Development of a rapid and high throughput yeast-based system for expression and purification of AID

by

Faezeh Borzooee

A thesis submitted to the school of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Science in Medicine

Graduate program in Immunology and Infectious disease, Faculty of Medicine

Memorial University of Newfoundland

October 2017

St. John's, Newfoundland

Abstract

The DNA mutator enzyme, activation–induced cytidine deaminase (AID) is expressed in mature B lymphocytes, and initiates the secondary antibody diversification events of somatic hypermutation (SHM) and class switch recombination (CSR) which mutates cytidine to uridine at Immunoglobulin loci. To date, the expression hosts for purification of AID/APOBECs include bacteria, insect or mammalian cell lines. Here we established and optimized a novel yeast–based expression/secretion system to express AID. This is compatible with different purification tags including GST and His positioned at either N- or C-termini or non-tag. This method also can be adapted for protein secretion into the media, allowing for a simpler purification method of concentrating and collecting the protein from the media, rather than current conventional purification methods. As proof of principle, we provide the first report of expression of functionally active untagged AID as well as GST-AID and AID-His which can address several important unanswered questions remaining in the field.

Acknowledgments

I would like to thank my thesis advisor Dr. Mani Larijani for his exceptional guidance, supervision and help during my graduate research career.

I take this opportunity to thank my thesis committee Dr. Dr. Michael Grant and Dr. Ken Kao for their comments and suggestions.

Also, I would like to extend my thanks to my present and former lab members for friendship and mutual support.

Table of Contents

List of tables

List of figures

List of Abbreviations and Symbols

µg	Microgram
µl	Microliter
°C	Degrees Celsius
A	Adenine
AOX	Alcohol oxidase
AOX1	Alcohol oxidase gene
AID	Activation-Induced Cytidine Deaminase
APOBEC	Apolipoprotein B m-RNA-editing, enzyme-catalytic polypeptide-like
A3A	Apolipoprotein B m-RNA-editing, enzyme-catalytic polypeptide- like 3A
A3B	Apolipoprotein B m-RNA-editing, enzyme-catalytic polypeptide-like 3B
A3D	Apolipoprotein B m-RNA-editing, enzyme-catalytic polypeptide-like 3D
A3F	Apolipoprotein B m-RNA-editing, enzyme-catalytic polypeptide-like 3F
A3G	Apolipoprotein B m-RNA-editing, enzyme-catalytic polypeptide- like 3G
A3H	Apolipoprotein B m-RNA-editing, enzyme-catalytic polypeptide-like 3H
bp	base pair
BSA	Bovine serum albumin
BMGY	Buffered Glycerol-complex Medium
BMMY	Buffered Methanol-complex Medium
C	Cytidine
CSR	Class switch recombination
DLCLs	Diffuse Large-cell Lymphomas

DHα5	*Escherichia coli* strain
DNA	Deoxyribonucleic acid
DSB	Double stranded break
DTT	Dithiothreitol
EDTA	Ethyldiamine tetraacetic acid
g	Gram
G	Guanine
GRAS	Generally Recognized As Safe
GAP	Glyceraldehyde 3-phosphate dehydrogenase
GST	Glutathione S-transferase
His	Histidine
IDT	Integrated DNA technologies
Ig	Immunoglobulin
kDa	Kilodalton
LB	Lysogeny broth
M	Molar

NEB	New England Biolabs
nt	Nucleotide
ORF	Open Reading Frame
PBS	Phosphate buffered saline
PAX-5	Paired box 5
PCR	Polymerase chain reaction
PIM1	Pim-1 proto-oncogene
RhoH	Ras homolog family member H
RNA	Ribonucleic acid
RESS	Repression under Secretion Stress
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SHM	Somatic hypermutation
S	Switch region
ssDNA	Single stranded DNA
T	Thymidine
TBE	Tris borate/EDTA electrophoresis buffer
TEMED	Tetramethylethylenediamine
SDS PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SHM	Somatic Hypermutation
ssDNA	Single stranded DNA
U	Uracil
UDG	Uracil-DNA glycosylase
V(D)J	Recombination of variable (V), diversity (D) and joining (J) genes

WRC	a ssDNA hotspot motif targeted by AID where W=A/T and R=A/G
YNB	Yeast nitrogen base without amino acids
YPD	Yeast Extract Peptone Dextrose Medium
YPDS	Yeast Extract Peptone Dextrose Medium with Sorbitol
Zn	Zinc

