACCTGCGACTCAGGATCTTCGTGTCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCAT
GTAGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTATGTGCGTCTGTCAAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Lampetra_tridentata

ATGGCCAACGATGAGTACGTGAGAGTCGGCGATAAGTTGGACAGCTGCACGTTTAGGACGC
AGTTTTTTAACTTTAAAAGATCCACGTCG---CATATATGCTGCGTTCTCTTTGAATTAAAACA
GCAGGATAGCGTCGCT----------TTTTGGGGCTATGCTGTGAATAAACCACGGAGCAATGCA
GACCTAGGAATTCACGCCGAAATTTTTTGCATTAAAAAAATC------AGAGAGTAC----------
----------CTGCACGAA---AACCCTGGAATATACACGATAAATTGGTACTCATCCTGGAGTTCG
TGTGCAGATTGCGCTGAAGAGATCTTAACATGGTATAAGAAGGAGGTGATGAAGATGGGC
CACACTTTGAATATCTGGGCTTGCAAACTCTATTTCGAGAACATT------ACGCGGAATCAAAT
TGGGTTGTGGAACCTCAGAAAAATCGGGGTTGGGTTGGAAATAATGCTTGGTGAACACTAC
CAATGGTGCTGGAACAACTACATCCAAACGTTGGACAGCAATTTGAATGAAAATAGATGGC
TTCAGAAGACTTCGAATCGAGCTCTTACACGACAGAACGAG---TTGTCCATTATGATTCAG--
----------GTAAAAAGACTCCACACCGCTAAGACTCCTGCTGTTTAG

Gasterosteus_aculeatus

ATGATTGCAAAGCTT---------GACAGTGTGCTTCTGCCCCGAAAAAAGTTCATCTACCACTAC
ACGAACATGCGCTGGGCGAGGGGCCGACACGAGACTTACCTCTGCTTTGTTGTGAAAAGGC
GAGTGGGGCCGGATTCCTTGTCCTTCGACTTTGGACACCTGCGCAAT----------CGCAGT
GGCTGCCATGTCGAGTTGTTGTTCCTGCGCCACCTC------GGAGCCTTGTGCCCTGGTTTCTT
GGGTTGTGGAGACACCGGAGGG---AGGAGGCTGAGTTACTCCATCACCTGGTTCTGCTCGTG
GTCTCCCTGCGTAAACTGCTCCATCAGTCTGTCCCAGTTCCTCAGCCGGACG---------CCCAA
CCTCCGCCTCAGGATCTTCGTCTCTCGCCTTTACTTTTGTGACATGGAGAACAGTCGGGAAA
GAGACGGCCTGAGAATGCTGAAAAAAGCTGGCGTGCAGGTCACAGTCATGAGTTACAAAG
ATTTCTTCTATTGCTGGCAGACTTTTGTAGATCGCAAACAAAGCCAGTTCAAGGCCTGGAAA
GAGCTTCACCAAAACTCTGTTCGCCTTTCCAGA---------AAG---CTCAAGCGCATCCTCCAGCC
TTGTGAAACAGAAGATTTAAGGGATGCCTTCAAGCTGCTTGGACTG----------TGA

Guentherus_altivela

ATGATTACTAAACTA---------GACAGCATACTTATGGCCCAGAAGAAGTTCATCTTCCACTAT
AAGAACATGCGATGGGCCAAGGGTCGAAATGAGACACACCTCTGCTTTGTGGTGAAGAGA
AGGCTGGGACCAAACTCCCTGTCCTTTGACTTTGGACACCTGCGTAAT----------CGCAC
TGGCTGCCATGTAGAGCTACTCTTCTTGCGCCACCTG------GGATTCCTGTGCCCTGGCTTGT
GGGGGTACGGAGAGCCAGGTGAA---GGGAGGCTGAATTACTCTGTCACCTGGTTCTGCTCCT
GGTCCCCCTGTGCAGATTGTTCCTTCACGCTGACCCACTTCCTCAGAGAGACT---------CCCA
ACCTCCGTCTTAGAATCTTTGTGTCTCGCCTCTACTTCTGTGACGAGGAGGACAGCAGTGCA
AGGGAAGGCCTGCGAATGTTGAAGAAAGCCGGTGTGAACATCACTGTCATGAGCTACAAA

GACTACTTCTATTGCTGGAAGACCTTTGTGGCTCACAGACAAAGGAACTTCAAGGCCTGGG
ATGGGCTAGACCAGAACTCTGTTCACCTAGCCAGG---------AAA---CTCAGCCACATCCTCCA
GCCCTGGGAAACAGCAGATTTAAGAGATGCCTTTAAACTTCTTGGACTG----------TGA

Helostoma_temminckii

ATGATTACAAAGCTA---------GACAGTGTGCTTTTGCCCCGAAAAAAGTTCATCTACCATTAC
AAAAATGTGCGCTGGGCAAGGGGTCGGCATGAGACATACCTCTGTTTTGTAGTGAAGAGGC
GGGTGGGCCCAGACTCCTTGACCTTTGACTTTGGGCATCTCCGCAAT----------CGCAAT
GGTTGCCATGTAGAGATGCTGTTCTTGCGATATCTG------GGAGCTTTGTGCCCTGGACTTTG
GGGGTGTGGAGGTACTGGAGAG---AGAAGGCTCAGTTACTCTATTACCTGGTTCTGCTCCTG
GTCTCCTTGTTCTAACTGCTCCCTTAGACTGGCCCAGTTCCTCAGTCAGACC---------CCAAAC
CTCCGCCTCAGGATCTTTGTGTCTCGCCTATACTTCTGTGACATGGAGGACAGTCGCGAGAG
GGAGGGTCTCAGGATCCTGAAAAACGCTGGAGTGCAGATCACAGTCATGAGTTACAAAGA
CTTCTTCTACTGCTGGCAGACATTTGTGGCACGTAAGCAGAGCAACTTCAAAGCATGGGAG
GAGCTGCACCAAAACTCTGTTCGCCTTACCAGA---------AAA---CTCCATCGCATCCTTCAGCC
TTGTGAAACAGAAGATTTAAGAGATGCTTTCAAGCTCCTTGGACTG----------TGA

Holocentrus_rufus

ATGATTACAAAACTA---------GACAGTGTGCTTTTGGCCAAGAAAAAGTTCATCTACCATTAT
AAGAACTTGCGCTGGGCAAAAGGCCGGCATGAGACATACCTCTGCTTTGTCGTGAAGAGGC
GGGCGGGGCCGGACTCCATCGCCTTCGACTTTGGACACCTCCGCAAC----------CGTGCT
GGCTGCCATGTAGAGCTGCTATTCCTTCGCTACCTG------GGAGCCTTGTGCCCTGGACTGTG
GGGCTACGGAGGAACTGGTGAG---AGGAAGATGAGCTACTCCATCACATGGTTCTGCTCCT
GGTCTCCTTGTGCCAACTGCTCCTACAGACTCGCCCAGTTCCTCAACCGGACG---------CCCA
ACCTCCGCCTCAGGCTCTTCGTCGCTCGCCTCTATTTCTGTGACATCGAGGACAGCCGTGAG
AGAGAGGGCCTGAGAATGCTGAAGAATGCCGGTGTGCACATCACTGTCATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACATTTGTGGCTCGTAAAACGAGCAACTTCAAGGCCTGGG
ATGGGCTGCACCAAAACTATGTTCGTCTGGCCAGG---------AAA---CTCAACCGCATCCTCCA
GCCTTGTGAGACAGAAGATTTAAGAGATGCATTCAGGCTTCTTGGCTTG----------TGA

Laemonema_laureysi

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTTCAATTAC
AAGAACATGCGCTGGGCAAGAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCCAATTCCCTGTCTTTCGACTTCGGACACCTACGCAAT----------CGCAC
AGGCTGCCATGTAGAGCTGCTGTTTTTGAGCTATCTG------GGGGCACTGTGCCCAGGCCTGT
GGGGGTGCAGAGGCGACGAAAAC---AGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCC
TGGTCTCCATGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACG---------CCC
AACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCCCA
TATAGAGGGCTTGAGGGACCTGAGGAGAGCAGGGGTGCGGGTCACCGTTATGAGCTACAA
AGACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGG
GAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------AAA---CTAAACCGCATCCTCC
AGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGGCTGTTAACC----------
-TGA

Lampris_guttauts

ATGATCAGCAAACTA---------GACAGTGTGCTTCTGACCCAGAAGAAGTTCCTCTACCATTAT
AAGAACGTGCGTTGGGCAAAAGGTCGGCATGAGACATATCTCTGCTTTGTGGTGAAGAGGA
GGGTGGGACCGGACTCCATGTCCTTCGATTTTGGACACCTCCGCAAT----------CGAGCT
GGCTGCCATGTAGAGCTGCTGTTCCTGCGCTACCTG------GGGGCCCTGTGTCCTGGACTGTG
GGGCTACGGGGACACCGGAGAC---AGGAGGCTCAGTTACTCGGTCACCTGGTTCTGCTCCT
GGTCTCCCTGCGCCAACTGCTCCTTCAGACTGGCCCAGTTCCTCCAAAGGACG---------CCCA
ACTTCCGCCTCAGGCTCTTTGTCTCCCGTCTGTACTTCTGTGACATGGAGGACAGCAGTGAG
AGGGACGGCCTGAGGTTGCTGAAAAACGCAGGGGTGCAGATCACCGTCATGAGCTACAAA
GACTACTTCTATTGCTGGCAGACTTTTGTGGCTCACAGAAAGAGCAGTTTCAAGGCCTGGG
ATGGGCTGCACCAAAACACTGTTCGCTTGGCCCGG---------TTA---CTCAACCGCATCCTCCAG
CCTTGTGAGGCAGAGGATTTGCGGGATGCGTTCAAACTTCTCGGGTTT----------TGA

Lamprogrammus_exutus

ATGATTGCAAAACTA---------GACAGTGTGCTTTTGCCCCGCAAAAAGTTCATCTTCCATTAC
AAGAACATGCGCTGGGCTAAGGGTCGGCACGAGACATACCTCTGCTTTGTAGTGAAGAGAC
GAGTGGGTCCAGACTCCCTGTCCTTTGACTTTGGACACCTCCGCAAT----------CGCAAT
GGCTGCCATGTAGAGCTACTGTTCCTGCGCTACCTG------GGAGCTCTATGCCCTGGACTGTG
GGGGTGTGGAGGTTCTGGTGAG---AGGAGACTCAGTTACTCCATCACCTGGTTCTGCTCTTG
GTCCCCCTGTGCCAACTGCTCCCAGAGACTATCCCAATTCCTCAGCCAGACA---------CCCAA
CCTTCGCCTCAGGATCTTTGTCTCTCGCCTCTACTTCTGTGACATGGAGAACAGCCGTGAGA
GAGAGGGCCTGAGGATGCTGAAAAATGCTGGTGTGCAAATCACAGTCATGAGCTACAAAG
ACTTTTTCTATTGCTGGCAAACCTTTGTGGCTTGTGGGAAAAGCAAATTCAAGGCCTGGGAT
GAGCTGCACCGAAACTCTGTTCGCCTCACCAGG---------AAA---CTGAACCGCATCCTCCAGC
CATGGGAGACAGAAGATTTAAGAGATGCATTCAGACTTCTTGGATTT----------TGA

Lesueurigobius_cf_sanzoi

ATGATTACCAAGCTA---------GACAGTGTACTTTTACCAAAGAAGAAGTTTATCTTCCATTAC
AAGAACGTGCGCTGGGCGAAGGGTCGGCATGAGACGTACCTCTGCTTTGTGGTCAAGAGGC
GCGTGGGGCCAAATTCTATGTCCTTTGACTTTGGACATCTTCGCAAT----------CGCAGC
GGCTGCCATGTGGAGATTCTGTTCCTGCGTTACCTT------GGTGCTCTGTGCCCTGGACTCTG
GGGGGCTGGAGGCTCGGAGGAG---AGGCGACTGAGTTACTCCATCACTTGGTTCTGCTCCT
GGTCTCCATGCGCCAACTGCTCCACGAAACTGTCGCAGTTCCTCGCCAAAACC---------CCAA
ACTTGCGTCTGCGGATATTTGTCTCACGCCTTTACTTCTGCGACCTGGAGGACAGCATAGAA
CGAGAGGGTCTGAGGATGCTAAAGAGAGCAGGCGTGCAGTTAACGGTCATGAAATACAAA
GACTACTTTTACTGCTGGCACACGTTTGTGGCTCGAAACCAAAGCAACTTCAAGGCCTGGG
AAGAGCTTCACCAAAACTCAGTGCGACTGACCAGG---------AAA---CTCAGTCGCATCCTTCA
GCCATGTGAGACAGAGGATTTAAGAGATGCCTTCAGACTTCTTGGTTTG----------TGA

Lota_lota

ATGATAAGTAAGCTA---------GACAGTGTGCTCTTAGCCCAGAAGAAATTCATATACAATTAC
AAGAACATAAGATGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGTAGAGCTGCTGTTTCTGAGCTACCTG------GGGGCGCTGTGCCCGGGCCTCT
GGGGGTGCGGAGGCGACAGAAAC---CGAAGACTCAGCTACTCGGTCACCTGGTTTTGCTCC
TGGTCTCCCTGTGCCAACTGTGCGGCTACACTGGCCCGGTTCCTGAGGCAGACG---------CCC
AACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCA
TATAGAGGGCTTGAGGGACCTGAGGAGAGCCGGGGTCCAGGTCAAAGTTATGAGCTACAA
AGACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCATGG
GAAGGGCTGCATACCAATTCGGTCCGTCTGTCAAGA---------AAA---CTAAACCGCATCCTCC
AGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTTTTGGACTGCTAACC----------
-TGA

Macrourus_berglax

ATGATTAGTAAGCTT---------GACAGCATACTCTTGGCCCAGAAGAAATTCAAGTACAATTAC
AATAACATGCGATGGGCAAAGGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGA
AGGCTCGGACCCAATTCACTGTCCTTTGACTTCGGACACCTACGCAAT----------CGTGC
TGGCTGCCACGTAGAGCTGCTGTTTCTGAGCCACCTG------GGGGCGCTGTGCCCGGGCCTGT
GGGGCTTTGGAGGGGCAGAAAAC---ATAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCT
GGTCTCCCTGCGCCAACTGTGCGGCCACACTGGCCCGGTTCCTGAGGCAGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGAACTGGCGGACAGTCCGCAC
TCAGAGGGCTTGAGGGAGCTGAGGAGAGCAGGGGTCCAGGTCAACGTTATGACCTACAAA
GACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------AAA---CTAAACCTCATCCTCCAG
CCATGTGAAACAGAAGATTTAAGAGACGCTTTCAGACTTATTGGCCTGTTAACC----------
TGA

Malacocephalus_occidentalis

ATGATTAGTAAGCTC---------GACAGCGTGCTCTTGGCCCAGAAGAAATTCATATACAATTAC
AAGAACATACGCTGGGCAAAGGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCCAATTCACTGTCCTTCGACTTCGGACACCTACGCAAC----------CGCAC

TGGCTGCCATGTAGAGCTGCTGTTTCTGAGCTACTTG------GGGGCGCTGTGCCCGGGCCTGT
GGGGCTGTGGAGGTGCAGATAAC---AGAAGACTCAACTACTCGGTCACCTGGTTCTGCTCC
TGGTCTCCCTGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACG---------CCC
AACCTGCGCCTCAGGATCTTTGTGGCTCGCCTCTACTTCTGCGACCTGGACGACAGTCCACA
CACAGAGGGCTTAAGGGAGCTGAGGAGAGCAGGGGTCCAGTTCACCGTAATGAGCTACAA
AGACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGG
GAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------AAA---CTAAACCGCATCCTCC
AGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGGCTGTTATCC----------
TGA

Melanogrammus_aeglefinus

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTGGCCCAGAAGAAATTCATCTACAATTAC
AAGAACATGCGATGGGCAAAGGGCCGCAACGAGACCTATCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCTGATTCCCTCTCTTTCGACTTTGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGCAGAGCTGCTGTTCCTGAGCTACCTG------GGGGCACTGTGCCCAGGCCTCT
GGGGCTGTGCAGGCGACAGAAAC---CGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT
GGTCGCCCTGTGCCAACTGTGCGACCACGCTGACCCGGTTCCTGAGGCAGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGTCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCAT
GTAGAGGGCTTGAGGGATCTAAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTATGTGCGTCTGTCAAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Melanonus_zugmayeri

ATGATTAGTAACCTA---------GACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTAC
AAGAACATGCATTGGGCAAAAGGCCGCAACGCGACCTACCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACACAAT----------CGCAC
TGGCTGCCACGCAGAGCTGCTGTTTCTCAGCCACCTG------GGGGCACTGTGCCCAGGCCTGT
GGGG?TGCGGAGGCGACAAAAAC---AGAAGACTCAGCTATTCGGTTACCTGGTTCTGCTCCT
GGTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGCTTTCTGAGGCAGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGGCTCGCGTCTACTTCTGTGAACAGGAGGACAGTCCGCAT
ATAGAGGGCTTGAGGGATCTGAGGAGAGCAGGGGTCCAGGTCACCGTTATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGACCTGGG
AAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGACTGTTAACC----------
TGA