Introduction

1.1 The activation- induced cytidine deaminase (AID) enzyme is required for antibody diversification.

There are a total of eleven family members AID/APOBEC enzymes including APOBEC1, activation-induced deaminase (AID), APOBEC2, APOBEC3s (A3A, A3B, A3C, A3D, A3F, A3G, and A3H), and APOBEC4 (Table 1). AID /APOBECs are DNA-editing cytidine deaminases with the activity of deaminating deoxycytidine (dC) in single stranded DNA (ssDNA), RNA and supercoiled transcribed DNA, thus converting it into deoxyuridine (dU) (Conticello et al., 2008). AID acts in the antigen-driven antibody diversification process by mutating genomes of activated lymphocytes while the A3 branch relatives mutate genomes of retroviruses, thus mediating adaptive/innate anti-viral defense (Honjo et al., 2004; Honjo et al,. 2002; Harris et al,. 2003; Conticello et al,. 2005 and Harris et al,. 2008). Human AID is a small enzyme of 198 amino acids that is normally expressed only in B lymphocytes that have been activated by their specific antigen and migrated to lymph node germinal centers (GC). *In vitro*, purified AID preferentially mutates dC within the trinucleotide WRC (W=A/T, R=A/G) motif, and this is the same sequence specificity that is observed when AID is expressed in bacteria or eukaryotic cells as well as the pattern found in B lymphocytes (Larijani et al., 2007; Rada et al., 2004; Martin et al., 2002; Honjo et al., 2008 and Conticello et al., 2008). In activated B lymphocytes, AID targets the antibody genes, and by creating C:G to T:A base changes, it initiates antibody genes hypermutation (SHM) and class switch recombination (CSR), leading to antibody diversification (Fig. 1a) (Longerich et al., 2006; Muramatsu et al., 1999; Muramatsu et al., 2000; Revy et al., 2000; Martin et al., 2002; Barreto et al., 2005; Neuberger et al., 2005). Although AID is somewhat preferentially targeting the antibody

genes in activated B cells, it often targets non-Ig genes in immune cells, or is aberrantly expressed in non-immune cells. As a result, there are a host of AID-driven genomic lesions that lead to the transformation into cancer. These include AID mutation of the proto-oncogenes PIM1, MYC, RhoH/TTF (ARHH), and PAX5 (Pasqualucci et al., 2008), tumor suppressor and genomic stability genes, and AID-driven IgH/cmyc, IgH/bcl2 and IgH/bcl6 chromosomal translocations (Shen et al., 2000 and Jankovic et al., 2009) that lead to lymphomas derived from germinal center. This genomic instability leads to rampant mutations and chromosomal translocations that transform normal cells into cancers. In epigenetics, AID may mediate so-called active demethylation by deamination of methylated deoxycytidines, which has the potential to alter gene expression(Fig. 1b) (Mechtcheriakova et al., 2012; Durandy et al., 2006; Dooley et al., 2006; Revy et al., 2000).

Table 1. A summary of known members in human APOBEC family and their properties

Enzyme	Tissue expression	Associated function	Targets
APOBEC1	GI tract	Unknown	Apolipoprotein B mRNA
APOBEC2	Heart and skeletal muscle	Viral and retrotransposon restriction	Unknown
APOBEC3A	Primary monocytes, keratinocytes	Viral and retrotransposon restriction	Adeno-associated virus, HPV, LINE-1
APOBEC3B	T cells, PBMC, Intestine, uterus, mammary gland, keratinocytes, other	Viral and retrotransposon restriction	Retroviruses, retrotransposons, HBV
APOBEC3C	Many tissues	Viral and retrotransposon restriction	Retroviruses, HBV, retrotransposons, HPV
APOBEC3D	Thyroid, spleen, blood	Viral restriction	Retroviruses,
APOBEC3F	Many tissues, T cells	Viral and retrotransposon restriction	Retroviruses, retrotransposons, HBV, HIV
APOBEC3G	Many tissues, T cells	Viral and retrotransposon restriction	Retroviruses, retrotransposons, HBV, HIV, HCV
APOBEC3H	Poorly expressed due to a premature stop codon	Viral restriction once expression is optimized	Retroviruses,
APOBEC4	Testis	Unknown	Unknown
AID	Activated B-cells, primordial germ cells, Breast Tissue	Antibody maturation	Immunoglobulin gene

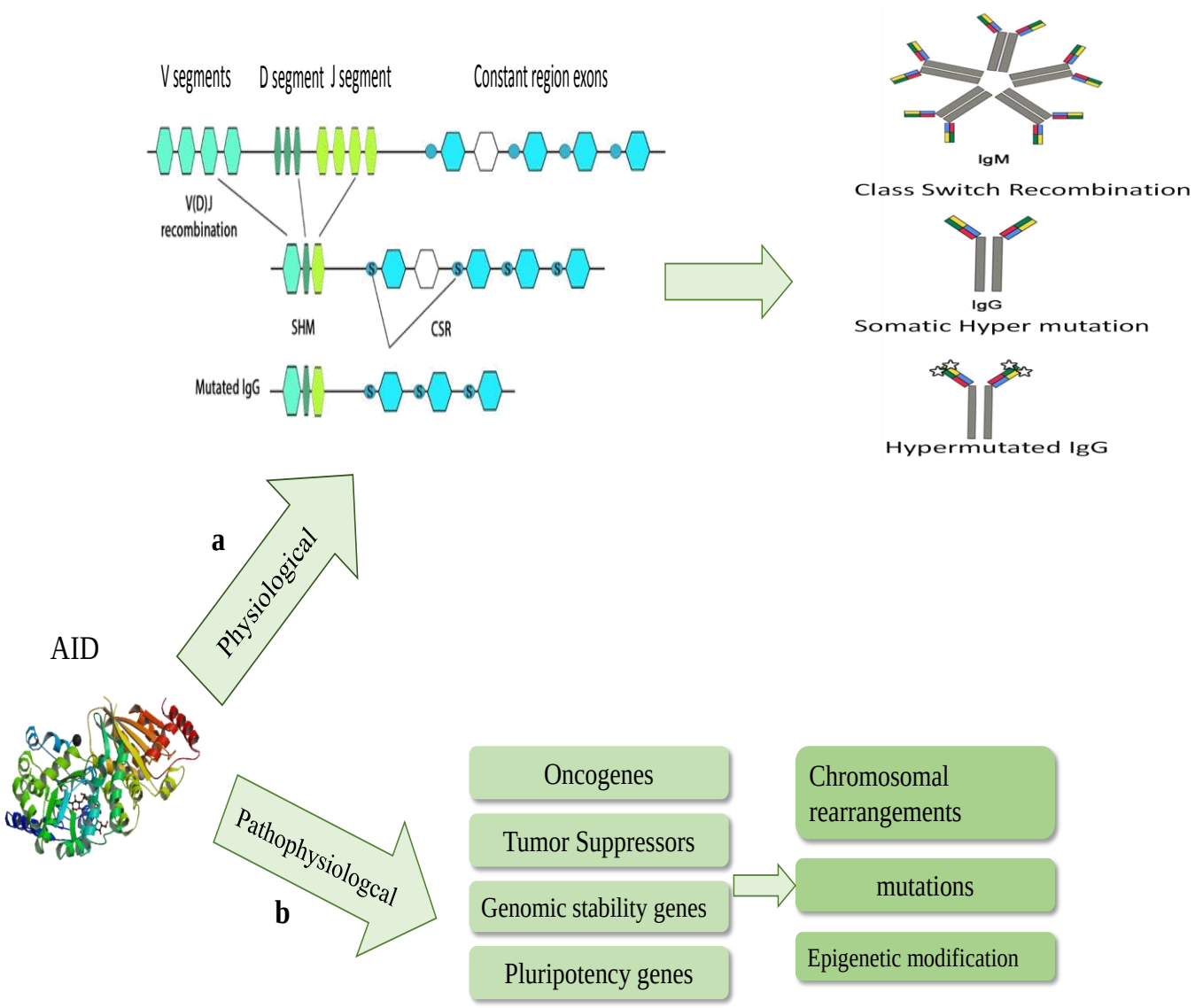

V segments
D segment
J segment
Constant region exons
V(D)J recombination
SHM
CSR
Mutated IgG
IgM
Class Switch Recombination
IgG
Somatic Hyper mutation
Hypermutated IgG
AID
a
Physiological
b
Pathophysiologcal
Oncogenes
Tumor Suppressors
Genomic stability genes
Pluripotency genes
Chromosomal rearrangements
mutations
Epigenetic modification

Figure 1. AID in immunity and cancer.

(**a**) Somatic hypermuation (SHM) and class switch recombination (CSR) of B-cell immunoglobulin (Ig) genes are initiated by AID-mediated mutations. Secondary antibody diversification process that occurs in the germinal center of the lymph nodes is SHM. It targets antigen binding regions in the variable sequences of both heavy and light chains and increases their affinity to specific antigens by introducing point mutations in the V(D)J exon. CSR process takes place in the germinal centers of the lymph nodes and is driven by AID which introduces double stranded breaks in the switch (S) regions resulting in the removal of the constant region and joining of a new constant region. (**b**) AID, although usually tightly regulated to avoid off-target activity, has been shown to target nuclear DNA and cause cancer. Under pathophysiological circumstances, AID may target non-Ig genes, including cancer-related genes such as oncogenes, tumor suppressor genes, genomic stability genes, and pluripotency genes, leading to accumulation mutations and chromosomal translocations and epigenetics modification.

1.2 Current methods for studying purified AID for function and structure

In order to study the enzymatic activity, molecular mechanisms and structure of AID as well interactions with other cellular proteins, it first needs to be purified. Unfortunately, AID is notoriously difficult to purify, and we are currently one of only a handful of labs with the expertise to work with a purified AID enzyme (Larijani et al., 2012; Robbiani et al., 2009; Cocker et al., 2016). The reason for this difficulty is two-fold: The first issue is that AID is by virtue of its activity and biological purpose, highly genotoxic. That is, when expressed at high levels, it can inflict mutations to the genome of the very host cells that it is expressed in, eventually causing cell death. For this reason, it can only be expressed for brief periods and at low levels, compared to other non-toxic proteins.

To date, we and others have developed expression systems for AID in both prokaryotic (bacteria), and eukaryotic (human, mouse or insect cells) hosts. Each has its own advantages: bacterial systems allow for the production of ample quantities of enzyme (in the milligram range per purification) because bacterial cultures are easily scaled up in volume. However, only a small fraction of the AID molecule is active because bacteria lack the sophisticated protein production machinery of eukaryotic cells and most often mis-fold human proteins. We estimated from classical Michaelis-Menten enzymes kinetic analysis enzyme assays that in the best cases only 1 in 1000-10000 AID molecule is folded correctly (Larijani et al., 2007), with the remainder acting as inert proteins, much like a non-specific protein such as Bovine Serum Albumin (BSA) in solution. On the other hand, eukaryotic cells produce a much higher quality of AID, but they are slow growing, much more prone to death due to heavy genome damage caused by AID, and time-consuming and costly to scale up. As a result the overall yield of AID protein is typically 2-3

orders of magnitude lower than in bacterial systems (in the nanogram-microgram range per purification).