Merlangius_merlangus

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTGGCCCAGAAGAAATTCATCTACAATTAC
AAGAACATGCGGTGGGCAAAAGGCCGCAACGAGACCTATCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCTGATTCCCTTTCTTTCGACTTCGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGCAGAGCTGCTGTTCCTGAGCTACCTG------GGGGCACTGTGCCCAGGCCTCT
GGGGCTGCGCAGGCGACAGAAAC---CGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT
GGTCGCCCTGTGCCAACTGTGCGACCACGCTGAGCCGGTTCCTGAGGCAGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGTCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCAT
GTAGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTATGTGCGTCTGTCAAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Merluccius_merluccius

ATGATTAGTAAGCTC---------GACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTAC
AAGAACATGCGCTGGGCAAAAGGCCGCAACCAGACCTACCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACACAAT----------CGCAC
TGGCTGCCACGCAGAGCTGCTGTTCCTGAGCCACCTA------GGGGCGCTGTGCCCGGGTCTG

TGGGGGTGCGGAGGTGACGAAAAC---CGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTC
CTGGTCTCCCTGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGACTCACG---------CC
CAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACGTGGAGGACAGTCCGC
ACAGGGAGGGCTTGAGGAACCTGAGGAGAGCAGGGGTCCTGGTCAACGTTATGAGCTACA
AAGACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTG
GGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------ACA---CTAAACCGCATCCTCC
AGCCATGTGAAACAGAAGATTTAAGAGACGCTTTCAGACTTATTGGGCTGTTAACC----------
-TGA

Merluccius_polli

ATGATTAGTAAGCTC---------GACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTAC
AAGAACATGCGCTGGGCAAAAGGCCGCAACCAGACCTACCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACACAAT----------CGCAC
AGGCTGCCACGCAGAGCTGCTGTTCCTGAGCCACCTA------GGGGCGCTGTGCCCGGGTCTG
TGGGGGTGCGGAGGTGACGAAAAC---CGAAGACTCAGCTACTCTGTCACCTGGTTCTGCTCC
TGGTCTCCCTGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGACTCACG---------CCC
AACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACGTGGAGGACAGTCCGCA
CAGGGAGGGCTTGAGGAACCTGAGGAGAGCAGGGGTCCTGGTCAACGTTATGAGCTACAA
AGACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGG
GAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------ACA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGACGCTTTCAGACTTATTGGGCTGTTAACC----------
TGA

Molva_molva

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTAGCCCAGAAGAAATTCATATACAACTAC
AAGAACATGCGATGGGCAAAAGGCCGCAATGAGACCTACCTCTGCTTCGTAGTGAAGAGA
AGGCTCGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGTAGAGCTGCTGTTTCTGAGCTACCTG------GGGGCGCTGTGCCCGGGCCTCT
GGGGGTGCGGAGGCGACACTAAC---CGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCT
GGTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCACACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCAT
ATAGAGGGCTTAAGGGACCTGAGGAGAGCAGGGGTCCAGGTCAAAGTTATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAAGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTATGTCCGTCTGTCAAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Monocentris_japonica

ATGATTACAAAACTA---------GACAGTGTGCTTTTGGCGCAGAAAAAGTTCATCTACCATTAT
AAGAACATGCGCTGGGCAAGGGGTCGGCATGAGACATACCTCTGCTTTGTAGTGAAGAGG
AGAGTGGGACCAGACTCCCTGTCCTTTGACTTTGGACACCTCCGCAAT----------CGCTC
TGGCTGCCATGTAGAGCTGCTGTTCCTGCGCCACCTG------GGAGCCTTGTGCCCTGGACTGT
GGGGGTATGGAGGCACTGGTGAG---AGGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCT
GGTCTCCCTGCGCTGACTGCTCCTTTAGATTGGTCCAGTTCCTCGGCCGGACG---------CCCAA
CCTCCGCCTCAGGATCTTCGTCTCTCGCCTCTACTTCTGTGACGTGGAGGACAGCCGCGAGA
GACAGGGCCTGAGAATGCTGAAAAAAGCCGGCGTGCAAATCACTGTCATGAGCTACAAAG
ACTACTTCTATTGCTGGCAGACCTTCGTGGCTCACAGACAGAGCAGTTTCAAGGCCTGGGA
TGAGCTGCACCAAAACTCTGTTCGCCTGGCCAGG---------AAA---CTCAACCGCATCCTCCAG
CCTTGTGAGACAGAAGATTTAAGAGATGCGTTCAAGCTTCTTGGGTTG----------TGA

Mora_moro

ATGATTAGTACACTA---------GACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTAC
AAGAACATGCGTTGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGTAGAGCTGCTGTTTCTGAGCCACCTG------GGGGCACTGTGCCCAGGCCTGT
GGGGGTGCGGAGGCGATGAAAAC---AGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCC
TGGTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACG---------CCC

AACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTATTTCTGTGACCTGGAGGACAGTCCGCA
TATAGAGGGCTTGAGGGACCTGAGGAGAGCAGGGGTGCAGGTCACTGTTATGAGCTACAA
AGACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGG
GAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------AAA---CTAAACCGCATCCTCC
AGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGGCTGTTAACC----------
-TGA

Muraenolepis_marmoratus

ATGATTAGCAAACTA---------GACAGTGTGCTCTTGGGCCAGAAGAAATTCATATACAATTAC
AAGAACATGCGTTGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTGGTGAAGAGA
AGGCTCGGACCCGATTCCATGTCTTTCGACTTCGGGCACCTACGCAAT----------CGCGC
AGGCTGCCACGTGGAGCTGCTGTTTCTCAGCCACCTG------GGGGCGCTGTGCCCGGGTCTGT
GGGGTTGCGGAGGCGACGAGAAC---AGACGGCTCAGCTACTCGGTCACCTGGTTCTGCTCCT
GGTCCCCCTGTGCCAACTGTGCCGCCACGCTGGCCCGGCTCCTGAGGCAGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGGCCCGCCTGTACTTCTGTGACCTGGAGGGCAGTCCGCAC
TCAGAGGGCCTGAGGGACCTGAGGAGGGCCGGGGTCCAGGTCAACGTTATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTTGTAGCGCACAGGGTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------AAA---CTAAACCGCATCCTCCA
GCCACGCGAAACAGACGATTTAAGAGATGCCTTCAGACTTATTGGTCTGTTAACC----------
TAA

Myoxocephalus_scorpius

ATGATTACAAAGCTA---------GACAGTGTGCTATTGCAGCAAAAAAAGTTCATCTACCATTAC
AAGAACATGCGCTGGGCAAGGGGCCGACATGAGACTTACCTCTGCTTTGTAGTGAAGAGGC
GAGTGGGGCCAGACTCCTTATCCTTTGACTTTGGACACCTCCGCAAT----------CGCACT
GGCTGCCATGTAGAGCTGTTGTTCCTACGCTACCTG------GGAGCCTTGTGCCCTGGTTTGTG
GGGTTACGGAGGCACTGGAGAG---AAGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG
GTCTCCCTGCATAAACTGCTCCATCAGTTTGTCCCAGTTCCTCAACCGGACG---------CCCAAC
CTTCGCCTCAGGATCTTTGTCTCTCGTCTTTACTTCTGTGACAAGGAGAACAGCCGGGAAAG
AGATGGCCTGAGAATGCTGAAAAATGCTGGCGTGCAGATCACAGTCATGAGTTACAAAGA
CTTCTTCTATTGCTGGCAGACATTTGTGGATCGCAAGAAAAGCAACTTCAAGGCCTGGGAA
GAGCTGCACCAGAACTCTGTTCGCCTTGCCAGA---------AAA---CTCAACCGCATCCTCCAGC
CTTGTGAAGCAGAAGATTTAAGGGATGCCTTCAAGCTTCTTGGACTG----------TGA

Myripristis_jacobus

ATGATTACAAAGCTA---------GACAGCATGCTTTTGGCCAAGAAAAAGTTCATTTACCATTAT
AAGAACATGCGCTGGGCTAAAGGTCGGCATGAGACATACCTGTGCTTTGTAGTGAAGAGAC
GAGTGGGGCCAGACTCCATGTCCTTTGACTTTGGACATCTCCGCAAT----------CGTGCT
GGCTGCCATGTAGAGCTGCTGTTCCTGCGCTACCTG------GGAGCGCTTTGCCCTGGACTGTG
GGGGTGTGGAGGCAACACTGAG---AAGAAGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG
GTCTCCCTGCGCCGACTGCTCTTTCAGACTGGCCCAGTTCCTCAACCGGACG---------CCCAA
CCTCCGCCTCAGGATCTTTGTCTCTCGCCTCTATTTCTGCGACCTGGAGGACAGCCGTGAGA
GAGAGGGCCTGAGGATGCTGAAAAAAGCCGGCGTGCAAATCACTGTTATGAGTTACAAAG
ATTACTTCTATTGCTGGCAGACATTTGTGGCACATAGAATGAGCAGCTTCAAGGCTTGGGAT
GGGCTGCACCAAAACTATGTTCGCCTGGCCAGG---------AAA---CTCAACCGCATCCTCCAGG
CTAGTGAGACAGAAGATTTAAGAGATGCATTCAAGCTTCTTGGATTG----------TGA

Neoniphon_sammara

ATGATTACAAAGCTA---------GACAGTGTGCTTTTGGCCAAGAAAAAGTTCATCTACCATTAT
AAGAACTTGCGCTGGGCAAAAGGCCGGCATGAGACATACCTCTGCTTTGTCGTGAAGAGGC
GGGTGGGGCCAGACTCCATTGCCTTCGACTTTGGACACCTCCGCAAT----------CGTGCT
GGCTGCCATGTAGAGCTGCTATTCCTTCGCTACCTG------GGAGCCTTGTGCCCTGGACTGTG
GGGGTATGGAGGAACTGGGGAG---AGGAAGCTGAGTTACTCCATCACGTGGTTCTGCTCCT
GGTCTCCCTGTGCCAACTGCTCCTTCAGACTCGCCCAGTTCCTCAACCGGACG---------CCCA
ACCTCCGCCTCAGGATCTTTGTCTCTCGCCTCTATTTCTGTGACGTGGAGGACAGCCGTGAG
AGAGAGGGCCTGAGAATGCTGAAAAATGCCGGCGTGCACATCACTGTTATGAGCTACAAA

GACTACTTCTACTGCTGGCAGACATTTGTGGCTCGTAAAACGAGCAGCTTCAAGGCTTGGG
ATGGGCTGCACCAAAACTATGTTCGCCTGGCCAGG---------AAA---CTCAACCGCATCCTCCA
GCCTTGTGACACAGAAGATTTAAGAGATGCATTCAGGCTTCTTGGATTG----------TGA

Oreochromis_niloticus

ATGATTGCAAAGCTA---------GACAGTATGCTTTTGCCCAGAAAAAAGTTCCTCTATCATTAC
AAGAATGTGCGCTGGGCGAGGGGCCGGAATGAAACATACCTCTGTTTTGTAGTAAAAAGAC
GAGTAGGGCCTGACTCCTTGTCCTTTGACTTTGGACACCTCCGCAAT----------CGCAAT
GGTTGCCACGTTGAGCTGCTGTTCCTGCGCCAACTT------GGTACATTATGCCCTGGCCTGTC
TGGGTATGGATTTCATGGGGAG---AGGAGGGTCAGCTACTCCATCACCTGGTTCTGCTCCTG
GTCTCCCTGTGCAAACTGCTCTTCCAGACTGGCCCAGTTCCTCAAACAGACA---------CCCAA
CCTTCGCCTCAGGATCTTTGTCTCTCGTCTTTACTTCTGTGACATGGAGGACAGTCGTGAAA
GAGAGGGTCTCAGGCTGCTTAAAAAAGTCGGCGTGCACATCACAGTCATGAGTTACAAAGA
CTTCTTCTACTGCTGGGAGAATTTTGTGGCC---CAGCAAAGCAAATTTAAGGCCTGGGAAGG
TCTGCATCAAAACACAGTACGCCTGGCCAGA---------AAA---CTCAACCGCATCCTCCAGCCC
TGTGACACAGAAGATTTAAGAGATGCCTTCAAGCTTCTTGGACTG----------TAA

Oryzias_latipes

ATGATTACTAAGCTC---------GACAGTGTGCTTCTTCCCAAAAAAAAGTTCATCTATCATTAC
AAGAACATGCGCTGGGCAAGAGGCAGACATGAAACATACCTCTGCTTTGTGGTGAAAAGG
AGAGTGGGCCCAGAATCTCTGTCCTTTGACTTCGGACATCTCCGCAAT----------CGTAA
TGGCTGCCATGTGGAGCTGTTGTTCCTGCGCCACCTG------AGTGCATTGTGCCCTGGTTTAT
GGGGATATGGAGCAACCGGACAG---GGAAGGGTCAGCTACTCCATCACCTGGTTTTGCTCTT
GGTCTCCTTGTGCTAACTGCTCTTTCAGACTAGCACAGTTTCTCAGCCAGACC---------CCCAA
CCTGCGCCTCAGGATTTTCGTCTCCCGTCTGTATTTCTGTGACTTGGAGGACAGCCGTGAAA
GAGAAGGTTTAAGGATGCTGAAAAAAGTTGGCGTGCACATCACAGTCATGAGTTACAAAG
ATTACTTCTATTGCTGGCAGACCTTTGTGGCTCGGAAACAAAGCAAATTCAAACCCTGGGA
TGGGCTGCACCAAAACTCTGTTCGGCTTTCAAGA---------AAA---CTCAACCGCATCCTACAG
CCATGTGAGACAGAAGATTTCAGAGATGCATTCAAGCTTCTTGGACTG----------TGA

Osmerus_eperlanus

ATGATCAGTACGCTA---------GACGGCGTGCTTCTGGCCCAGAAGAAGTTCATCTACCACTAC
AAGAACATGCGCTGGGCCAGAGGTCGACACGAGACCTACCTGTGCTTTGTGATCAAGAGGA
GGGTTGGGCCCGACTCGCTCTCCTTTGACTTCGGACACCTGCGCAAT----------CGCACC
GGCTGCCATGTAGAGCTGCTGTTTCTACGCCACCTG------GGGGCTCTGTGTCCCGGCCTGTG
GGGTACGGGTGGTGCCGGTGGTGGGGTGAGGTTGAGCTACTCCATCACCTGGTTCTGCTCC
TGGTCTCCCTGCTCCGCCTGCTCCCACAGGCTGTCTGACTTCCTCAGCCGGACC---------CCCA
ACCTCCGCCTCCGGATCTTCGTGTCTCGTCTCTACTTCTGCGACCCGGAAGACAGCCTGGAG
AGGGAGGGGCTCCGTATGCTGAAGAGAGCCGGAGTAAACATCACTGTCATGAGTTATAAA
GACTATTTCTACTGCTGGGAGACTTTTGTAGCTCGCAGAAAGACAGGCTTCAAGGCCTGGG
ACGGGCTTCACCACAACTCGGTTCGCCTGGCCAGG---------AAG---CTCTACCGTATCCTACA
GCCTTGTGAGACAGAAGATCTGAGAGATGCTTTCACGCTGCTGGGACTG----------TGA

Parablennius_parvicornis

ATGATTGCCAAGCTC---------GACAGTATGCTCCTGCCCAGAAAAAAGTTCATCTATCATTAC
AAGAACATGCGCTGGGCGAAGGGTCGGCATGAGACTTACCTCTGCTTCGTGGTGAAGCGGC
GACTGGGCCCAGACTCTTTGTCCTTTGACTTCGGGCATCTCCGAAAT----------CGCAAT
GGTTGCCATGTAGAGTTGCTGTTCCTGCGCCACCTG------GGGACTTTGTGCCCTGGTCTGTC
GGGGTACGGAGTACATGGAGAA---AAAAGGCTTAGCTACTCCATCACCTGGTTCTGCTCCTG
GTCTCCCTGTTCCAACTGTTCTCACCGACTAGCCCAGTTCCTGAGCCGAACG---------CCCAA
CATTCGACTCAGAATCTTTGTCTCCCGCCTGTACTTCTGCGACTTGGAGGACAGCCGCGAGA
GAGAGGGTCTCCGGCTGCTGAAAAAAACTGGCGTGCATATCACGGTCATGAGCTACAAAG
ATTATTTCTATTGCTGGCAAACTTTTGTGGCAAGTAATCAGAGCAGGTTTAAGCCTTGGGAT
GAGCTGCAGCGAAACTCCATCCGCCTCACCAGA---------AAA---CTCAACCGCATCCTCCAGC
CCTGCGAAACAGAAGATTTAAGAGATGCCTTCAAGCTTCTTGGACTC----------TGA

Parasudis_fraserbrunneri

ATGATTACTAATCTA---------GACAGTGTGCTTCTGGCCCAGAAGAAGTTCATCTACCATTAC

AAGAACATGCGGTGGGCAAGGGGCCGGCATGAGACTTATCTCTGCTTTGTAGTGAAGAGAA
GGTTGGGTCCAGACTCCTTGTCCTTCGACTTTGGACACCTTCGCAAT----------CGCTCT
GGCTGCCATGTAGAGCTGCTTTTCCTGCGTCACCTG------GGCGCCCTTTGCCCTGGCCTGTG
GGGATATGGAGGAGAG---------AAGAGGCTGAGCTACTCTGTCACCTGGTTCTGCTCCTGGTC
GCCCTGCGCCGACTGCTCCACCAGACTGTCCCAGTTCCTCAGCAGGACG---------CCCAACCT
CCGCCTGAGGATCTTCGTCTCGCGCCTCTACTTCTGCGACCTGGAGGACAGCCTCGCAAGA
GAGGGCCTGAGGACACTGAAGAGAGTCGGCGTGCAGGTCACTGTCATGAGCTACAAAGAC
TACTTCTACTGCTGGCAGACCTTCGTGGCTCGCAGACAGAGCAGCTTCAAGGCTTGGGATG
GGCTGCAGCAGAACTCTGTCCGCCTGGCCAGG---------AAA---CTCAACCGCATCCTCCAGCC
TTGTGAGACAGAAGACTTACGAGATGCATTCAAGCTTCTTGGACTG----------TGA

Perca_fluviatilis

ATGATTACAAAGCTA---------GACAGTGTGCTTTTGCCCCGAAAAAAGTTCATCTACCATTAC
AAGAACATGCGCTGGGCAAGGGGTCGCCATGAGACATATCTCTGCTTTGTAGTGAAGAGGC
GAGTGGGGCCAGACTCCTTATCCTTTGACTTTGGACACCTCCGCAAT----------CGCAAT
GGCTGCCATGTAGAGCTGCTGTTCCTGCGCTACATT------GGAGCCTTGTGCCCTGGTTTGTG
GGGATGCAGCGGTACTGGAGAG---AGGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCT
GGTCTCCTTGTGCCAACTGCTCCATCAGACTGTCCCAGTTCCTCAGCCAGACG---------CCCA
ACCTTCGCCTAAGGATTTTCGTCTCTCGCCTTTACTTCTGTGACACGGAGAACAGCCCTGAA
AGAGACGGCCTAAGAATGCTGAAAAAAGCTGGCGTGCAGATCACAGTCATGAGTTACAAA
GACTTCTTTTATTGCTGGCAGACCTTTGTGGATCGTAAGCAAAGCAACTTCAAGGCCTGGGA
AGAGCTGCACTCAAACTCTGTTCGCCTTTCCAGA---------AAA---CTCAACCGCATCCTCCAGC
CTTTTGAAACAGAAGATTTAAGAGATGCCTTCAAGCTTCTTGGACTG----------TGA

Percopsis_transmontana

ATGATTACCAAGCTA---------GACAGTGTGCTTCTGGCGCAGAAGAAATTCATCTTCCACTAC
AAGAACATGCGCTGGGCAAGGGGTCGCCATGAGACATATCTCTGCTTTGTCATTAAGAGGA
GAGTGGGGCCAAACTCCCTGTCCTTTGACTTTGGACACCTCCGCAAT----------CGCTCC
GGTTGCCATGTAGAGATCCTGTTCCTGCGCCACTTG------GGAGCGCTGTGCCCTGGACTGTG
GGGAGAGGGGGGTACTGGTGAG---AGAAGATTAAGTTACTCCATCACCTGGTTCTGCTCCTG
GTCTCCCTGTGCCAACTGCTCCCTCAGACTGGCCCAGATCCTCAGACAGCTG---------CCCAA
CCTCCGCCTGAGGATCTTTGTGTCCCGCCTCTACTTCTGTGACCTGGAGGACAGCAAAGAGA
GAGATGGCCTCAGAATGCTGAAGAACGTGGGTGTGCAGATCACCGTCATGAGCTACAAAG
ACTATTTCTATTGCTGGCAGACCTTTGTAGCTCACAGAAAGAGTAACTTCAAAGCCTGGGA
CGGGCTGCACCAAAACTCTGTTCGCCTGGCTCGG---------AAA---CTCAACCGCATCCTCCAG
CCTTGTGAGATAGAAGATTTAAGAGATGCCTTCAAACTTCTTGGGTTT----------TGA

Phycis_blennoides

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTAGCCCAGAAGAAATTCATATACAATTAC
AAGAACATACGATGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTTGTAGTGAAGAGA
AGGCTCGGACCCAATTCCCTGTCCTTCGACTTCGGTCACCTACGCAAT----------CGCGC
TGGCTGCCACGTAGAGCTGCTGTTTCTGAGCCACCTG------GGGGCGCTGTGCCCGGGCCTCT
GGGGGTGCGTGGATGACAGCAAC---AGGAGACTGAGCTACTCGGTCACCTGGTTCTGCTCCT
GGTCTCCCTGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTACGGATGACA---------CCCA
ACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCGCAT
ATTGAGGGCTTGAGGCACCTGAGGAGAGCAGGGGTTGAGGTCAAAGTTATGAGCTACAAA
GACTACTTCTACTGTTGGCAGACCTTCGTAGCTCACAGGCTGAGTCGCTTCAAGGCCTGGGA
AGGGCTGCATACCAATTCTGTCCGTCTGTCAAGA---------AAA---CTAAACCGCATCCTCCAG
CCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Phycis_phycis

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTAGCCCAGAAGAAATTCCTATACAATTAC
AAGAACATACGATGGGCAAAAGGCCGCAACGAGACCTTCCTCTGCTTTGTAGTGAAGAGA
AGGCTCGGACCCAATTCCTTGTCCTTCGACTTCGGTCACCTACGCAAT----------CGCGC
TGGCTGCCACGTAGAGCTGCTGTTTCTGAGCCACCTG------GGGGCGCTGTGCCCGGGCCTCT
GGGGGTGCGTAGATGACAGCAAC---AGGAGACTGAGCTACTCGGTCACCTGGTTCTGCTCCT

GGTCTCCATGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTCAGGATGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACCTGGAAGACAGTCCGCAT
ATTGAGGGCTTGAGGCACTTGAGGAGAGCGGGGGTCGAGGTCAAAGTTATGAGCTATAAA
GACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGTCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTCTGTCCGTCTGTCAAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Poecilia_formosa

ATGATTACAAAGCTA---------GACAGGGCACTATTACCCAGAAAAAAATTCATCTATCATTAC
AAGAACTTGCGCTGGGCAAGAGGTCGATGTGAGACGTACCTCTGTTTTGTGGTGAAGAAGC
GAGTGGGACCAGACTCCCTGTCCTTTGACTTTGGGCATCTCCGCAAC----------CGCAAC
AACTGCCATGTGGAGCTGCTGTTCCTGCGCCACCTG------GGAGCGTTGTGCCCTGGCCTGTG
GGGTTATGGAGTCACTGGTGAA---AGAAAAGTCAGCTACTCTGTCACCTGGTTTTGCTCCTG
GTCTCCCTGTGCAAACTGCTCCATCCGACTGGCTCAGTTCCTCCACCAGACC---------CCCAA
CCTCCGCCTCAGGATCTTTGTATCCCGGCTTTATTTCTGCGACTTGGAGGACAGCCGTGAAA
GAGAGGGACTTAGAATACTGAAAAAAGCTGGCGTGCACATCACAGTCATGAGTTACAAAG
ATTACTTTTACTGCTGGCAGACCTTTGTGGCAAAAAGCCAAAGCAAGTTCAAGCCGTGGGA
TGGGCTGCACCAAAACTATATCCGGCTGTCAAGG---------AAA---CTCAACCGCATTCTTCAG
CCATGTGAGACAGAAGATTTAAGAGATGCCTTCAGGCTTCTTGGACTG----------TGA

Pollachius_virens

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTGGCCCAGAAGAAATTCATCTACAATTAC
AAGAACATGCGATGGGCAAAAGGCCGCAACGAGACCTATCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCTGATTCCCTCTCTTTCGACTTCGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGCAGAGCTGCTGTTCCTGAGCTACCTG------GGGGCGCTGTGCCCAGGCCTCT
GGGGCTGCGCAGACGACAGAAAC---CGAAGACTAATTTACTCCGTCACCTGGTTCTGCTCCT
GGTCGCCCTGTGCCAACTGTGCGACCACGCTGGCCCGGTTCCTGAGGCAGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGTCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCAT
GTAGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTATGTGCGTCTGTCAAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGTTTTCGGACTTTTTGGACTGTTAACT----------
TGA

Polymixia_japonica

ATGATTACTAAACTA---------GACAGTGTGCTTTTGGCCCAGAAGAAATTCATCTACCATTAT
AAGAACATGCGCTGGGCGAAGGGTCGACACGAGACGTATCTCTGCTTTGTAGTCAAGAGGA
GGGTGGGACCGGACTCCATGTCCTTTGATTTTGGACACCTACGCAAT----------CGCTCT
GGCTGCCATGTAGAGCTGCTGTTCCTGCGCCACCTG------GGGGCCTTGTGCCCTGGACTGTG
GGGATACGGAGGTACTGGTGAG---AAGAGGCTCAGTTACTCCGTCACCTGGTTCTGCTCCTG
GTCGCCCTGCTCCAACTGCTCCTACAGACTGGCCCAGTTCCTCAGCCAGACG---------CCCAA
CCTCCGCCTCAGGATCTTCGTCTCTCGACTTTACTTCTGCGACCTGGAGGACAGCCGGGAGC
GAGACGGCCTCAGAATGCTCAAAAGGGCTGGAGTGCAAATCACAGTCATGACCTACAAAG
ACTACTTCTATTGCTGGCAGACCTTTGTGGCTCACAGAACAAGCAAGTTCAAGGCCTGGGA
TGAGCTGCACCGGAACTCTGTCCGCCTGTCCAGG---------ATA---CTCAACCGCATCCTCCAGC
CTTGTGAGACAGAAGATTTAAGAGATGCCTTCAGACTTCTTGGGTTG----------TGA

Pseudochromis_fuscus

ATGATTGCAAAGCTT---------GACAGTGTGCTTTTGCCAAAAAAGAAATTCATCTTTCATTAC
AAGAACATGCGCTGGGCAAGGGGCCGACATGAGACATACCTCTGCTTTGTGGTGAAAAGG
CGAAGGGGCCCAGACTCTCTGTCCTTTGACTTTGGACATCTCCGCAAT----------CGCAA
CGGCTGCCATGTAGAGCTGCTATTCCTACGGTACCTG------GGAGCCTTGTGCCCTGGTCTGT
GGGGGTATGGGGCTACTGGGGCG---AGCAGGCTCAGCTACTCCATCACGTGGTTCTGCTCCT
GGTCTCCTTGTGCCAACTGCTCTTTCAGACTGGCCCAGTTCCTCAGCCAGACG---------CCCA
ATCTTCGCCTCAGGATCTTCGTCTCTCGCCTTTACTTTTGTGACATGGAGGACAGCCGTGAA
AGAGAGGGTCTAAGGCAGCTGAAAAAAGCCGGAGTGCACATCACAGTCATGAGTTACAAA

GACTACTTCTACTGCTGGCAGACCTTTGTGGCTCGTAATCAAAGCAAATTCAAGCCCTGGG
ATGAATTGCACCAAAACTCTGTCCGCCTGTCCAGA---------AAA---CTCAACCGCATCCTCCA
GCCTTGTGAGACAGAAGATTTAAGAGATGCCTTCAAGCTTCTTGGACTG----------TGA

Rondeletia_loricata

ATGATTACAAAACTA---------GACAGTGTGCTTTTGGCCAAGAAAAAGTTCATCTACCATTAT
AAGAACATGCGCTGGGCAAGGGGTCGGCATGAGACATACCTCTGCTTTGTAGTGAAGAGGC
GAGTGGGGCCAGACTCCCTGTCCTTCGACTTTGGACACCTCCGCAAC----------CGCACT
GGCTGCCATGTAGAGCTGCTGTTCCTGCGCCACCTG------GGAGCCTTGTGCCCTGGACTGTG
GGGGCATGGAGGCACTGGAGAG---AGGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCTG
GTCTCCCTGCGCTGACTGCTCCTTCAGACTGGCCCAGTTCCTCGGCCGGATG---------CCCAA
CCTCCGCCTCAGGATCTTCGTCTCTCGCCTCTACTTCTGCGACCTGGAGGACAGCCGCGAGA
GAGAGGGCCTGAGGTTGCTGAAAAAAGCCGGCGTGCAGATCACTGTCATGAGCTACAAAG
ACTTCTTCTATTGCTGGCAGACCTTTGTGGCTCATAGAAATTGCAGCTTCAAGGCCTGGGAT
GAGATGCATCAAAACTCTGTTCGCCTGGCCAGG---------AAA---CTCAACCGCATCCTGCAGC
CTTGTGAGACAGAAGATTTAAGAGATGCGTTCAAGCTTCTTGGGTTG----------TGA

Salmo_salar_1

ATGATCAACAAATTT---------GACAGTGTTCTGTTGGCCCAGAAGAAGTTTATCTACCACTAT
AAGAACATGCGCTGGGCCAAGGGCCGACACGAAACCTACCTGTGCTTCGTGGTCAAGAGG
CGGGTGGGACCAAACTCACTCTCCTTCGACTTTGGACACCTGCGCAAC----------CGGTC
CGGCTGTCATGTTGAGCTGCTGTTCCTGCGCCTCTTGGAAGCAGGCGCCCTGTGTCCAGGCC
TGTGGGGTTATGGAGCTCCAGACAGT---GTGGGACTGTGTTACTCTGTCACCTGGTTCTGTTC
CTGGTCCCCCTGCTCAGACTGCTCCTACAGGCTGGCCCAGTTCCTCAGCCAGACC---------CC
CAACCTCCGCCTCAGGATCTACGTCTCCAGGCTCTACTTCTGTGACCCGGAGGACAGCAGT
GCTAGAGAGGGTCTCCGCATGCTGCAGAGAGCCGGGGTGCAGATCACTGTCATGAACTATG
AAGACTATTTCTACTGTTGGCAGACCTTTGTGGCTTGCAGACAGCGTGTTTTTAAGGCCTGG
GATGGACTGCATCAGAACTCTGTTCAACTGGCTAGG---------AAA---CTTAACGACATCCTCC
AGCCTGGAGAGGCAGAAGATTGGGGAGATGCTTTCGAGCTACTTGGACTG----------
TGA

Salmo_salar_2

ATGATCAACAAATTT---------GACAGTGTTCTGTTGGCCCAGAAGAAGTTTATCTACCACTAT
AAGAACATGCGCTGGGCCAAGGGCCGACACGAAACCTACCTGTGCTTCGTGGTCAAGAGG
CGGGGGGGACCAAACTCACTCTCCTTCGACTTTGGACACCTGCGCAAC----------CGGT
CCGGCTGTCATGTTGAGTTGCTGTTCCTGCGCCTCCTGGAAGCAGGCGCCCTGTGTCCAGGC
CTGTGGGGTTATGGAGCTCCAGACAGT---GTGGGACTGTGTTACTCTGTCACCTGGTTCTGTT
CCTGGTCCCCCTGCTCAGACTGCTCCTACAGGCTGGCCCAGTTCCTCAGCCAGACC---------C
CCAACCTCCGCCTCAGGATCTACGTCTCCAGGCTCTACTTCTGTGACCCGGAGGACAGCAGT
GCTAGAGAGGGTCTCCGCATGCTGCAGAGAGCCGGGGTGCAGATCACTGTCATGAACTATG
AAGACTATTTCTACTGTTGGCAGACTTTTGTAGCTTGCAGACAGCGTGTGTTTAAGGCCTGG
GACGGACTGCATCAAAACTCTGTTCAACTGGCCAGG---------AAA---CTTAACGACATCCTCC
AGCCTGGTGAGGCAGAAGATTGGGGAGATGCTTTCGAGCTACTTGGACTG----------
TGA

Sebastes_norvegicus

ATGATTACAAAGCTA---------GACAGTGTGCTTTTGCCTCGAAAAAAGTTCATCTTCCATTAC
AAGAACATGCGCTGGGCAAGAGGCCGGCATGAGACATACCTCTGCTTCGTAGTGAAGAGG
CGAGTGGGGCCAGACTCCTTAACCTTTGACTTTGGACACCTCCGCAAT----------CGCAA
TGGCTGCCATGTAGAGCTGCTGTTCATGCGCTACCTG------GGAGCCTTGTGCCCTGGTTTGT
GGGGGCAGGGAGTCCCCGGAGAG---AAGAGGCTCAGTTACTCCATCACCTGGTTTTGCTCCT
GGTCTCCCTGCGTCAACTGCTCCGTCACACTGTCCCAGTTCCTCAGCAAAACG---------CCCA
ACCTTCGCCTCAGGATCTTCGTCTCTCGCCTTTACTTCTGTGACATGGAGAACAGCCGTGAA
AGAGATGGACTAAGAATGCTGAAAAAAGCTGGCGTGCAGATCTCAGTCATGAGTTACAAA
GACTACTTCTATTGCTGGCAGACCTTTGTGGATCGGAAGCAGAGCAAGTTCAAGGCCTGGG
ATGAGATGCACCAAAACTCTGTTCGCCTTACCAGA---------AAA---CTCAGCCGCATCCTCCA
GCCTAGTGAAACAGAAGATTTAAGGGATGCCTTCAAGCTTCTTGGACTG----------TGA