The second issue is that AID, like some other APOBECs, has a surface that is highly positively charged. In fact, amongst AID/APOBECs, AID carries the highest positive charge (+14) which we and others have shown is due to an over-abundance of positively-charged lysine and arginine residues, particularly on the surface (King et al., 2015 and 2017; Pham et al., 2016). This is in line with its function, which requires binding to negatively charged ssDNA in order to mutate it; however, it presents a challenge in purification because it results in considerable non-specific protein-protein interactions with other cellular proteins (King et al., 2017). One way to mitigate this issue has been to attach bulky purification tags, to either the N-terminal or C-terminal of AID, and use these tags as a bait to purify AID. Examples of these protein tags range from the 6 amino acid long poly histidine (His) tag to the 200 amino acid long Glutathione S transferase (GST) tag. We and others have successfully used both tags with their corresponding affinity binding columns (His-binding or GST-binding) to purify various AID/APOBEC enzymes (Bransteitter et al., 2009; Abdouni et al., 2013; Abdouni et al., 2017 submitted). This method has the advantage that it greatly facilitates purification, but the disadvantage that the resulting AID enzyme is modified, because the presence of a tag at their N- or C-termini can exert subtle influences on the folding of the AID itself. This is especially relevant for AID, since we and others have shown that even the residues at the ends of the N- and C-termini have critical functions in regulating the enzymatic activity, cellular trafficking and structural folding integrity (King et al., 2015 and 2017; Zahn et al., 2014 ; Mondal et al., 2016).

Another disadvantage of purification tags is that they can interfere with the natural oligomerization of AID. Certain tags (e.g. GST) are self-attracting thus resulting in artificial dimerization of the

protein they are attached to, whilst others (e.g. His-) are self-repelling thus hindering the natural dimerization of proteins they are attached to (Larijani et al., 2012). In principle, it is possible to cleave purification tags after or during the purification; however we and others have found that the extended time and incubation conditions of these procedures often result in the inactivation and aggregation of AID. A peculiarity of purifying enzymes, in contrast to other non-enzyme proteins, is that due to their innately mobile and transient protein structures, they typically ought to be frozen as soon as purified in order to preserve structural integrity and maintain enzymatic activity for further study (Larijani et al., 2009).

1.3 Yeast protein expression system

A wide range of host cells are available to produce recombinant proteins. The most common expression systems are prokaryotic and mammalian cell based (Rosano et al., 014; Wingfield 2015; Li et al., 2013; Terpe et al., 2006; Aricescu et al., 2006; Jia et al., 2016). For the last two decades, bacteria such as *Escherichia coli* have been the factory of choice for the simple low cost expression of the majority of proteins. However, the prokaryote lacks intracellular organelles needed for post-translational modifications and therefore, is not able to properly fold and express many eukaryotic proteins, especially larger and more complex proteins (Jia et al., 2016; Sahdev et al., 2008). Alternatively, mammalians cells and other eukaryotic cells such as insect cells are popular choices to express eukaryotic proteins with high quality, but in lower amount since scaling up the cultures requires expensive infrastructure and media. In this light, yeast has been viewed as an attractive alternative eukaryotic expression system (Macauley-Patrick et al., 2005; Shobayashi et al., 2006). Compared with mammalian cells, yeast does not require a complex growth medium or culture conditions, is genetically relatively easy to manipulate, and has a similar protein synthesis pathway

(Wang et al., 2016; Ahmad et al., 2014). In the past decade, the methylotrophic yeast, *Pichia pastoris* (*Komagataella phaffii*), has been genetically engineered to produce hundreds of recombinant proteins important to industrial, pharmaceutical, and basic research purposes (Kurtzman 2009; Macauley-Patrick et al., 2005).

There are many advantages of using *P. Pastoris* over other yeast strains for expression of heterologous proteins: it can be easily manipulated on a genetic level (Wang et al., 2016; Ahmad et al., 2014; Graf et al., 2009), does not undergo fermentation like *Saccharomyces cerevisiae*, can add many eukaryotic posttranslational modifications, and has a number of strong promoters available to drive protein expression intracellularly as well as in secreted form (Lin-Cereghino et al., 2005). Since *P. pastoris* secretes only small amounts of endogenous proteins, the secreted recombinant protein constitutes the vast majority of the total protein in the medium (Lin-Cereghino et al., 2008). *P. pastoris* grows quickly in a wide range of temperatures from 15°C to 30°C, and a pH ranging from 3.0 to 7.0 (Lin-Cereghino et al., 2008), thus providing a wide window of optimization for expression of each specific protein. Moreover, *P. pastoris* has an efficient methanol-inducible promoter from the alcohol oxidase I gene (AOX1) which is commonly used for foreign gene expression. A strong methanol inducible promotor from the AOX 1 gene, PAOX1 can express a high level of recombinant protein compared to Alcohol Oxidase 2 promotor PAOX2. While AOX2 is about 97% homologous to AOX1, growth on methanol is much slower than with AOX1 (Tschopp et al., 1987; Kurtz et al., 1989; Cos et al., 2006; Graf et al., 2009). A characteristic attribute of this promoter is that it is strongly repressed in the presence of some carbon sources such as glucose and glycerol, but induced by methanol. Furthermore, *P. pastoris'* preference for respiratory rather than fermentative metabolism, even at high-biomass, prevents the accumulation of secondary metabolites like ethanol and acetic acid which can be detrimental for cell growth as

well as damaging to the exogenously expressed protein. Finally, *P. pastoris* is a GRAS (Generally Recognized As Safe) organism, and can be grown in many different types of bioreactors (Noseda et al., 2014). Because of these characteristics, some proteins that cannot be expressed efficiently in bacteria, *S. cerevisiae* or the insect cell/baculovirus system, have been successfully produced in functionally active form in *P. pastoris*. (Lin-Cereghino et al., 2013).