Selene_dorsalis
ATGATTACTAAGCTA---------GACAGTGTGCTTTTGCCCCGCAAAAAGTTTATCTTCCATTAC
AAGAATGTGCGCTGGGCGAAGGGCCGGCATGAGACATACCTCTGCTTTGTTGTGAAGAGAC
GAGTGGGCCCAGACTCCATGACTTTTGACTTTGGACACCTCCGCAAT----------CGTAAT
GGCTGCCATGTAGAGATTCTGTTCCTGCGTTACCTT------GGTGCCTTGTGTCCTGGTCTATG
GGGTTATGGGGTTGGTGGAGAG---AAGAGACTCAGTTACTCCATCACCTGGTTCTGCTCCTG
GTCTCCCTGTGCCAACTGCTCCAGCAGGCTGGCCCAGTTCTTAAAGCAGACG---------CCCAA
CCTTCGCCTAAGGATCTTCGTTTCACGCCTTTATTTCTGTGACTTGGAGGACAGCCAAGAGA
GAGAGGGCCTGAGGATATTGAAAAAAGCTGGAGTGCACATAACAGTCATGACTTACAAAG
ACTTCTTCTATTGCTGGCAGACCTTTGTGGCTCGTAAACAGAGTAGCTTCAAAGCCTGGGAT
GAGCTGCACCAAAATTCTGTTCGTCTTGCTAGA---------AAA---CTTCAGCGTATCCTCCAGCC
ATGTGAAACAGAAGATTTGAGGGATGCCTTCAAACTTCTTGGACTG----------TGA
Spondyliosoma_cantharus
ATGATTACAAAGCTA---------GACAGTGTGCTTTTGCCTAAAAAAAAATTCATCTACCACTAC
AAGAATGTGCGCTGGGCAAGGGGCCGACATGAGACTTACCTGTGCTTTGTAGTGAAGAGGC
GAGTGGGGCCAGACACCTTAACCTTTGACTTCGGACACCTCCGCAAT----------CGCAA
CGGCATCCATGTCGAGTTGCTGTTCCTGCGCTATCTG------GGAGCCTTGTGCCCTGGTTTGT
GGGGGTATGGAGGCACTGGAGAG---AAGAGGCTGAGTTACTCTATCACCTGGTTCTGCTCCT
GGTCTCCCTGTGCCAACTGCTCACTCAGACTGTGCCAGTTCCTCAGCCAGACT---------CCCA
ACCTTCGCCTTAGGATCTTCGTCTCTCGCCTCTACTTCTGTGACATGGAGGACAGCCGTGAA
AGAGAGGGCCTAAGAATGCTGAAAAAAGCCGGCGTGCAGATCACAGTCATGAGTTACAAA
GACTTCTTCTATTGCTGGCAGACATTTGTGGCTCGTAGGGCGAGCCAATTCAAGGCCTGGG
AAGAGCTGCAACGTAACTCTGTTCGCCTTACCAGA---------AAA---CTGAACCGCATCCTCCA
GCCCTGTGAAACAGAAGATTTAAGAGATGCCTTTAAGCTTCTTGGACTG----------TGA
Stylephorus_chordatus
ATGATTGCAAAACTA---------GACAGTGTGCTTCTGGCCCGGAATAAATTCATCTACCATTAT
AAGAACATGCGCTGGGCGAAAGGGCGCAACGAGACCTACCTCTGCTTTGTAGTGAAGAGA
AGGGTTGGACCTGATTCCCTGGCTTTCGACTTTGGACACCTCCGCAAT----------CGTAC
CGGTTGCCACGTAGAGCTCCTGTTCCTGCGCCACCTG------GGGGCCCTGTGCCCTGGACTGT
GGGGA---GGTGCTGCTGGTGAT---AAAAGGCTCAGCTATTCAGTCACCTGGTTCTGCTCTTGG
TCTCCCTGTGCCAACTGTGCTTCCACGCTGGCCCAATTCCTGAGACAGACG---------CCAAAC
CTCCGTCTCAGGCTCTTTGTGGCTCGTCTCTACTTCTGTGACCTGGAGGATAGTCCTGACAG
AGAGGGCCTACGGATTTTGAGAAGAGCTGGGGTGCATATCACAGTTATGAGATACAAAGA
CTACTTCTACTGCTGGCAGACCTTTGTGGCTCACAACCAGAGCCGCTTCAAAGCCTGGGAA
GGACTGCACCCAAACTCTGTCCGTTTGTCCAGA---------ACA---TTAAACCGCATCCTCCAGC
CTTGTGAAACAGAAGATTTAAGAGATGCTTTCAAACTCCTTGGATTG----------TAA
Symphodus_melops
ATGAATACAAAACTC---------GACAGCGTGCTTTTGCCACGAAAGAAGTTCATTTACCATTAC
AAGAACGTGCGCTGGGCAAGGGGCCGGCATGAGACATACCTTTGCTTTGTAATCAAGAGAC
GGGTGGGGCCGGACACCTTAACCTTTGATTTTGGACACCTCCGCAAT----------CGCAAT
GGCTGCCATGTAGAGCTGCTGTTCCTGCGCTACCTG------GGGGCCTTGTGTCCTGGTTTATT
GGGGTATGGAGGCGCCGGAGAG---AAGAGGCTCAGCTACTCTATCACCTGGTTCTGCTCCTG
GTCTCCATGCTCCAACTGCTCCACAATACTTTGCCAGTTCCTCAGTAAGATG---------CCCAA
CCTTCGCCTCCGGCTCTTCGTCTCTCGCCTTTACTTCTGTGACATGGAGGATAGTCGTGAAA
GAGAGGGCTTAAGAATGCTGAAAAAAGTCGGGGTGCAGATCACAATCATGAGTTACAAAG
ATTTCTTCTATTGTTGGCAGAAATTTGTGGCACGTAGGCAAAGCAACTTCAAGGCATGGGA
AGAGCTGCACCAGAACTCTGTTCGTCTTTCCAGG---------AAA---CTCAACCGCATCCTACAG
CCCTGTGAAACAGAAGACTTGAGAGATGCGTTCAAGCTTCTTGGACTT----------TGA
Takifugu_rubripes
ATGATCACCAAGCTA---------GACAGTATGCTTTTGCCAAGAAAGAAGTTCATCTACCATTAC
AAGAATGTGCGATGGGCACGGGGCCGGCACGAGACCTACCTCTGCTTTGTGGTGAAGAGGC
GAGTGGGCCCAGACACGCTAACCTTTGACTTTGGACACCTCCGCAAT----------CGCAG
TGGCTGCCACGTAGAGCTGCTTTTCCTGCGTTACCTG------GGAGCCTTATGCCCTGGGTTGT

GGGGTTACGGAGCCGCCGGAGAG---AAGAGGCTCAGTTACTCCGTTACCTGGTTCTGCTCCT
GGTCTCCCTGTGTCAACTGCTCCATCCAACTCTGCCAGTTTCTCAATAACACC---------CCCAA
CCTTCGCCTCAGGATCTTTGTCTCTCGCCTTTACTTTTGTGACCTGGAGGACAGCCTTGAAA
GAGAAGGCCTGAGGATGCTGACCAAAGCCGGCGTGAGGATCTCAGTGATGAGCTACAAAG
ACTATTTCTATTGCTGGCAGAAATTTGTGGATTGCAAAAAGAGCAACTTCAAGGCCTGGGA
AGAGCTGCATCAAAACTCTGTACGCCTCACCCGC---------AAA---CTCAATCGCATTCTGCAG
GCCTGGGATTTAGAAGACTTACGAGATGCTCTTAAGCTCCTCGGATTC----------TGA

Tetraodon_nigroviridis

ATGATTACCAAGCTA---------GACAGTATGCTTTTGCCAAGAAAAAAGTTCCTCTACCATTAC
AAGAACGTGCGATGGGCGCGGGGCCGACACGAGACCTACCTCTGCTTTGTTGTGAAGCGGA
GAGTGGGCCCAGACACGCTAACCTTTGACTTCGGGCACCTCCGCAAT----------CGCAA
CGGTTGCCACGTAGAGCTGCTCTTCCTGCGCTACCTG------GGGGCCCTGTGCCCGGGTTTGT
GGGGGTTATGGCGCTGCCGGGGAG---AAGAGGCTCAGTTACTCCATCACCTGGTTCTGCTCCT
GGTCCCCCTGCGCCAACTGCTCCATCCAACTTTCCCAGTTTCTGAGGAACACG---------CCCA
ACCTTCGCCTCAGAATCTTTGTCTCCCGCCTTTACTTCTGTGACATGGAGGACAGCCTTGAA
CGGGAAGGCCTGAGGATGCTGTCCAGGGCCGGCGTGAGGATTTCAGTGATGAGCTACAAA
GACTTTTTCTATTGCTGGCAGAAATTTGTGGATAGCAAAACGAGCAGCTTTAAAGCCTGGG
AAGAGCTGCACCAGAACTCTGTACGCCTCACTCGA---------AAA---CTCAACCGCATTCTCCA
GAGCTGGGATTTAGAAGATTTACGAGACGCCCTTAAGCTTCTTGGACTC----------TAA

Theragra_chalcogramma

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTGGCCCAGAAAAAATTCATCTACAATTAC
AAGAACATGCGATGGGCAAAAGGCCGCAACGAGACCTATCTCTGCTTCGTAGTGAAGAGA
AGGCTTGGACCTGATTCCCTCTCTTTCGACTTCGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGCAGAGCTGCTGTTCCTGAGCTACCTG------GGGGCGCTGTGCCCGGGCCTCT
GGGGCTGCGCAGACGACAGAAAC---CGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT
GGTCGCCCTGTGCCAACTGTGCGACCACGCTGACCCGGTTCCTGAGGCAGACA---------CCCA
ACCTGCGACTCAGGATCTTCGTGTCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCAT
GTAGAGGGCTTGAGGGACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTGATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTTGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTATGTGCGTCTGTCAAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGTTTTCAGACTTTTTGGACTGTTAACC----------
TGA

Thunnus_albacares

ATGATTACAAAACTA---------GACAGTGTGCTTTTGCCCCGGAAAAAGTTCATCTACCATTAC
AAGAACGTGCGCTGGGCAAGAGGACGGCATGAAACATACCTCTGCTTTGTAGTGAAGAGG
CGAGTGGGGCCAGACTCTTTATCCTTTGACTTTGGACACCTGCGCAAT----------CGCAA
TGGCTGCCATGTAGAGCTGCTGTTCCTGCGATATCTG------GGAGCCTTGTGCCCTGGTGTGT
GGGGGTATGGAAATACTGGACAG------AGGATCAGTTACTCCATCACCTGGTTCTGCTCTTG
GTCTCCCTGTGCCAACTGCTCTCGCAGACTGGCCCAGTTCCTCAGCCAGGTA---------CCCAA
CGTTCGCCTTAGGATCTTCGTATCACGCCTCTACTTTTGTGACTTGGAGGACAGCCGTGAGA
GAGACGGCCTGAGGTTGCTAAAAAACGCCGGCGTGCAGATCACAGTCATGAGTTACAAAG
ACTTCTTCTACTGCTGGCAGACTTTTGTAGCTCGTAATCAGAGCAAATTCAAGGCCTGGGAA
GAGCTGCACCGAAACTCTGTTCGCCTAACAAGA---------ACA---CTCAACCGCATACTCCAGC
CCTGTGACATTGATGATTTAAGAGATGCCTTCAAGCTTCTTGGGCTG----------TGA

Trachyrincus_murrayi

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTGGCCCAGAAGAAATTCATCTACAATTAC
AAGAACATGCGTTGGGCAAAAGGCCGCAACGAGACCTACCTATGCTTTGTGGTGAAGAGA
AGGCTTGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAAT----------CGCCT
TGGCTGCCACGTAGAGCTGCTGTTTCTGAGCCACCTG------GGGGCGCTGTGCCCGGGCCTGT
GGGGGTGTGGAGGCGACGTAAAC---AGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCT
GGTCTCCCTGCGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCGCAT
ATAGAGGGCTTGAGGGATCTGAGGAGAGCAGGGGTCCAGGTCACCGTTATGAGCTACAAA

GACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGACTGTTAACC----------
TGA

Trachyrincus_scabrus

ATGATAAGTAAGCTA---------GACAGTGTGCTCTTGGCTCAGAAGAAATTCATCTACAATTAC
AAGAACATGCGTTGGGCAAAAGGCCGGAATGAGACCTACCTATGCTTTGTGGTGAAGAGA
AGGCTTGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAAT----------CGCCT
TGGCTGCCACGTAGAGCTGCTGTTTCTGAGCCACCTG------GGGGCACTGTGCCCGGGCCTG
TGGGGGTGCGGAGGCGACGAAAAC---AGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTC
CTGGTCTCCCTGCGCCAACTGTGCGGCCACACTGGCCCGGTTCCTGAGGCACACG---------CC
CAACCTGCGCCTCAGGATCTTCGTGGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCGC
ATATAGAGGGCTTGAGGGATCTGAGGAGAGCAGGGGTCCAGGTCACTGTTATGAGCTACA
AAGACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTG
GGAAGGGCTGCATACCAATTCTGTCCGTCTGTCCAGA---------AAA---CTAAACCGCATCCTC
CAGCCATGTGAAACAGAAGATTTAAGAGATGCTTTCAGACTTATTGGACTGTTAACC---------
---TGA

Trisopterus_minutus

ATGATTAGTAAGCTA---------GACAGTGTGCTCTTGGCCCAGAAGAAATTTATATACAATTAC
AAGAACCTACGATGGGCAAAAGGACGCAACGAGACCTACCTCTGCTACGTAGTGAAGAGG
AGGCTCGGACCTGATTCCCTCTCCTTCGACTTCGGACACCTACGCAAT----------CGCAC
TGGCTGCCACGTAGAGCTGCTGTTCCTCAGCTACCTT------GGGGCACTATGCCCGGGCCTCT
GGGGCTGCACCGATGACAGAAAC---CGAAGACTGAGCTACTCCGTCACCTGGTTCTGCTCCT
GGTCTCCCTGTGCCAACTGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACG---------CCCA
ACCTGCGCCTCAGGATCTTCGTCGCCCGCCTCTACTTCTGCGACCTGGAGGGCAGTCCGCAC
ATAGAGGGCTTGAGGCACCTGAGGAGGGCAGGGGTCCAGGTCAAAGTCATGAGCTACAAA
GACTACTTCTACTGCTGGCAGACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGG
AAGGGCTGCATACCAATTCTGTGCGTCTGTCAAGA---------AAA---CTAAACCGCATCCTCCA
GCCATGTGAAACAGAAGATTTAAGAGATGTTTTCGGACTTTTTGGACTGTTAACC----------
TGA

Typhlichthys_subterraneus

ATGATTAGCAAGCTA---------GACAGTGTCCTTCTGGCGCAGAAGAAATTCATCTTCCACTAT
AAGAATATGCGCTGGGCAAGGGGGCGCAATGAGACATATCTCTGCTTTGTCATTAAGAGGA
GAGTGGGGCCGGACTCCCTGTCCTTTGACTTTGGACACCTCCGCAAT----------CGCTCC
GGCTGCCATGTAGAGCTGCTGTTCCTGCGCCACTTG------GGGGCGCTGTGCCCTGGCCTGTG
GGGACAGGGGGGTACAGGTGAC---AACAGACTCAGTTACTCCATCACCTGGTTCTGCTCCTG
GTCCCCCTGTTCCAACTGCTCTCACAGACTGGCCCAGTTCCTCAGCCAGCTG---------CCCAA
CCTCCGCCTGAGGATCTTTGTGTCCCGTCTGTACTTCTGTGACCTGGAGGACAGCAGGGAGA
GAGAGGGCCTGAGAATGCTGAAGAATGCGGGCGTGCACATAACCGTCATGAGCTACAAAG
ACTATTACTATTGCTGGCAGACCTTTGTAGCTCGCAGAAAGAGTAAATTCAAAGCATGGGA
AGGGCTGCACCAAAACTCTGTTCGCCTGGCCAGG---------AAA---CTCAACCGCATCCTCCAG
CCGTGCGAGATAGAAGATTTAAGAGATGCCTTCAAACTTCTTGGGTTT----------TGA

Xiphophorus_maculatus

ATGATTACAAAGCTA---------GACAGGGTACTATTACCCAAAAAAAAATTCATCTATCATTAC
AAGAACATGCGCTGGGCAAGAGGTCGATGTGAGACATACCTCTGCTTTGTGGTGAAGAAGC
GAGTGGGACCAGACTCCCTGTCCTTTGACTTTGGACATCTCCGCAAC----------CGCAAC
AACTGTCATGTGGAGCTGCTGTTCCTGCGCCACCTG------GGAGCGTTGTGCCCTGGCCTGTG
GGGTTATGGAGTCACTGGTGAG---AGAAAAGTCAGCTACTCCATCACCTGGTTTTGCTCCTG
GTCTCCCTGTGCAAACTGCTCCTTCCGACTGGCTCAGTTCCTCCACCAGACC---------CCCAA
CCTCCGCCTCAGGATCTTTGTATCCCGGCTTTATTTCTGTGACTTGGAGGACAGCCGTGAAA
GAGAGGGACTTAGAATGCTGAAAAAAGCTGGCGTGCACATCACAGTCATGAGTTACAAAG
ATTACTTTTACTGCTGGCAGACCTTTGTGGCAAAAAGTCAAAGCAAGTTCAAGCCGTGGGA
TGGGCTGCACCAAAACTGTATCCGGCTGACAAGG---------AAA---CTCAACCGCATACTTCAG

```
CCATGTGAGACAGAAGATTTAAGAGATGCCTTCAGGCTTCTTGGACTG----------TGA
        Zeus_faber
ATGATAACTAAACTA---------GACAGTGTGCTTCTGGCTCGGAAGAAATTCATTTACCACTAT
AAGAACATGCGCTGGGCAAAAGGCCGCTGTGAGACGTACCTCTGCTTTGTCGTCAAGAGGA
GAGTTGGACCCAATTCCCTGTCCTTTGACTTTGGACACCTTCGCAAT----------CGGACC
GGCTGCCATGTAGAGCTCCTGTTTCTACGTCACCTG------GGAGCCTTGTGCCCTGGACTGTG
GGGACACGGAGGCCCCTATGGA---GGGCGGCTCAGTTACTCAGTCACCTGGTTCTGCTCGTG
GTCTCCCTGCGCCAACTGCTCCTTCAGACTGGCCCAATTCCTCGGGCAGACG---------CCCAA
CCTCCGCCTCAGGATCTTTGTCTCCCGCCTCTACTACTGCGACCTTGAAGATAGCCGCGAGA
GAGAGGGCTTACGGATCCTGAAAAGAGCCGGAGTCCAAATCACAGTCATGAGCTACAAAG
ACTACTTCTATTGCTGGCAAACCTTCGTGGCTCACAGACAGACCAGCTTCAAGGCGTGGGA
TGAGCTGCACCAAAACTCAGTTCGCCTGGCCAGG---------AAA---CTAAACCGCATCCTCCAG
CCTTGTGAAACAGAAGATTTAAGAGATGCCTTCAAACTTCTTGGGTTCTTG----------TGA
;
end;

begin trees;

        tree speciestree =
(((Astyanax_mexicanus:121.77122741279602,Danio_rerio:121.77122741279602):101.99953637657163
,(((((((((((((((Gasterosteus_aculeatus:55.40747769565583,Myoxocephalus_scorpius:55.4074776956558
3):22.627589007329938,(Sebastes_norvegicus:71.41027871780396,Chaenocephalus_aceratus:71.41027
871780396):6.624787975823878):5.948298200333113,Perca_fluviatilis:83.98336487731933):27.23499
7279977804,((((Takifugu_rubripes:46.743550480651855,Tetraodon_nigroviridis:46.743550480651855)
:52.31449333152771,Antennarius_striatus:99.05804377441406):4.952094098567969,Spondyliosoma_c
antharus:104.0101378944397):3.8940865391969623,Symphodus_melops:107.90422446670532):3.3141
37740588194):4.0983751179695105,(((Oreochromis_niloticus:90.12568147794833,(Oryzias_latipes:69
.76327566986083,(Poecilia_formosa:18.384480726242064,Xiphophorus_maculatus:18.3844807262420
64):51.37879498329163):20.362405779966025):6.229750216332775,((Chromis_chromis:86.03637034
606933,Pseudochromis_fuscus:86.03637034606933):6.271233430540818,Parablennius_parvicornis:92.
30760380917813):4.047827894961586):15.657636823177327,((Helostoma_temminckii:64.1932831497
1924,Anabas_testudineus:64.19328314971924):42.49635722122191,Selene_dorsalis:106.68964036407
47):5.32342815576196):3.303668787118795):2.6957501523196754,(Thunnus_albacares:103.80823429
222107,Lesueurigobius_cf_sanzoi:103.80823429222107):14.204253155434131):5.149373299789431,C
hatrabus_melanurus:123.16186078948975):4.574033951210964,((Lamprogrammus_exutus:66.9588872
9228973,Carapus_acus:66.95888729228973):23.925274275398266,Brotula_barbata:90.8841615993499
8):36.85173311395644):9.028980841016761,((Myripristis_jacobus:59.01862996520996,(Holocentrus_r
ufus:14.652058449554444,Neoniphon_sammara:14.652058449554444):44.366571516036984):70.1287
3428974152,((Rondeletia_loricata:89.71703486652375,Acanthochaenus_luetkenii:89.71703486652375)
:23.455124383091928,Beryx_splendens:113.17215923690796):15.97520504798888):7.6175112740576
2):3.0117428180396644,Monocentris_japonica:139.776618334198):7.268487315320982,Lampris_gutta
uts:147.04510569152832):3.839686785376074,(((((((Molva_molva:42.47743926963806,(Brosme_bros
me:39.03891726341247,(((((Arctogadus_glacialis:5.222854929506778,Boreogadus_saida:5.222854929
506778):2.4513389710009097,(Theragra_chalcogramma:3.346329225230217,Gadus_morhua:3.346329
225230217):4.327864675396681):5.726030785477162,(Melanogrammus_aeglefinus:10.395505192184
448,Merlangius_merlangus:10.395505192184448):3.0047194936364896):4.48665917098522,Pollachiu
s_virens:17.886883866405487):9.021208262825013,(Trisopterus_minutus:22.696635680580137,Gadic
ulus_argenteus:22.696635680580137):4.211456434738636):12.130825152540204):3.43852200285792
97):3.19094077802896,Lota_lota:45.6683800485611):13.098774013638497,(Physis_blennoides:16.400
560005474087,Physis_physis:16.400560005474087):42.366594044685364):12.221774325680741,((Me
rluccius_merluccius:5.7998921918630600,Merluccius_polli:5.7998921924829485):61.73268190526217
,(Melanonus_zugmayeri:63.95622381646633,((((Macrourus_berglax:29.67249945344925,Malacocepha
lus_occidentalis:29.67249945344925):20.17204357004166,Bathygadus_melanobranchus:49.844543035
```

```
12574):10.645187737723738,(Mora_moro:36.98188234682083,Laemonema_laureysi:36.98188234682
083):23.5078484090443):1.939285897175786,(Muraenolepis_marmoratus:56.88453414344788,(Trachy
rincus_scabrus:12.07161490740776,Trachyrincus_murrayi:12.07161490740776):44.81291924285889):
5.5444825193190255):1.5272070992565432):3.576350340722499):3.4563542779281846):33.4718737
1263503,Stylephorus_chordatus:104.46080207824707):21.041643343019487,(Zeus_faber:32.85098531
341553,Cyttopsis_roseus:32.85098531341553):92.65146017112733):19.170966317129142,(Polymixia_
japonica:135.7494994041443,(Percopsis_transmontana:60.18133554153442,Typhlichthys_subterraneus
:60.18133554153442):75.56816387424469):8.923912356662754):6.211380692934995):7.56075037511
5873,Benthosema_glaciale:158.44554285736086):10.299416859668469,(Parasudis_fraserbrunneri:161.
78278560620342,Guentherus_altivela:161.78278560077345):6.962174173392896):25.78392385015487
8,(Osmerus_eperlanus:117.3724450843811,Borostomias_antarcticus:117.3724450843811):77.15643841
142654):17.481817657327667,(Salmo_salar_1:0.006629789359941943 95,Salmo_salar_2:0.008576995
53333491141):212.0107011795044):11.760062609915053):10,Lampetra_tridentata:232);

end;

begin mrbayes;
        log start filename=log.out;
        charset 1st_pos=1-.\3;
        charset 2nd_pos=2-.\3;
        charset 3rd_pos=3-.\3;
        partition by_codon=3:1st_pos,2nd_pos,3rd_pos;
        set partition=by_codon;
        lset applyto=(all) nst=6 rates=gamma;
        unlink revmat=(all) statefreq=(all) shape=(all) ratemultiplier=(all);
        prset applyto=(all) ratepr=variable;
        constraint gadiformes = 32 31 6 26 40 70 71 41 34 36 37 20 68 9 4 35 33 55 19 72 11 38 30 53
52;
        prset applyto=(all) topologypr=constraints(gadiformes);
        report applyto=(all) ancstates=yes siterates=yes;
        outgroup 21;
        showmodel;
        taxastat;
        mcmcp samplefreq=5000 printfreq=5000 nruns=24 nchains=10 starttree=current nperts=4
nswaps=3 temp=0.01;
        mcmcp savebrlens=yes filename=asr_gadiformes_gtr_outgroup_tree ngen=6000000;
        sump;
        sumt;

End;
```

Appendix 11: ProtASR setting and input files

ProtASR setting file

```
##########################################################################################
#####################################
##### Settings file for ProtASR 2.0
##### Ancestral sequence reconstruction of proteins under structurally constrained substitution models
##### Miguel Arenas, David Liberles & Ugo Bastolla
##### (c) 2014-2015
##### Contact: miguelmmmab@gmail.com
#####
##### Parameters with an "*" are mandatory (must be specified)
##### Text with an "#" is not read. Parameter values must be introduced immediately after the "="
##########################################################################################
#####################################


#########################################################
#### Alignment of amino acid sequences and tree ####
#########################################################

### Target alignment file with a rooted tree ### # nexus format with a rooted tree, see documentation and examples
*NameOfNexusFile=60_teleosts.nex


#############################
### Settings for the ASR  ###
#############################

# Substitution model: MEANFIELD (requires specification of settings in the following section), Blosum62, cpREV64, Dayhoff, DayhoffDCMUT, G1974a, G1974c, G1974p, G1974v, Grantham, HIVb, HIVw, JTT, JonesDCMUT, LG, Miyata, MtArt, MtMam, MtRev24, MtZoa, RtRev, VT, WAG
*SubsModel=MEANFIELD


# Consider frequencies from the model (+F) (0: No, 1: Yes)
*ModelFreqs=0


# Estimate (0) or fix (1) gamma shape parameter
*TypeG=1

# Gamma shape parameter value. Initial or fixed alpha, 0:infinity (constant rate)
*AlphaG=0

# Different alphas for genes, introduce a number
*Malpha=0


# Estimate (0) or fix (1) rho (correlation parameter)
*TypeRho=1
```

```
# Rho (correlation parameter). initial or fixed rho, 0:no correlation
*RhoCorr=0



##############################################################################################
###########
### Settings to compute the substitution model based on the protein structure - MEANFIELD - ###
##############################################################################################
###########

### Input files defining the protein ###
# PDB file (must be placed in the current directory)
*PDBfile=model3.pdb

# Chain of the PDB file
*CHAIN=A


### Thermodynamic model ###
# Temperature
*TEMP=0.5

# Configurational entropy per residue (unfolded)
*SU1=0.065

# Configurational entropy per residue (misfolded)
*SC1=0.065

# Configurational entropy offset (misfolded)
*SC0=0.0

# Use up to 1,2,3 moments of misfolding energy?
*REM=2

# Contacts map (must be placed in the ProtEvol directory)
*FILE_STR=structures.in


### Mean field model ###
# Number of substitutions to simulate data (0 by default, not required for ASR)
*TMAX=0

# LAMBDA~ NPOP*exp(-DELTA G/TEMP)
*LAMBDA_par=0.90

# Optimize LAMBDA? (0: No, 1: Yes, default)
OPT_LAMBDA=1

# Target value of DeltaG if OPT_LAMBDA
DG_OPT=-1

# Optimization criterion. Allowed: NAT ALL DG
```

```
*MODEL=ALL


### Mutation model ###
# Global matrix. Mean (0) or mean weighted by frequencies (1)
*GLOBALMATRIX=0

# Exchangeability. Allowed: MUT, EXCH, FLUX, RATE
*EXCHANGE=FLUX

# Rate matrix. Allowed: JTT, WAG
*MATRIX=JTT

# Get nucleotide frequencies from sequence? (0: Use input nucleotide frequencies, 1: Fit nucleotide frequencies from prot sequences, 2: Fit amino acid frequencies from prot sequences)
*GET_FREQ=2


## DNA Parameters for model MUT ##
# Frequency for A
*fA=0.25

# Frequency for T
*fT=0.25

# Frequency for C
*fC=0.25

# Frequency for G
*fG=0.25

# Transition transversion ratio (Kappa, >1)
*TT_RATIO=1.3

# Ratio between 1-nuc and 2-nuc mutations (0-1)
*TWONUCMUT=0.25
```

```
###############################################################################################
########################################
##### Settings file for ProtASR 2.2
##### Ancestral sequence reconstruction of proteins under structurally constrained substitution models
##### Miguel Arenas & Ugo Bastolla
##### (c) 2014-2018
##### Contact: miguelmmmab@gmail.com
#####
##### Parameters with an "*" are mandatory (must be specified)
##### Text with an "#" is not read. Parameter values must be introduced immediately after the "="
###############################################################################################
########################################


#####################################################
#### Alignment of amino acid sequences and tree ####
```

```
#######################################################

### Target alignment file with a rooted tree ### # nexus format with a rooted tree, see documentation and examples
*NameOfNexusFile=60_teleosts.nex


##############################
### Settings for the ASR  ###
##############################

# Substitution model: MEANFIELD (requires specification of settings in the following section), Blosum62, cpREV64, Dayhoff, DayhoffDCMUT, G1974a, G1974c, G1974p, G1974v, Grantham, HIVb, HIVw, JTT, JonesDCMUT, LG, Miyata, MtArt, MtMam, MtRev24, MtZoa, RtRev, VT, WAG
*SubsModel=MEANFIELD


# Consider frequencies from the model (+F) (0: No, 1: Yes)
*ModelFreqs=0


# Estimate (0) or fix (1) gamma shape parameter
*TypeG=0

# Gamma shape parameter value. Initial or fixed alpha, 0:infinity (constant rate)
*AlphaG=0.5

# Different alphas for genes, introduce a number
*Malpha=0


# Estimate (0) or fix (1) rho (correlation parameter)
*TypeRho=1

# Rho (correlation parameter). initial or fixed rho, 0:no correlation
*RhoCorr=0



#############################################################################################
###########
### Settings to compute the substitution model based on the protein structure - MEANFIELD - ###
#############################################################################################
###########

### Input files defining the protein ###
# PDB file (must be placed in the current directory)
*PDBfile=Gadus_morhua.pdb

# Chain of the PDB file
*CHAIN=A
```

```
### Thermodynamic model ###
# Temperature
*TEMP=0.5

# Configurational entropy per residue (unfolded)
*SU1=0.065

# Configurational entropy per residue (misfolded)
*SC1=0.065

# Configurational entropy offset (misfolded)
*SC0=0.0

# Use up to 1,2,3 moments of misfolding energy?
*REM=2

# Coefficient of local interactions
*A_LOC=0

# Contacts map (must be placed in the ProtEvol directory)
*FILE_STR=structures.in


### Mean field model ###
# Number of substitutions to simulate data (0 by default, not required for ASR)
*TMAX=0

# LAMBDA~ NPOP*exp(-DELTA G/TEMP)
*LAMBDA_par=0.90

# Optimize LAMBDA? (0: No, 1: Yes, default)
OPT_LAMBDA=1

# Target value of DeltaG if OPT_LAMBDA
DG_OPT=-1

# Optimization criterion. Allowed: NAT ALL DG
*MODEL=ALL

# WildType model: No (0) or yes (1)
*WTmodel=1


### Mutation model ###
# Global matrix. Mean (0) or mean weighted by frequencies (1)
*GLOBALMATRIX=0

# Exchangeability. Allowed: MUT, EXCH, FLUX, RATE
*EXCHANGE=FLUX

# Rate matrix. Allowed: JTT, WAG
*MATRIX=JTT
```

```
# Get nucleotide frequencies from sequence? (0: Use input nucleotide frequencies, 1: Fit nucleotide frequencies from prot sequences, 2: Fit amino acid frequencies from prot sequences)
*GET_FREQ=2

# Improve mutation parameters after selection?
*REMUT=0


## DNA Parameters for model MUT ##
# Frequency for A
*fA=0.25

# Frequency for T
*fT=0.25

# Frequency for C
*fC=0.25

# Frequency for G
*fG=0.25

# CpG transition ratio
*kCpG=2

# Transition transversion ratio (Kappa, >1)
*TT_RATIO=1.3

# Ratio between 1-nuc and 2-nuc mutations (0-1)
*TWONUCMUT=0.25
```

ProtASR input file

```
#NEXUS

[
Real data set from NCBI

PDBtaxa=Gadus_morhua
]

Begin data;
Dimensions ntax=74 nchar=217;
	Format datatype=protein gap=- missing=? matchchar=.;
	Matrix
Acanthochaenus_luetkenii
MITKLDRVLLAKETFIFHYENMRWAKGRHETYLCFVVKRRVGPDSLSFDFGHLRNRTGCHVE-
LLFLRHL--GTLCPGLWGYGGAGE-
RRLSYSITWFCSWSPCADCAFRVAQLIGRTPNLRLRIFVSRLYFCDLEDSRERGGLRLLKKAGVQ
ITVMSYKDFFYCWQTFVANGGSSFKAWDEMHQNSVRLASQLNHILQPCDTEDLRDAFKLLGL--
Anabas_testudineus
MITKLDSVLLPRKKFIYHYKNVRWARGRHETYLCFVVKRRVGPDSLTFDFGHLRNRNGCHVE-
MLFLRYL--GALCPGIWGYGGAGE-
KRLSYSITWFCSWSPCANCSLRLTQFLSQTPNLRLRIFVSRLYFCDMEDSREREGLRILKNAGVQI
TVMTYKDFFYCWQTFVDRKQSSFKAWDELHQNSVRLTRKLYRILQPCEIEDLRDAFKLLGL--
```