As with any expression system, choosing an optimal selection marker and promoter is essential for obtaining maximum protein quality and yield. To minimize the cytotoxic effects of AID we used the AOX1 promotor which has been the most widely utilized inducible promotor in pPICZ/pPICZα *P. Pastoris*-specific expression vectors. However, other promoter options such as constitutive GAP (glyceraldehyde 3-phosphate dehydrogenase) are also available for the production of foreign proteins in *Pichia* (Cos et al., 2006). Zeocin is a broad spectrum antibiotic that acts as a strong antibacterial and anti-tumor drug, shows strong toxicity against bacteria, fungi (including yeasts), plants and mammalian cells and is often used as a selection marker in yeast expression systems (Baron et al., 1992; Drocourt et al., 1990). *P. pastoris* is capable of integrating multiple copies of the vector DNA via recombination into its genome at sites of sequence homology. Like in *S. cerevisiae*, linear DNA can generate stable transformants of *P. pastoris* via homologous recombination between the vector DNA and regions of homology within the genome. Once integrants are identified by using a selection marker such as Zeocin on growth plates, they typically exhibit very stable gene expression even in the absence of selective pressure even when present as multiple copies (Pichia Expression kit Manual, Invitrogen, USA). Fig. 2 demonstrates selection marker and promotor in *Pichia* pPICZ/pPICZα vectors. The *S. cerevisiae* α- mating factor prepro-signal is the most widely used secretion signal, being in some cases a more efficient director of secretion than the leader sequence of the native heterologous protein (Cregg et al., 1993; Scorer

et al., 1993). However, variability in the number of N-terminal amino acids is commonly reported with heterologous proteins secreted using the α-factor prepro-leader (Kurjan et al., 1982). The α-mating factor prepro-peptide signal leader consists of a 19-amino acid signal (pre) sequence followed by a 67-residue (pro) sequence containing three consensus N-linked glycosylation sites and a dibasic Kex2 endopeptidase processing site (Kurjan et al., 1982). Processing of the α-mating factor secretion signal occurs in three steps: first, the pre-signal is removed by signal peptidases in the endoplasmic reticulum; second, Kex2 endopeptidase severs the pro leader sequence between the arginine and lysine; and finally Ste13 protein rapidly cleaves the Glu-Ala repeats in the Golgi. As demonstrated in Fig.2, the small pre-signal is believed to be important for interaction with the signal recognition particle, subsequent translocation into the Endoplasmic Reticulum (ER), and folding of the nascent protein. (Lin-Cereghino et al., 2013; Kurjan et al., 1982).