Antennarius_striatus
MITKLDSVLLPRKKFIYHYKNMRWARGRCETYLCFVVKRREGPDTLTFDFGHLRNRNGCHVE-
LLFLRYL--GALCPGLWGSGGTGE-
KRLSYSITWFCSWSPCANCSIRQCEFLSRTPNLRLRIFVSRLYFCDLEDSREREGLRMLKKAGVQI
SVMSYKDFFYCWQTFVASKQSSFKAWEELHQNSVRLARKLNRILQPCEAEDLRDAFKLLGL--
Arctogadus_glacialis
MISKLDSVLLAQNKFIYNYKNMRWAKGRNETYLCFVMKRRLGPDSLSFDFGHLRNRTGCHAE-
LLFLSYL--GALCPGLWGCADDRN-
RRLSYSVTWFCSWSPCANCATTLTRFLRQTPNLRLRIFVSRLYFCDLEGSPHVEGLRDLRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNYVRLSRK?NRILQPCETEDLRDVFRLFGL
LT
Astyanax_mexicanus
MTSKLDSILLTQKKFIYHYKNVRWARGRHETYLCFVVKRRIGPNSLSFDFGHLRNRSGCHVE-
LLFLRYL--GALCPGLGGLGVDGV--
KVGYAVTWFCSWSPCSNCAQRIAHILSQTPSLRLRIFVSRLYFCDNEDSLEREGLRHLLRAGVQI
TVMTYKDFFYCWQTFVARRESRFKAWDGLHQNSVRLSRKLKRILQPCQTEDLRDVFALLGL--
Bathygadus_melanobranchus
MISKLDSVLLAQKKFMYNYKNVRWAKGRHETYLCFVVRRRLGPNSLSFDFGHLRNRTGCHVE-
LLFLSHL--GALCPGLWGCVGDDN-
RRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLEDSPNIEGLRELRRAGV
QVIVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDAFRVIGL
LS
Benthosema_glaciale
MITKLDSVLLGQKKFLFHYKNVRWAWGRNETYLCFVVKRRVGPNSLSFDFGHLRNRSSCHAE-
LLFLRHL-
GGALCPGLWGYGGDGGEGRFNYSVTWFCSWSPCADCSLRLAQFLSRTPNLRLRIFVSRLYFCD
AEDSREREGLRTLKRAGVQITVMNYKDYYYCWQTFVAHRQSSFKAWADLHQNSVRLARKLH
RILQPCETEDFRDAFKLLGL--
Beryx_splendens
MITKLDSVLLAKKKFIYHYKNMRWAKGRHETYLCFVVKRRVGPDSLSFDFGHLRNRAGCHVE-
LLFLRHL--GALCPGLWGHGGSGE-
RKLSYSITWFCSWSPCADCSFRLAQFLNRTPNLRLRIFVSRLYFCDQEDSRERDGLRLLKKAGV
NITVMSYKDFFYCWQTFVANRTSRFKAWDLLHQNSVRLARKLNRILQPYEIEDLRDAFRLLGF--
Boreogadus_saida
MIRKLDSVLLAQNKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLHNRTGCHAE-
LLFLSYL--GALCPGLWGCADDRN-
RRLSYSVTWFCSWSPCANCATTLTRFLRQTPNLRLRIFVSRLYFCDLEGSPHVEGLRDPRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNYV?LSRKLNRILQPCETEDLRDVFRLFGL
LT
Borostomias_antarcticus
MISKLDSVLLAQKKFLFHYKNVRWARGRHETYLCFVVKRRVGPDSLTFDFGHLRNRTGCHVE-
LLFLRHL--
GVLCPGLSASGGAGGGRGLNYSITWFCSWSPCFDCSARLAQFLRRTPNLRLRLFVSRLYFCDPE
DRHEREGLRALKRAGVHITVMSYKDYFYCWQTFVAHRQRAFKAWEDLQQNSVRLARKLNSIL
LPCETEDLRDPFRLLGL--
Brosme_brosme
MMSKLDSVLLAQKKFIYNYKNLRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHVE-
LLFLSYL--GALCPGLWGCGGDRN-
QRLSYSVTWFCSWSPCANCAATLARFLRHTPNLRLRIFVARLYFCDLEGSPHIEGLRDLRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRALNRILQPCETEDLRDPFRLFGL
LT
Brotula_barbata
MIAKLDSVLLPRKKFIYHFKNMRWAKGRHETYLCFVVKRRVGPDSLSFDFGHLRNRNGCHVE-

LLFLRYL--GALCPGLWGCGNSGQ--
RLCYSITLFCSWSPCANCSERLAKFLGRTPNLRLRIFVSRLYFCDMEDSREREGLRMLKNAGVNI
TVMSYKDYFYCWQTFVARGASNFKAWDGLQENSIRLARKLTHILQPGETEDLRDAFKLLGM--
Carapus_acus
MTAKLDRVLLPRKKFLFHYKNVRWAKGRHETYLCFVVKRRVGPDSMSFDFGHLRNRSGCHVE
-LLFLRYL--GALCPGLWGYEGSGQ-
RRLSYSITWFCSWSPCANCSERLAQFLNRTPNLRLRIFVSRLYFCDLEDSREREGLRTLEKAGVH
ITIMSYKDYFYCWQTFVACGTSKFKAWDELHQNTTRLKRKLNRILQPCETEDLRDAFKLLGLL-
Chaenocephalus_aceratus
MITKLDSMLLPRKKFIYHYKNMRWARGRCETYLCFVVKRRVGPDSLTFDFGHLRNRNGCHVE-
MLFLRYL--DALCPGLLGCEGTGE-
KRLSYSITWFCSWSPCANCSIRLSQFLSQTPNLRLRIFVSRLYFCDMENSPARDGLIMLKKAGVQ
TSVMSYKDFFYCWHNFVDCKQSKFKPWEDLHQNSVRLARKLKRILQLCETEDLRDAFKLLGL--
Chatrabus_melanurus
MITKLDSVLLPRKKFIYHYKNMRWAKGRHETYLCFVVKRRMGPDSLSFDFGHLRNRNGCHVE-
LLFLRYL--GALCPGLWGYGIAGE-
RKLSYSVTWFCSWSPCVNCSLRLTQFLMQTPNLRLRIFVSRLYFCDMEDSREREGLRMLKKAG
VHITVMSYKDFFYCWQTFVACKESKFKAWEALHQNSVRLARKLNRILQPCETEDFRDAFKLLG
L--
Chromis_chromis
MITKLDSVLLPQKKFIYHYKNMRWARGRCETYLCFVIKKRAGPDSISFDFGHLRNRNGCHVE-
LLFLRYL--GALCPGLWGYGQ----
NRISYSITWFCSWSPCANCSLRLAQFLNQTPNLRLRIFVSRLYFCDMEDSREREGLRILKKAGVNI
TVMSYKDYFYCWQTFVARRLSKFKPWDGLQQNYVRLSRKLNRILQPCETEDFRDAFRLLGL--
Cyttopsis_roseus
MITKLDSVLLARKTFIYHYKNMRWAKGRHETYLCFVVKRRVGPDSLSFDFGHLRNRTGCHVE-
LLFLRHL--GALCPGLWGQGGADE-
RRLSYSVTWFCSWSPCANCSLRLVQFLGQTPNLRLRIFVSRLYYCDLEDSREREGLRTLKRAGV
QITVMSYKDYFYCWQTFVARRQTRFKAWDELHQNSVRLARKLNRILQPCETEDLRDAFKLLGF
L-
Danio_rerio
MICKLDSVLMTQKKFIFHYKNVRWARGRHETYLCFVVKRRIGPDSLSFDFGHLRNRSGCHVE-
LLFLRHL--GALCPGLSASSVDGA--
RLCYSVTWFCSWSPCSKCAQQLAHFLSQTPNLRLRIFVSRLYFCDEEDSVEREGLRHLKRAGVQ
ISVMTYKDFFYCWQTFVARRERSFKAWDGLHENSVRLVRKLNRILQPCETEDLRDVFALLGL--
Gadiculus_argenteus
MISKLDSVLLAQKKFIYNYNNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHAE-
VLFLSYL--GALCPGLWGCAGDRS-
LRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLEGSPHVEGLRDLRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDVFRLFGL
LT
Theragra_chalcogramma
MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHAE-
LLFLSYL--GALCPGLWGCADDRN-
RRLSYSVTWFCSWSPCANCATTLTRFLRQTPNLRLRIFVSRLYFCDLEGSPHVEGLRDLRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNYVRLSRKLNRILQPCETEDLRDVFRLFG
LLT
Gadus_morhua
MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHAE-
LLFLSYL--GALCPGLWGCADDRN-
RRLSYSVTWFCSWSPCANCATTLTRFLRQTPNLRLRIFVSRLYFCDLEGSPHVEGLRDLRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNYVRLSRKLNRILQPCETEDLRDVFRLFG
LLT

Gasterosteus_aculeatus
MIAKLDSVLLPRKKFIYHYTNMRWARGRHETYLCFVVKRRVGPDSLSFDFGHLRNRSGCHVE-
LLFLRHL--GALCPGFLGCGDTGG-
RRLSYSITWFCSWSPCVNCSISLSQFLSRTPNLRLRIFVSRLYFCDMENSRERDGLRMLKKAGVQ
VTVMSYKDFFYCWQTFVDRKQSQFKAWKELHQNSVRLSRKLKRILQPCETEDLRDAFKLLGL--
Guentherus_altivela
MITKLDSILMAQKKFIFHYKNMRWAKGRNETHLCFVVKRRLGPNSLSFDFGHLRNRTGCHVE-
LLFLRHL--GFLCPGLWGYGEPGE-
GRLNYSVTWFCSWSPCADCSFTLTHFLRETPNLRLRIFVSRLYFCDEEDSSAREGLRMLKKAGV
NITVMSYKDYFYCWKTFVAHRQRNFKAWDGLDQNSVHLARKLSHILQPWETADLRDAFKLLG
L--
Helostoma_temminckii
MITKLDSVLLPRKKFIYHYKNVRWARGRHETYLCFVVKRRVGPDSLTFDFGHLRNRNGCHVE-
MLFLRYL--GALCPGLWGCGGTGE-
RRLSYSITWFCSWSPCSNCSLRLAQFLSQTPNLRLRIFVSRLYFCDMEDSREREGLRILKNAGVQI
TVMSYKDFFYCWQTFVARKQSNFKAWEELHQNSVRLTRKLHRILQPCETEDLRDAFKLLGL--
Holocentrus_rufus
MITKLDSVLLAKKKFIYHYKNLRWAKGRHETYLCFVVKRRAGPDSIAFDFGHLRNRAGCHVE-
LLFLRYL--GALCPGLWGYGGTGE-
RKMSYSITWFCSWSPCANCSYRLAQFLNRTPNLRLRLFVARLYFCDIEDSREREGLRMLKNAGV
HITVMSYKDYFYCWQTFVARKTSNFKAWDGLHQNYVRLARKLNRILQPCETEDLRDAFRLLG
L--
Laemonema_laureysi
MISKLDSVLLAQKKFMFNYKNMRWARGRNETYLCFVVKRRLGPNSLSFDFGHLRNRTGCHVE-
LLFLSYL--GALCPGLWGCRGDEN-
RRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLEDSPHIEGLRDLRRAGV
RVTVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGL
LT
Lampris_guttauts
MISKLDSVLLTQKKFLYHYKNVRWAKGRHETYLCFVVKRRVGPDSMSFDFGHLRNRAGCHVE
-LLFLRYL--GALCPGLWGYGDTGD-
RRLSYSVTWFCSWSPCANCSFRLAQFLQRTPNFRLRLFVSRLYFCDMEDSSERDGLRLLKNAGV
QITVMSYKDYFYCWQTFVAHRKSSFKAWDGLHQNTVRLARLLNRILQPCEAEDLRDAFKLLGF
--
Lamprogrammus_exutus
MIAKLDSVLLPRKKFIFHYKNMRWAKGRHETYLCFVVKRRVGPDSLSFDFGHLRNRNGCHVE-
LLFLRYL--GALCPGLWGCGGSGE-
RRLSYSITWFCSWSPCANCSQRLSQFLSQTPNLRLRIFVSRLYFCDMENSREREGLRMLKNAGV
QITVMSYKDFFYCWQTFVACGKSKFKAWDELHRNSVRLTRKLNRILQPWETEDLRDAFRLLGF
--
Lesueurigobius_cf_sanzoi
MITKLDSVLLPKKKFIFHYKNVRWAKGRHETYLCFVVKRRVGPNSMSFDFGHLRNRSGCHVE-
ILFLRYL--GALCPGLWGAGGSEE-
RRLSYSITWFCSWSPCANCSTKLSQFLAKTPNLRLRIFVSRLYFCDLEDSIEREGLRMLKRAGVQ
LTVMKYKDYFYCWHTFVARNQSNFKAWEELHQNSVRLTRKLSRILQPCETEDLRDAFRLLGL--
Lota_lota
MISKLDSVLLAQKKFIYNYKNIRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHVE-
LLFLSYL--GALCPGLWGCGGDRN-
RRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLEGSPHIEGLRDLRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLFGL
LT
Macrourus_berglax
MISKLDSILLAQKKFKYNYNNMRWAKGRNETYLCFVVKRRLGPNSLSFDFGHLRNRAGCHVE-

LLFLSHL--GALCPGLWGFGGAEN-
IRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCELADSPHSEGLRELRRAGVQ
VNVMTYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNLILQPCETEDLRDAFRLIGLL
T
Malacocephalus_occidentalis
MISKLDSVLLAQKKFIYNYKNIRWAKGRNETYLCFVVKRRLGPNSLSFDFGHLRNRTGCHVE-
LLFLSYL--GALCPGLWGCGGADN-
RRLNYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLDDSPHTEGLRELRRAGV
QFTVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGL
LS
Melanogrammus_aeglefinus
MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHAE-
LLFLSYL--GALCPGLWGCAGDRN-
RRLSYSVTWFCSWSPCANCATTLTRFLRQTPNLRLRIFVSRLYFCDLEGSPHVEGLRDLRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNYVRLSRKLNRILQPCETEDLRDVFRLFG
LLT
Melanonus_zugmayeri
MISNLDSVLLAQKKFMYNYKNMHWAKGRNATYLCFVVKRRLGPDSLSFDFGHLHNRTGCHA
E-LLFLSHL--GALCPGLWGCGGDKN-
RRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARVYFCEQEDSPHIEGLRDLRRAGV
QVTVMSYKDYFYCWQTFVAHRLSRFKTWEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGL
LT
Merlangius_merlangus
MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHAE-
LLFLSYL--GALCPGLWGCAGDRN-
RRLSYSVTWFCSWSPCANCATTLSRFLRQTPNLRLRIFVSRLYFCDLEGSPHVEGLRDLRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNYVRLSRKLNRILQPCETEDLRDVFRLFG
LLT
Merluccius_merluccius
MISKLDSVLLAQKKFMYNYKNMRWAKGRNQTYLCFVVKRRLGPDSLSFDFGHLHNRTGCHA
E-LLFLSHL--GALCPGLWGCGGDEN-
RRLSYSVTWFCSWSPCANCAATLARFLRLTPNLRLRIFVARLYFCDVEDSPHREGLRNLRRAGV
LVNVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRTLNRILQPCETEDLRDAFRLIGL
LT
Merluccius_polli
MISKLDSVLLAQKKFMYNYKNMRWAKGRNQTYLCFVVKRRLGPDSLSFDFGHLHNRTGCHA
E-LLFLSHL--GALCPGLWGCGGDEN-
RRLSYSVTWFCSWSPCANCAATLARFLRLTPNLRLRIFVARLYFCDVEDSPHREGLRNLRRAGV
LVNVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRTLNRILQPCETEDLRDAFRLIGL
LT
Molva_molva
MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHVE-
LLFLSYL--GALCPGLWGCGGDTN-
RRLSYSVTWFCSWSPCANCAATLARFLRHTPNLRLRIFVARLYFCDLEGSPHIEGLRDLRRAGV
QVKVMSYKDYFYCWQTFVAHKLSRFKAWEGLHTNYVRLSRKLNRILQPCETEDLRDAFRLFG
LLT
Monocentris_japonica
MITKLDSVLLAQKKFIYHYKNMRWARGRHETYLCFVVKRRVGPDSLSFDFGHLRNRSGCHVE-
LLFLRHL--GALCPGLWGYGGTGE-
RRLSYSITWFCSWSPCADCSFRLVQFLGRTPNLRLRIFVSRLYFCDVEDSRERQGLRMLKKAGV
QITVMSYKDYFYCWQTFVAHRQSSFKAWDELHQNSVRLARKLNRILQPCETEDLRDAFKLLGL
--

Mora_moro
MISTLDSVLLAQKKFMYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHVE
-LLFLSHL--GALCPGLWGCGGDEN-
RRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLEDSPHIEGLRDLRRAGV
QVTVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGL
LT
Muraenolepis_marmoratus
MISKLDSVLLGQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSMSFDFGHLRNRAGCHVE
-LLFLSHL--GALCPGLWGCGGDEN-
RRLSYSVTWFCSWSPCANCAATLARLLRQTPNLRLRIFVARLYFCDLEGSPHSEGLRDLRRAGV
QVNVMSYKDYFYCWQTFVAHRVSRFKAWEGLHTNSVRLSRKLNRILQPRETDDLRDAFRLIGL
LT
Myoxocephalus_scorpius
MITKLDSVLLQQKKFIYHYKNMRWARGRHETYLCFVVKRRVGPDSLSFDFGHLRNRTGCHVE-
LLFLRYL--GALCPGLWGYGGTGE-
KRLSYSITWFCSWSPCINCSISLSQFLNRTPNLRLRIFVSRLYFCDKENSRERDGLRMLKNAGVQI
TVMSYKDFFYCWQTFVDRKKSNFKAWEELHQNSVRLARKLNRILQPCEAEDLRDAFKLLGL--
Myripristis_jacobus
MITKLDSMLLAKKKFIYHYKNMRWAKGRHETYLCFVVKRRVGPDSMSFDFGHLRNRAGCHV
E-LLFLRYL--GALCPGLWGCGGNTE-
KKLSYSITWFCSWSPCADCSFRLAQFLNRTPNLRLRIFVSRLYFCDLEDSREREGLRMLKKAGV
QITVMSYKDYFYCWQTFVAHRMSSFKAWDGLHQNYVRLARKLNRILQASETEDLRDAFKLLG
L--
Neoniphon_sammara
MITKLDSVLLAKKKFIYHYKNLRWAKGRHETYLCFVVKRRVGPDSIAFDFGHLRNRAGCHVE-
LLFLRYL--GALCPGLWGYGGTGE-
RKLSYSITWFCSWSPCANCSFRLAQFLNRTPNLRLRIFVSRLYFCDVEDSREREGLRMLKNAGV
HITVMSYKDYFYCWQTFVARKTSSFKAWDGLHQNYVRLARKLNRILQPCDTEDLRDAFRLLG
L--
Oreochromis_niloticus
MIAKLDSMLLPRKKFLYHYKNVRWARGRNETYLCFVVKRRVGPDSLSFDFGHLRNRNGCHVE
-LLFLRQL--GTLCPGLSGYGFHGE-
RRVSYSITWFCSWSPCANCSSRLAQFLKQTPNLRLRIFVSRLYFCDMEDSREREGLRLLKKVGV
HITVMSYKDFFYCWENFVA-QQSKFKAWEGLHQNTVRLARKLNRILQPCDTEDLRDAFKLLGL-
-
Oryzias_latipes
MITKLDSVLLPKKKFIYHYKNMRWARGRHETYLCFVVKRRVGPESLSFDFGHLRNRNGCHVE-
LLFLRHL--SALCPGLWGYGATGQ-
GRVSYSITWFCSWSPCANCSFRLAQFLSQTPNLRLRIFVSRLYFCDLEDSREREGLRMLKKVGV
HITVMSYKDYFYCWQTFVARKQSKFKPWDGLHQNSVRLSRKLNRILQPCETEDFRDAFKLLGL
--
Osmerus_eperlanus
MISTLDGVLLAQKKFIYHYKNMRWARGRHETYLCFVIKRRVGPDSLSFDFGHLRNRTGCHVE-
LLFLRHL--
GALCPGLWGTGGAGGGVRLSYSITWFCSWSPCSACSHRLSDFLSRTPNLRLRIFVSRLYFCDPED
SLEREGLRMLKRAGVNITVMSYKDYFYCWETFVARRKTGFKAWDGLHHNSVRLARKLYRILQ
PCETEDLRDAFTLLGL--
Parablennius_parvicornis
MIAKLDSMLLPRKKFIYHYKNMRWAKGRHETYLCFVVKRRLGPDSLSFDFGHLRNRNGCHVE-
LLFLRHL--GTLCPGLSGYGVHGE-
KRLSYSITWFCSWSPCSNCSHRLAQFLSRTPNIRLRIFVSRLYFCDLEDSREREGLRLLKKTGVHI
TVMSYKDYFYCWQTFVASNQSRFKPWDELQRNSIRLTRKLNRILQPCETEDLRDAFKLLGL--

Parasudis_fraserbrunneri
MITNLDSVLLAQKKFIYHYKNMRWARGRHETYLCFVVKRRLGPDSLSFDFGHLRNRSGCHVE-
LLFLRHL--GALCPGLWGYGGE---
KRLSYSVTWFCSWSPCADCSTRLSQFLSRTPNLRLRIFVSRLYFCDLEDSLAREGLRTLKRVGVQ
VTVMSYKDYFYCWQTFVARRQSSFKAWDGLQQNSVRLARKLNRILQPCETEDLRDAFKLLGL-
-
Perca_fluviatilis
MITKLDSVLLPRKKFIYHYKNMRWARGRHETYLCFVVKRRVGPDSLSFDFGHLRNRNGCHVE-
LLFLRYI--GALCPGLWGCSGTGE-
RRLSYSITWFCSWSPCANCSIRLSQFLSQTPNLRLRIFVSRLYFCDTENSPERDGLRMLKKAGVQI
TVMSYKDFFYCWQTFVDRKQSNFKAWEELHSNSVRLSRKLNRILQPFETEDLRDAFKLLGL--
Percopsis_transmontana
MITKLDSVLLAQKKFIFHYKNMRWARGRHETYLCFVIKRRVGPNSLSFDFGHLRNRSGCHVE-
ILFLRHL--GALCPGLWGEGGTGE-
RRLSYSITWFCSWSPCANCSLRLAQILRQLPNLRLRIFVSRLYFCDLEDSKERDGLRMLKNVGV
QITVMSYKDYFYCWQTFVAHRKSNFKAWDGLHQNSVRLARKLNRILQPCEIEDLRDAFKLLGF
--
Phycis_blennoides
MISKLDSVLLAQKKFIYNYKNIRWAKGRNETYLCFVVKRRLGPNSLSFDFGHLRNRAGCHVE-
LLFLSHL--GALCPGLWGCVDDSN-
RRLSYSVTWFCSWSPCANCAATLARFLRMTPNLRLRIFVARLYFCDLEDSPHIEGLRHLRRAGV
EVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLFGL
LT
Phycis_phycis
MISKLDSVLLAQKKFLYNYKNIRWAKGRNETFLCFVVKRRLGPNSLSFDFGHLRNRAGCHVE-
LLFLSHL--GALCPGLWGCVDDSN-
RRLSYSVTWFCSWSPCANCAATLARFLRMTPNLRLRIFVARLYFCDLEDSPHIEGLRHLRRAGV
EVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLFGL
LT
Poecilia_formosa
MITKLDRALLPRKKFIYHYKNLRWARGRCETYLCFVVKKRVGPDSLSFDFGHLRNRNNCHVE-
LLFLRHL--GALCPGLWGYGVTGE-
RKVSYSVTWFCSWSPCANCSIRLAQFLHQTPNLRLRIFVSRLYFCDLEDSREREGLRILKKAGVH
ITVMSYKDYFYCWQTFVAKSQSKFKPWDGLHQNYIRLSRKLNRILQPCETEDLRDAFRLLGL--
Pollachius_virens
MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHAE-
LLFLSYL--GALCPGLWGCADDRN-
RRLIYSVTWFCSWSPCANCATTLARFLRQTPNLRLRIFVSRLYFCDLEGSPHVEGLRDLRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNYVRLSRKLNRILQPCETEDLRDVFGLFG
LLT
Polymixia_japonica
MITKLDSVLLAQKKFIYHYKNMRWAKGRHETYLCFVVKRRVGPDSMSFDFGHLRNRSGCHVE
-LLFLRHL--GALCPGLWGYGGTGE-
KRLSYSVTWFCSWSPCSNCSYRLAQFLSQTPNLRLRIFVSRLYFCDLEDSRERDGLRMLKRAGV
QITVMTYKDYFYCWQTFVAHRTSKFKAWDELHRNSVRLSRILNRILQPCETEDLRDAFRLLGL--
Pseudochromis_fuscus
MIAKLDSVLLPKKKFIFHYKNMRWARGRHETYLCFVVKRRRGPDSLSFDFGHLRNRNGCHVE-
LLFLRYL--GALCPGLWGYGATGA-
SRLSYSITWFCSWSPCANCSFRLAQFLSQTPNLRLRIFVSRLYFCDMEDSREREGLRQLKKAGVH
ITVMSYKDYFYCWQTFVARNQSKFKPWDELHQNSVRLSRKLNRILQPCETEDLRDAFKLLGL--
Rondeletia_loricata
MITKLDSVLLAKKKFIYHYKNMRWARGRHETYLCFVVKRRVGPDSLSFDFGHLRNRTGCHVE-
LLFLRHL--GALCPGLWGHGGTGE-

RRLSYSITWFCSWSPCADCSFRLAQFLGRMPNLRLRIFVSRLYFCDLEDSREREGLRLLKKAGV
QITVMSYKDFFYCWQTFVAHRNCSFKAWDEMHQNSVRLARKLNRILQPCETEDLRDAFKLLG
L--
Salmo_salar_1
MINKFDSVLLAQKKFIYHYKNMRWAKGRHETYLCFVVKRRVGPNSLSFDFGHLRNRSGCHVE-
LLFLRLLEAGALCPGLWGYGAPDS-
VGLCYSVTWFCSWSPCSDCSYRLAQFLSQTPNLRLRIYVSRLYFCDPEDSSAREGLRMLQRAGV
QITVMNYEDYFYCWQTFVACRQRVFKAWDGLHQNSVQLARKLNDILQPGEAEDWGDAFELL
GL--
Salmo_salar_2
MINKFDSVLLAQKKFIYHYKNMRWAKGRHETYLCFVVKRRGGPNSLSFDFGHLRNRSGCHVE-
LLFLRLLEAGALCPGLWGYGAPDS-
VGLCYSVTWFCSWSPCSDCSYRLAQFLSQTPNLRLRIYVSRLYFCDPEDSSAREGLRMLQRAGV
QITVMNYEDYFYCWQTFVACRQRVFKAWDGLHQNSVQLARKLNDILQPGEAEDWGDAFELL
GL--
Sebastes_norvegicus
MITKLDSVLLPRKKFIFHYKNMRWARGRHETYLCFVVKRRVGPDSLTFDFGHLRNRNGCHVE-
LLFMRYL--GALCPGLWGQGVPGE-
KRLSYSITWFCSWSPCVNCSVTLSQFLSKTPNLRLRIFVSRLYFCDMENSRERDGLRMLKKAGV
QISVMSYKDYFYCWQTFVDRKQSKFKAWDEMHQNSVRLTRKLSRILQPSETEDLRDAFKLLGL
--
Selene_dorsalis
MITKLDSVLLPRKKFIFHYKNVRWAKGRHETYLCFVVKRRVGPDSMTFDFGHLRNRNGCHVE-
ILFLRYL--GALCPGLWGYGVGGE-
KRLSYSITWFCSWSPCANCSSRLAQFLKQTPNLRLRIFVSRLYFCDLEDSQEREGLRILKKAGVHI
TVMTYKDFFYCWQTFVARKQSSFKAWDELHQNSVRLARKLQRILQPCETEDLRDAFKLLGL--
Spondyliosoma_cantharus
MITKLDSVLLPKKKFIYHYKNVRWARGRHETYLCFVVKRRVGPDTLTFDFGHLRNRNGIHVE-
LLFLRYL--GALCPGLWGYGGTGE-
KRLSYSITWFCSWSPCANCSLRLCQFLSQTPNLRLRIFVSRLYFCDMEDSREREGLRMLKKAGV
QITVMSYKDFFYCWQTFVARRASQFKAWEELQRNSVRLTRKLNRILQPCETEDLRDAFKLLGL-
-
Stylephorus_chordatus
MIAKLDSVLLARNKFIYHYKNMRWAKGRNETYLCFVVKRRVGPDSLAFDFGHLRNRTGCHVE
-LLFLRHL--GALCPGLWG-GAAGD-
KRLSYSVTWFCSWSPCANCASTLAQFLRQTPNLRLRLFVARLYFCDLEDSPDREGLRILRRAGV
HITVMRYKDYFYCWQTFVAHNQSRFKAWEGLHPNSVRLSRTLNRILQPCETEDLRDAFKLLGL-
-
Symphodus_melops
MNTKLDSVLLPRKKFIYHYKNVRWARGRHETYLCFVIKRRVGPDTLTFDFGHLRNRNGCHVE-
LLFLRYL--GALCPGLLGYGGAGE-
KRLSYSITWFCSWSPCSNCSTILCQFLSKMPNLRLRLFVSRLYFCDMEDSREREGLRMLKKVGV
QITIMSYKDFFYCWQKFVARRQSNFKAWEELHQNSVRLSRKLNRILQPCETEDLRDAFKLLGL--
Takifugu_rubripes
MITKLDSMLLPRKKFIYHYKNVRWARGRHETYLCFVVKRRVGPDTLTFDFGHLRNRSGCHVE-
LLFLRYL--GALCPGLWGYGAAGE-
KRLSYSVTWFCSWSPCVNCSIQLCQFLNNTPNLRLRIFVSRLYFCDLEDSLEREGLRMLTKAGV
RISVMSYKDYFYCWQKFVDCKKSNFKAWEELHQNSVRLTRKLNRILQAWDLEDLRDALKLLG
F--
Tetraodon_nigroviridis
MITKLDSMLLPRKKFLYHYKNVRWARGRHETYLCFVVKRRVGPDTLTFDFGHLRNRNGCHVE
-LLFLRYL--GALCPGLWGYGAAGE-

KRLSYSITWFCSWSPCANCSIQLSQFLRNTPNLRLRIFVSRLYFCDMEDSLEREGLRMLSRAGVR
ISVMSYKDFFYCWQKFVDSKTSSFKAWEELHQNSVRLTRKLNRILQSWDLEDLRDALKLLGL--
Thunnus_albacares
MITKLDSVLLPRKKFIYHYKNVRWARGRHETYLCFVVKRRVGPDSLSFDFGHLRNRNGCHVE-
LLFLRYL--GALCPGVWGYGNTGQ--
RISYSITWFCSWSPCANCSRRLAQFLSQVPNVRLRIFVSRLYFCDLEDSRERDGLRLLKNAGVQI
TVMSYKDFFYCWQTFVARNQSKFKAWEELHRNSVRLTRTLNRILQPCDIDDLRDAFKLLGL--
Trachyrincus_murrayi
MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRLGCHVE-
LLFLSHL--GALCPGLWGCGGDVN-
RRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLEDSPHIEGLRDLRRAGV
QVTVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGL
LT
Trachyrincus_scabrus
MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRLGCHVE-
LLFLSHL--GALCPGLWGCGGDEN-
RRLSYSVTWFCSWSPCANCAATLARFLRHTPNLRLRIFVARLYFCDLEDSPHIEGLRDLRRAGV
QVTVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDAFRLIGL
LT
Trisopterus_minutus
MISKLDSVLLAQKKFIYNYKNLRWAKGRNETYLCYVVKRRLGPDSLSFDFGHLRNRTGCHVE-
LLFLSYL--GALCPGLWGCTDDRN-
RRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIFVARLYFCDLEGSPHIEGLRHLRRAGV
QVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSVRLSRKLNRILQPCETEDLRDVFGLFGL
LT
Typhlichthys_subterraneus
MISKLDSVLLAQKKFIFHYKNMRWARGRNETYLCFVIKRRVGPDSLSFDFGHLRNRSGCHVE-
LLFLRHL--GALCPGLWGQGGTGD-
NRLSYSITWFCSWSPCSNCSHRLAQFLSQLPNLRLRIFVSRLYFCDLEDSREREGLRMLKNAGVH
ITVMSYKDYYYCWQTFVARRKSKFKAWEGLHQNSVRLARKLNRILQPCEIEDLRDAFKLLGF--
Xiphophorus_maculatus
MITKLDRVLLPKKKFIYHYKNMRWARGRCETYLCFVVKKRVGPDSLSFDFGHLRNRNNCHVE-
LLFLRHL--GALCPGLWGYGVTGE-
RKVSYSITWFCSWSPCANCSFRLAQFLHQTPNLRLRIFVSRLYFCDLEDSREREGLRMLKKAGV
HITVMSYKDYFYCWQTFVAKSQSKFKPWDGLHQNCIRLTRKLNRILQPCETEDLRDAFRLLGL-
-
Zeus_faber
MITKLDSVLLARKKFIYHYKNMRWAKGRCETYLCFVVKRRVGPNSLSFDFGHLRNRTGCHVE-
LLFLRHL--GALCPGLWGHGGPYG-
GRLSYSVTWFCSWSPCANCSFRLAQFLGQTPNLRLRIFVSRLYYCDLEDSREREGLRILKRAGV
QITVMSYKDYFYCWQTFVAHRQTSFKAWDELHQNSVRLARKLNRILQPCETEDLRDAFKLLGF
L-
;
End;