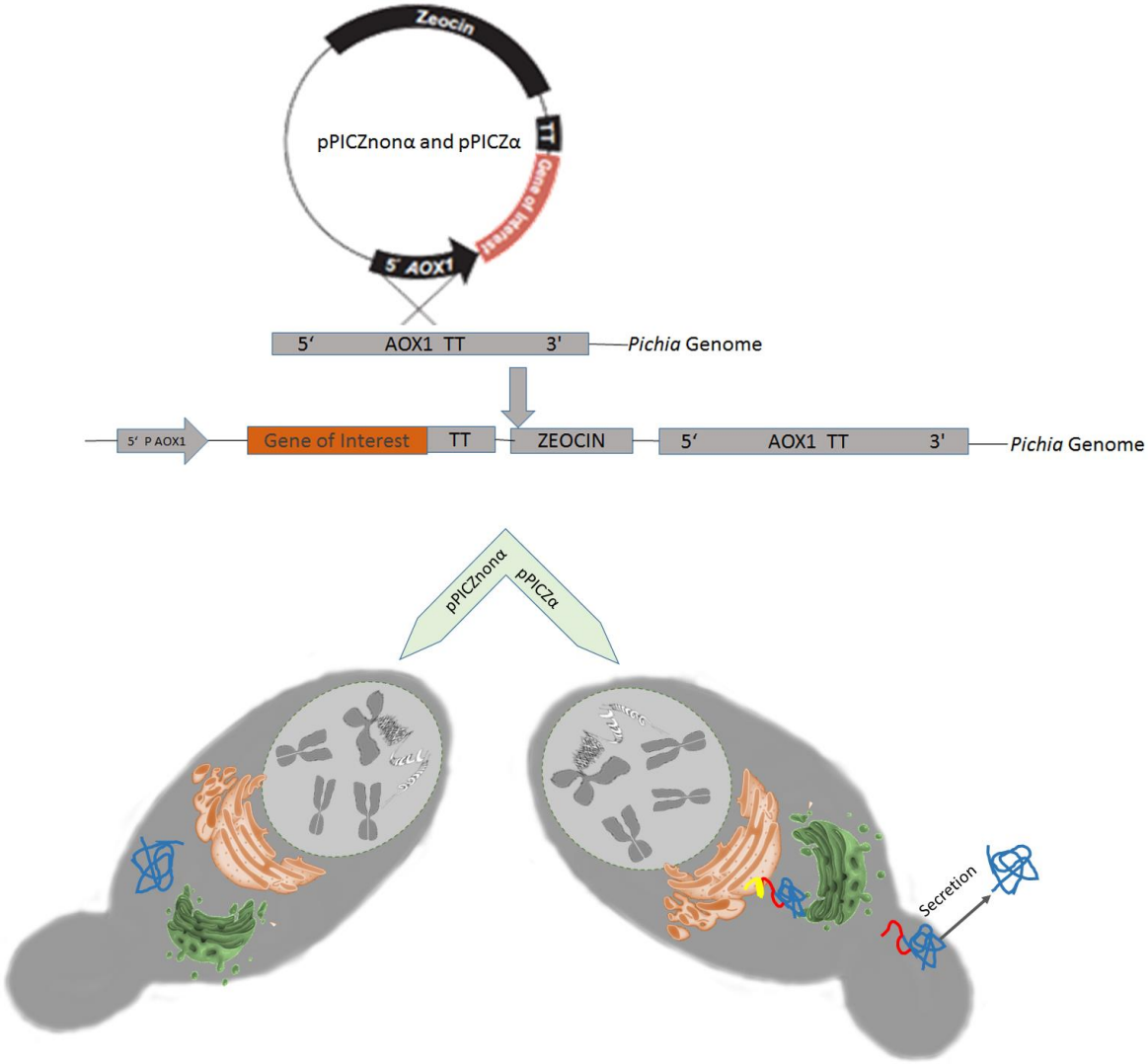

Zeocin
pPICZnonα and pPICZα
Gene of Interest
5' AOX1
5' AOX1 TT 3'
Pichia Genome
5' P AOX1
Gene of Interest
TT
ZEOCIN
5' AOX1 TT 3'
Pichia Genome
pPICZnonα
pPICZα
Secretion

Figure 2. Schematic representation of the yeast intracellular (pPICZnonα vector) and secretory expression (pPICZα vector) pathways.

Integration of the pPICZnon**α** and pPICZα vectors into the *P. pastoris* genome. Linear DNA can integrate into the genome of *P. pastoris* via homologous recombination to generate stable transformants. Gene insertion events at the AOX1 loci arise from a single crossover event between the loci and either of the two AOX1 regions on the pPICZnon**α** or pPICZα vectors (Double headed arrows). This results in the insertion of one or more copies of the vector upstream or downstream of the AOX1. Recombinant protein enters the endoplasmic reticulum (ER). Correctly folded proteins are transported to the Golgi for further processing including additional glycosylation. Proper signal sequences will guide recombinant protein for intracellular or secretory expression (yellow line: pre sequence, red line: pro sequence and blue line: recombinant protein).