BEGIN TREES;
TREE part_1 =
((Astyanax_mexicanus:121.77122741279602,Danio_rerio:121.77122741279602):101.99953637657163,
((((((((((((((Gasterosteus_aculeatus:55.40747769565583,Myoxocephalus_scorpius:55.4074776956558
3):22.627589007329938,(Sebastes_norvegicus:71.41027871780396,Chaenocephalus_aceratus:71.41027
871780396):6.624787975823878):5.948298200333113,Perca_fluviatilis:83.98336487731933):27.23499
7279977804,((((Takifugu_rubripes:46.743550480651855,Tetraodon_nigroviridis:46.743550480651855)

```
:52.31449333152771,Antennarius_striatus:99.05804377441406):4.952094098567969,Spondyliosoma_c
antharus:104.0101378944397):3.8940865391969623,Symphodus_melops:107.90422446670532):3.3141
37740588194):4.0983751179695105,(((Oreochromis_niloticus:90.12568147794833,(Oryzias_latipes:69
.76327566986083,(Poecilia_formosa:18.384480726242064,Xiphophorus_maculatus:18.3844807262420
64):51.37879498329163):20.362405779966025):6.229750216332775,((Chromis_chromis:86.03637034
606933,Pseudochromis_fuscus:86.03637034606933):6.271233430540818,Parablennius_parvicornis:92.
30760380917813):4.047827894961586):15.657636823177327,((Helostoma_temminckii:64.1932831497
1924,Anabas_testudineus:64.19328314971924):42.49635722122191,Selene_dorsalis:106.68964036407
47):5.32342815576196):3.303668787118795):2.6957501523196754,(Thunnus_albacares:103.80823429
222107,Lesueurigobius_cf_sanzoi:103.80823429222107):14.204253155434131):5.149373299789431,C
hatrabus_melanurus:123.16186078948975):4.574033951210964,((Lamprogrammus_exutus:66.9588872
9228973,Carapus_acus:66.95888729228973):23.925274275398266,Brotula_barbata:90.8841615993499
8):36.85173311395644):9.028980841016761,((Myripristis_jacobus:59.01862996520996,(Holocentrus_r
ufus:14.652058449554444,Neoniphon_sammara:14.652058449554444):44.366571516036984):70.1287
3428974152,((Rondeletia_loricata:89.71703486652375,Acanthochaenus_luetkenii:89.71703486652375)
:23.455124383091928,Beryx_splendens:113.17215923690796):15.97520504798888):7.6175112740576
2):3.0117428180396644,Monocentris_japonica:139.776618334198):7.268487315320982,Lampris_gutta
uts:147.04510569152832):3.839686785376074,(((((((Molva_molva:42.47743926963806,(Brosme_bros
me:39.03891726341247,(((((Arctogadus_glacialis:5.222854929506778,Boreogadus_saida:5.222854929
506778):2.4513389710009097,(Theragra_chalcogramma:3.346329225230217,Gadus_morhua:3.346329
225230217):4.327864675396681):5.726030785477162,(Melanogrammus_aeglefinus:10.395505192184
448,Merlangius_merlangus:10.395505192184448):3.0047194936364896):4.48665917098522,Pollachiu
s_virens:17.886883866405487):9.021208262825013,(Trisopterus_minutus:22.696635680580137,Gadic
ulus_argenteus:22.696635680580137):4.211456434738636):12.130825152540204):3.43852200285792
97):3.19094077802896,Lota_lota:45.6683800485611):13.098774013638497,(Phycis_blennoides:16.400
560005474087,Phycis_phycis:16.400560005474087):42.366594044685364):12.221774325680741,((Me
rluccius_merluccius:5.7998921918630600,Merluccius_polli:5.7998921924829485):61.73268190526217
,(Melanonus_zugmayeri:63.95622381646633,(((((Macrourus_berglax:29.67249945344925,Malacocepha
lus_occidentalis:29.67249945344925):20.17204357004166,Bathygadus_melanobranchus:49.844543035
12574):10.645187737723738,(Mora_moro:36.98188234682083,Laemonema_laureysi:36.98188234682
083):23.5078484090443):1.939285897175786,(Muraenolepis_marmoratus:56.88453414344788,(Trachy
rincus_scabrus:12.07161490740776,Trachyrincus_murrayi:12.07161490740776):44.81291924285889):
5.5444825193190255):1.5272070992565432):3.576350340722499):3.456354277928184 6):33.4718737
1263503,Stylephorus_chordatus:104.46080207824707):21.041643343019487,(Zeus_faber:32.85098531
341553,Cyttopsis_roseus:32.85098531341553):92.65146017112733):19.170966317129142,(Polymixia_
japonica:135.7494994041443,(Percopsis_transmontana:60.18133554153442,Typhlichthys_subterraneus
:60.18133554153442):75.56816387424469):8.923912356662754):6.211380692934995):7.56075037511
5873,Benthosema_glaciale:158.44554285736086):10.299416859668469,(Parasudis_fraserbrunneri:161.
78278560620342,Guentherus_altivela:161.78278560077345):6.962174173392896):25.78392385015487
8,(Osmerus_eperlanus:117.3724450843811,Borostomias_antarcticus:117.3724450843811):77.15643841
142654):17.481817657327667,(Salmo_salar_1:0.0066297893599419 4395,Salmo_salar_2:0.008576995
53333491141):212.0107011795044):11.760062609915053);

END;

BEGIN ASSUMPTIONS;
        CHARSET span_1 = 1-217;

END;
```

Appendix 12: Ancestral AID sequences predicted in this thesis

Ancestral sequences predicted by RAxML based on the *aicda* gene tree
Gd-ANC: ATGATTAGTAAGCTAGACAGTGTGCTCTTAGCCCAGAAGAAATTCATATACAATTACAAGA ACATGCGATGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGAAGGC TCGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTA GAGCTGCTGTTTCTGAGCTACCTGGGGGCGCTGTGCCCGGGCCTCTGGGGGTGCGGAGGCG ACAGAAACCGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTGGTCTCCCTGTGCCAAC TGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCG TGGCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCATATAGAGGGCTTGAGGGACCT GAGGAGAGCAGGGGTCCAGGTCAAAGTTATGAGCTACAAAGACTACTTCTACTGCTGGCAG ACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTG TCCGTCTGTCAAGAAAACTAAACCGCATCCTCCAGCCATGTGAAACAGAAGATTTAAGAGA TGCTTTCAGACTTTTTGGACTGTTAACCTGA
Gf-ANC: ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTACAAGA ACATGCGTTGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGAAGGCT TGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTAG AGCTGCTGTTTCTGAGCCACCTGGGGGCACTGTGCCCAGGCCTGTGGGGGTGCGGAGGCGA CGAAAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTGGTCTCCCTGTGCCAACT GTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGT GGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCGCATATAGAGGGCTTGAGGGACCTG AGGAGAGCAGGGGTGCAGGTCACCGTTATGAGCTACAAAGACTACTTCTACTGCTGGCAGA CCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGT CCGTCTGTCCAGAAAACTAAACCGCATCCTCCAGCCATGTGAAACAGAAGATTTAAGAGAT GCTTTCAGACTTATTGGGCTGTTAACCTGA
Zg-ANC: ATGATTACTAAACTAGACAGTGTGCTTCTGGCTCGGAAGAAATTCATTTACCACTATAAGA ACATGCGCTGGGCAAAAGGCCGCAATGAGACATACCTCTGCTTTGTCGTCAAGAGAAGAGT TGGACCCGATTCCCTGTCCTTTGACTTTGGACACCTTCGCAATCGGACTGGCTGCCATGTAG AGCTCCTGTTTCTACGTCACCTGGGGGCCCTGTGCCCTGGACTGTGGGGACACGGAGGCGC TGATGAAAGAAGGCTCAGTTACTCAGTCACCTGGTTCTGCTCCTGGTCTCCCTGCGCCAACT GCTCCTTCAGACTGGCCCAATTCCTCGGGCAGACGCCCAACCTCCGTCTCAGGATCTTTGTC TCCCGTCTCTACTACTGTGACCTTGAAGATAGCCGCGAGAGAGAGGGCTTAAGGATCCTGA AAAGAGCCGGAGTCCAAATCACAGTCATGAGCTACAAAGACTACTTCTATTGCTGGCAGAC CTTCGTGGCTCACAGACAGACCCGCTTCAAGGCGTGGGATGAGCTGCACCAAAACTCAGTT CGTCTGGCCAGGAAACTAAACCGCATCCTCCAGCCTTGTGAAACAGAAGATTTAAGAGATG CTTTCAAACTTCTTGGGTTCTTGACCTAA
Ancestral sequences predicted by RAxML based on the species tree previously published
Gd-ANC: ATGATTAGTAAGCTAGACAGTGTGCTCTTAGCCCAGAAGAAATTCATATACAATTACAAGA ACATGCGATGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGAAGGC TCGGACCTGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTA GAGCTGCTGTTTCTGAGCTACCTGGGGGCGCTGTGCCCGGGCCTCTGGGGGTGCGGAGGCG ACAGAAACCGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTGGTCTCCCTGTGCCAAC TGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCG TGGCTCGCCTCTACTTCTGTGACCTGGAGGGCAGTCCGCATATAGAGGGCTTGAGGGACCT GAGGAGAGCAGGGGTCCAGGTCAAAGTTATGAGCTACAAAGACTACTTCTACTGCTGGCAG ACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTG TCCGTCTGTCAAGAAAACTAAACCGCATCCTCCAGCCATGTGAAACAGAAGATTTAAGAGA TGCTTTCAGACTTTTTGGACTGTTAACCTGA
Gds-ANC:

ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTACAAGA ACATGCGTTGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGAAGGCT TGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTAG AGCTGCTGTTTCTGAGCCACCTGGGGGCGCTGTGCCCGGGCCTGTGGGGGTGCGGAGGCGA CGAAAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTGGTCTCCCTGCGCCAACT GTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGT GGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCGCATATAGAGGGCTTGAGGGACCTG AGGAGAGCAGGGGTCCAGGTCACCGTTATGAGCTACAAAGACTACTTCTACTGCTGGCAGA CCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGT CCGTCTGTCCAGAAAACTAAACCGCATCCTCCAGCCATGTGAAACAGAAGATTTAAGAGAT GCTTTCAGACTTATTGGGCTGTTAACCTGA
Gf-ANC: ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTACAAGA ACATGCGTTGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGAAGGCT TGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTAG AGCTGCTGTTTCTGAGCCACCTGGGGGCGCTGTGCCCGGGCCTGTGGGGGTGCGGAGGCGA CGAAAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTGGTCTCCCTGTGCCAACT GTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGT GGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCGCATATAGAGGGCTTGAGGGACCTG AGGAGAGCAGGGGTCCAGGTCACCGTTATGAGCTACAAAGACTACTTCTACTGCTGGCAGA CCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGT CCGTCTGTCCAGAAAACTAAACCGCATCCTCCAGCCATGTGAAACAGAAGATTTAAGAGAT GCTTTCAGACTTATTGGACTGTTAACCTGA
Zg-ANC: ATGATTACTAAACTAGACAGTGTGCTTCTGGCTCGGAAGAAATTCATTTACCACTATAAGA ACATGCGCTGGGCAAAAGGCCGCAATGAGACATACCTCTGCTTTGTCGTCAAGAGAAGAGT TGGACCCGATTCCCTGTCCTTTGACTTTGGACACCTTCGCAATCGGACCGGCTGCCATGTAG AGCTCCTGTTTCTACGTCACCTGGGGGCCCTGTGCCCTGGACTGTGGGGACACGGAGGCGC TGATGAAAGAAGGCTCAGTTACTCAGTCACCTGGTTCTGCTCCTGGTCTCCCTGCGCCAACT GCTCCTTCAGACTGGCCCAATTCCTCGGGCAGACGCCCAACCTCCGTCTCAGGATCTTTGTC TCCCGTCTCTACTACTGTGACCTTGAAGATAGCCGCGAGAGAGAGGGCTTACGGATCCTGA AAAGAGCCGGAGTCCAAATCACAGTCATGAGCTACAAAGACTACTTCTATTGCTGGCAGAC CTTCGTGGCTCACAGACAGACCCGCTTCAAGGCGTGGGATGAGCTGCACCAAAACTCAGTT CGTCTGGCCAGGAAACTAAACCGCATCCTCCAGCCTTGTGAAACAGAAGATTTAAGAGATG CTTTCAAACTTCTTGGGTTCTTGACCTAA
Ancestral sequences predicted by MrBayes
Gd-ANC: ATGATTAGTAAGCTAGACAGTGTGCTCTTAGCCCAGAAGAAATTCATAATCAATTACAAGA ACATGCGATGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGAAGGC TTGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTA GAGCTGCTGTTTCTGAGCCACCTGGGGGCGCTGTGCCCGGGCCTCTGGGGGTGCGGAGGCG ACAGAAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTGGTCTCCCTGTGCCAAC TGTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCG TGGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCGCATATAGAGGGCTTGAGGGACCT GAGGAGAGCAGGGGTCCAGGTCAAAGTTATGAGCTACAAAGACTACTTCTACTGCTGGCAG ACCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTG TCCGTCTGTCAAGAAAACTAAACCGCATCCTCCAGCCATGTGAAACAGAAGATTTAAGAGA TGCTTTCAGACTTTTTGGACTGTTAACCTGA
Gds-ANC: ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATATACAATTACAAGA ACATGCGTTGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGAAGGCT TGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTAG

AGCTGCTGTTTCTGAGCCACCTGGGGGCGCTGTGCCCGGGCCTGTGGGGGTGCGGAGGCGA CGAAAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTGGTCTCCCTGTGCCAACT GTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGT GGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCGCATATAGAGGGCTTGAGGGACCTG AGGAGAGCAGGGGTCCAGGTCACCGTTATGAGCTACAAAGACTACTTCTACTGCTGGCAGA CCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGT CCGTCTGTCCAGAAAACTAAACCGCATCCTCCAGCCATGTGAAACAGAAGATTTAAGAGAT GCTTTCAGACTTATTGGACTGTTAACCTGA
Gf-ANC: ATGATTAGTAAGCTAGACAGTGTGCTCTTGGCCCAGAAGAAATTCATGTACAATTACAAGA ACATGCGTTGGGCAAAAGGCCGCAACGAGACCTACCTCTGCTTCGTAGTGAAGAGAAGGCT TGGACCCGATTCCCTGTCTTTCGACTTCGGACACCTACGCAATCGCACTGGCTGCCACGTAG AGCTGCTGTTTCTGAGCCACCTGGGGGCACTGTGCCCAGGCCTGTGGGGGTGCGGAGGCGA CGAAAACAGAAGACTCAGCTACTCGGTCACCTGGTTCTGCTCCTGGTCTCCCTGTGCCAACT GTGCGGCCACGCTGGCCCGGTTCCTGAGGCAGACGCCCAACCTGCGCCTCAGGATCTTCGT GGCTCGCCTCTACTTCTGTGACCTGGAGGACAGTCCGCATATAGAGGGCTTGAGGGACCTG AGGAGAGCAGGGGTGCAGGTCACCGTTATHAGCTACAAAGACTACTTCTACTGCTGGCAGA CCTTCGTAGCTCACAGGCTGAGCCGCTTCAAGGCCTGGGAAGGGCTGCATACCAATTCTGT CCGTCTGTCCAGAAAACTAAACCGCATCCTCCAGCCATGTGAAACAGAAGATTTAAGAGAT GCTTTCAGACTTATTGGGCTGTTAACCTGA
Zg-ANC: ATGATTACTAAACTAGACAGTGTGCTTCTGGCCCGGAAGAAATTCATCTACCATTATAAGA ACATGCGCTGGGCAAAAGGCCGGCATGAGACATACCTCTGCTTTGTAGTGAAGAGGAGAGT TGGACCCGATTCCCTGTCCTTTGACTTTGGACACCTCCGCAATCGCACTGGCTGCCATGTAG AGCTGCTGTTCCTGCGCCACCTGGGGGCCCTGTGCCCTGGACTGTGGGGATACGGAGGCGC TGGTGAAAGGAGGCTCAGTTACTCAGTCACCTGGTTCTGCTCCTGGTCTCCCTGCGCCAACT GCTCCTTCAGACTGGCCCAATTCCTCAGGCAGACGCCCAACCTCCGCCTCAGGATCTTCGTC TCTCGCCTCTACTTCTGTGACCTGGAGGACAGCCGCGAGAGAGAGGGCCTAAGGATCCTGA AAAGAGCCGGAGTGCAAATCACAGTCATGAGCTACAAAGACTACTTCTATTGCTGGCAGAC CTTTTGTGCTCACAGACAGAGCAGCTTCAAGGCCTGGGATGGGCTGCACCAAAACTCTGTT CGCCTGGCCAGGAAACTAAACCGCATCCTCCAGCCTTGTGAAACAGAAGATTTAAGAGATG CTTTCAAACTTCTTGGGTTGTTGTGA
Ancestral sequences predicted by ProtASR
Gd-ANC: MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHVEL LLFLSHLEGGALCPGLWGCGGDENGRRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIF VARLYFCDLEDSPHIEGLRDLRRAGVQVKVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSV RLSRKLNRILQPCETEDLRDAFRLFGLLT
Gds-ANC: MISKLDSVLLAQKKFMYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHVE LLLFLSHLEGGALCPGLWGCGGDENGRRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIF VARLYFCDLEDSPHIEGLRDLRRAGVQVTVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSV RLSRKLNRILQPCETEDLRDAFRLIGLLT
Gf-ANC: MISKLDSVLLAQKKFIYNYKNMRWAKGRNETYLCFVVKRRLGPDSLSFDFGHLRNRTGCHVEL LLFLSHLEGGALCPGLWGCGGDENGRRLSYSVTWFCSWSPCANCAATLARFLRQTPNLRLRIF VARLYFCDLEDSPHIEGLRDLRRAGVQVTVMSYKDYFYCWQTFVAHRLSRFKAWEGLHTNSV RLSRKLNRILQPCETEDLRDAFRLIGLLT
Zg-ANC: MITKLDSVLLAQKKFIYHYKNMRWAKGRHETYLCFVVKRRVGPDSLSFDFGHLRNRTGCHVE LLLFLRHLEGGALCPGLWGYGGTGEGRRLSYSVTWFCSWSPCANCSFRLAQFLSQTPNLRLRIF

VSRLYFCDLEDSREREGLRILKRAGVQITVMSYKDYFYCWQTFVAHRQSRFKAWDELHQNSVR
LARKLNRILQPCETEDLRDAFKLLGLLT