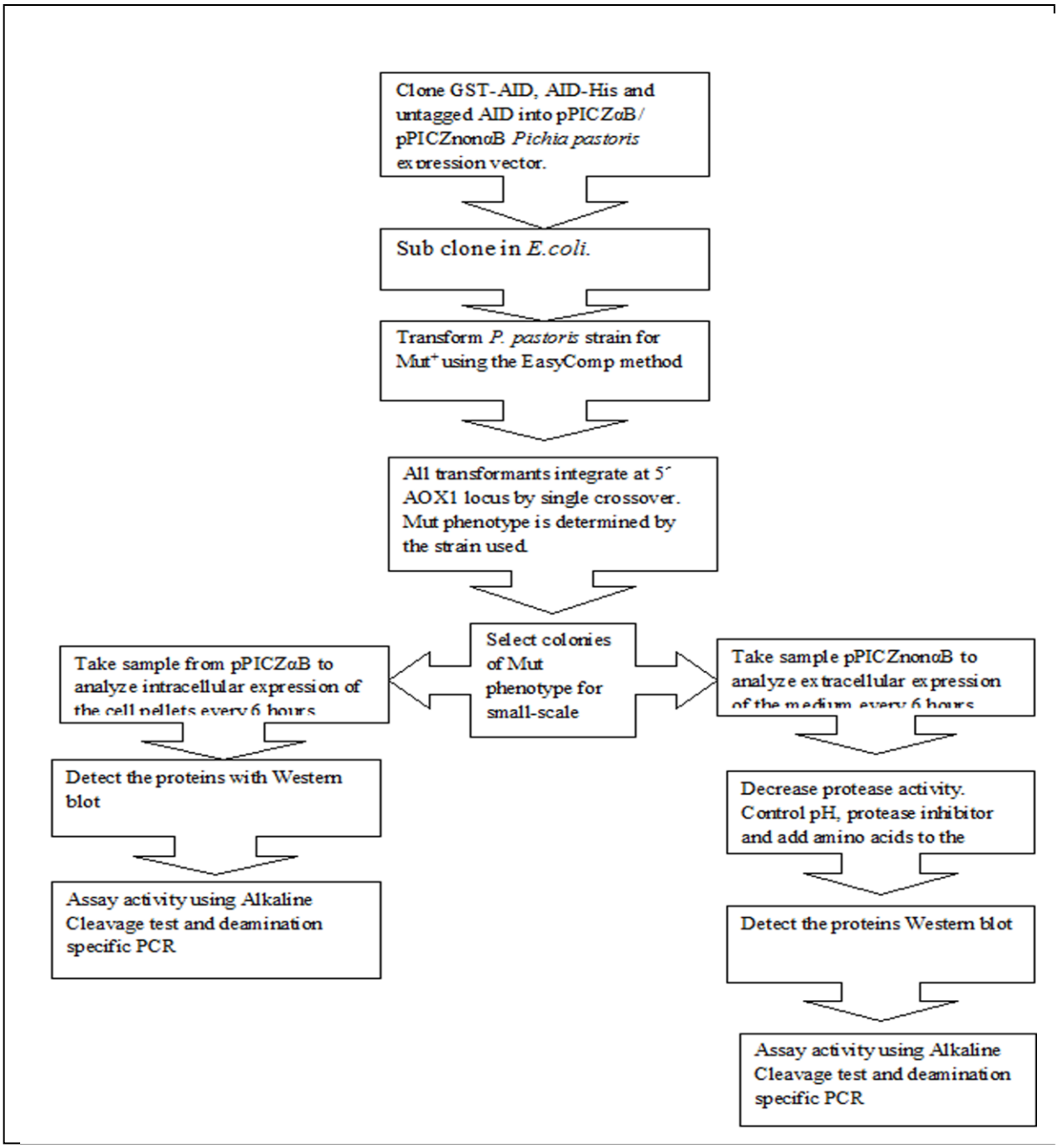
1.4 Rationale and hypotheses

Given the challenges of the bacterial and mammalian expression systems for AID expression, we sought to develop a novel yeast-based expression system for the study of AID. There are 4 basic reasons we believed this system could solve both the genotoxicity as well as purification hurdles that make AID purification difficult: first, yeast is a simple eukaryotic organism but its transcription and protein translation machinery has nearly the same level of sophistication as human cells. Second, compared with mammalian cells, yeast is easier to manipulate genetically (Macauley-Patrick et al., 2005) and can be grown quickly (Shobayashi et al., 2006) with log-phase doubling times almost as low as that of bacterial cultures. Third, yeast has approximately a quarter of the total number of proteins as mammalian cells resulting in fewer non-specific interactions with AID thus facilitating purification. Fourth, yeast is an efficient secretor of proteins, allowing us to develop a novel secretion system. Our first hypothesis was that AID produced in yeast will be at least as robust in terms of activity as those produced in bacteria and mammalian cells. Our second hypothesis was that the secretion system will be immensely beneficial in the case of AID firstly because secreted AID can no longer inflict significant damage to the expression host cell genome, and secondly because yeast media is relatively free of other proteins and AID could potentially be harvested by simple concentration of the media, circumventing the need for protein tags, as is the current gold standard for AID/APOBEC purification. Our third hypothesis was that by studying untagged AID we will discover its true molecular properties which may prove to be different from the more artificially tagged versions studied thus far.

Material and Methods

2.1 Workflow

The overall experimental process used in this study is presented below.



2.2 Strains and plasmids

P. pastoris (Komagataella pastoris) wild type strain (ATCC 28485) was used as a host for expression of fusion and non-fusion proteins and sub-cloning. *E. coli* TOP10 (Invitrogen, USA) and *E. coli* DH5α were used for the plasmid amplification. For this study we used pPICZα (Invitrogen, USA) and pPICZnonαB yeast expression vectors (Fig. 4). The vectors are identical except pPICZα contains an additional *S. cerevisiae* α-factor secretion signal sequence at the beginning of the insert protein gene, encoding for the yeast secretion marker α-factor protein. To eliminate the α-factor and create the pPICZnonαB vector, Site-Directed Mutagenesis method was used. Primers were designed to anneal to the parent plasmid and make an *EcoR* I digestion site (GAATTC) at the beginning of the α-factor sequence. We used *EcoR* I because there was already an *EcoR* I downstream of alpha factor gene, and because it does not cut elsewhere on the plasmid Mutagenesis was performed according to the manufacturer's directions. Briefly, primers were diluted to a concentration of 60 ng/µl and the following reaction mixture was set up: 10 µl of 5× reaction buffer, 5 ng parent plasmid, 2 µl of each primer, 1 µl of 10 mM dNTP, 1 µl of Pfu DNA polymerase (NEB), and sterile water to a final volume of 50 µl. Cycling parameters for the reaction were: 1 cycle of 96 °C for 1 min; 35 cycles of 95 °C for 30 s, 58 °C for 30 s, 72 °C for 4 min, followed by a single cycle of 68 °C for 10 min. After the reaction was complete, 0.75 µl of DpnI was added directly to the reaction mix and incubated at 37 °C for 4 h to destroy the methylated, parental DNA. Finally, 5 µl of the digested reaction mix was transformed into *E. coli* DH5α chemically competent cells and transformants were plated on LB agar plates supplemented with Zeocin. To confirm the site-directed mutagenesis at the α- factor region, several colonies from each construct were sequenced and then digestion was preformed using *EcoR* I to delete full α-factor sequence.

2.3 Construction of GST-AID, AID-V5-His and untagged AID expression vectors

Full sequences and details of vector construction are shown in the appendix of this thesis. We used PCR to amplify AID containing either an N-terminal GST tag, a C-terminal His tag as well as only the native ORF sequence with no tags. To control for the protein expression and purification, we used a GST-only expressing construct that produced a GST-V5-His fragment. We have previously generated this control protein, since it is an inert protein and contains 3 epitopes making it useful as a universal control for expression of any protein that is tagged with GST, His, or V5 epitopes, or any combination. PCR amplicons were cloned into the pPICZα/nonαB expression vectors using GAATTC restriction sites for *EcoR* I. The genes of interests and pPICZα/nonαB were digested with *EcoR* I in the separate reactions. After 1% agarose gel electrophoresis, AID fragments and the pPICZα/nonαB fragments were gel recovered and ligated overnight at 16 ˚C with T4 DNA ligase. The resultant recombinant expression GST-AID, AID-V5-His and untagged AID-pPICZα/nonαB were finally transformed into *E. coli* DH5α and selected on Low Salt LB agar plates containing 25 μg/mL Zeocin at 37 ˚C for overnight. We then sequenced these plasmids and identified multiple correct clones bearing either the GST-AID, AID-V5-His or untagged AID inserts in each vector.

2.4 Transformation of GST-AID, AID-V5-His and untagged AID- pPICZα/nonαB into

P. pastoris

The constructed GST-V5-His, GST-AID, AID-V5-His and untagged AID- pPICZα/nonαB plasmids were linearized by *Sac* I and transformed into *P. pastoris* cell which is capable of integrating this linearized DNA into its genome in a long term stable fashion by chemical

transformation (Invitrogen, USA) according to a previously described procedure. Briefly, *P. pastoris* cells were chemically competent and transformed with 3 μg of linearized plasmids harboring the gene of interests using heat shock and chemical solutions II (PEG solution) and III (salt solution) from EasyCom kit (Invitrogen, USA). Single colonies of the transformants were selected on selective YPDS (1% yeast extract, 2% peptone, 2% dextrose, 1mol/L sorbitol and 2% agar) plates containing 100 μg/mL Zeocin for expression. We identified colonies that had successfully integrated our expression vectors into their genomes. Subsequently, AID expression from the positive colonies was confirmed the expression of our AID proteins using Western blot with either anti GST- , V5- or -AID specific antibodies. To lyse the cells, cultures were centrifuged and yeast cell pellets bearing pPICZnonαB vectorwere lysed using glass beads (425-600 μm, Sigma-Aldrich). An equal volume of glass beads was then added, by estimating volume by displacement. The mixture was then vortexed for 30 seconds, then incubated on ice for 30 seconds. This was repeated 5-7 more times in standard protein lysis buffer. Then, 20-25 μl of cell lysate or 20-25 μl of concentrated cell (bearing pPICZα vector) free medium (5-fold concentrated using a Vivaspin-2 column) were applied to Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS–PAGE).

SDS–PAGE was performed on 10% stacking and running gels. After electrophoresis, the proteins were transfered to Nitrocellulose membrane (Biorad). The membrane was washed with Tris buffer saline Tween 20 (TBST). 5% Skim milk powder in TBST was used as a blocking solution for 1h at room temperature. Western blots were probed with anti-V5 and anti-rabbit polyclonal AID (Abcam) or anti-GST (SantaCruz) antibodies (1/5000 diluted of each), followed by the secondary detection by Goat anti-Rabbit IgG (SantaCruz) in TBST and 5% Skim milk powder. The reaction was stopped by repeated TBST wash after each probing. Color development was carried out by

western blotting detection reagents. The Band Analysis tools of ImageLab software version 4.1 (Bio-Rad) were used to select and determine the background-subtracted density of the bands in all the gels and blots. We used SDS-PAGE as a semi-quantitative estimate of the GST-V5-His piece concentration. We created a standard curve with a protein of known concentration bovine serum albumin (BSA) standard curves to compare against. Quantification was done by densitometry. Different concentration of BSA 0.5, 1, 1.5 and 2 μl of 0.5mg/ml as standard and 15, 17.5, 20 and 25 μL of GST-V5-His protein were used. Gels were stained with Coomassie and the band intensities estimated using ImageLab software version 4.1 (Bio-Rad). A standard curve was created by plotting intensity as a function of mass, from which concentration of the GST-V5-His was found.

Subsequently, known amounts of GST-V5-His (0.02 μg/μl) were used in Western blots and utilized to quantitate the yeast-expressed AIDs. We used this standard protein, since it can be probed with both anti-GST and anti-V5 antibodies and it is thus suitable for quantification of any GST- or V5-tagged protein. For measurement of untagged AID, we first quantified AID-V5-His based on GST-V5-His on another gel with either anti-V5 or GST antibodies. Then untagged AID was measured based on the known AID-V5-His concentration using anti-AID antibody.

2.5 AID expression and purification

The expression vectors pPICZα/nonαB harbor GST-AID, AID-V5-His, untagged AID and GST-V5-His were induced to express AID or GST piece with methanol in yeast. The recombinant *P. pastoris* clone, was inoculated into 10 mL YPD (1% yeast extract, 2% peptone and 2% dextrose) at 30 ˚C, 250 rpm, for about 16 h. Then 1 mL culture supernatants was inoculated into 100mL BMGY medium 1% yeast extract, 2% peptone, 100 mmol/L potassium phosphate, pH 6.0, 1.34%

YNB (yeast nitrogen base without amino acids), 0.005% biotin and 1% glycerol) in 250 mL baffled flasks. The cultures were grown at 30 ˚C, 250 rpm for 24–32 h. Cells grown in BMGY were harvested and re-suspended in 100 mL BMMY (1% yeast extract, 2% peptone, 100 mmol/L potassium phosphate, pH 6.0, 1.34% YNB, 0.005% biotin and 0.5% methanol), then 0.5% final concentration of methanol was added to induce protein expression at 30 ˚C with constant shaking at 300 rpm. To induce the expression of the target protein, 100% methanol was added to a final concentration of 0.5% (v/v) every 24 h due to methanol evaporation. At each of the times: 0, 6, 12, 18 and 24 hours, 1 ml of the expression culture was taken. Cell pellets bearing pPICnonαB vector and medium separately were used to analyze expression levels and determine the optimal time post-induction to harvest. Cells were re-suspended in PBS with 0.2 mM PMSF, 50 μg/ml RNAse and lysed using a French Pressure cell, followed using the French pressure cell press (Thermospectronic) and purification using Glutathione Sepharose high-performance beads (Amersham) as per manufacturer's recommendations. Beads were washed with PBS and were stored in 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM dithiothreitol.

For AID-V5-His cells were resuspended in the lysis buffer 50 mM Phosphate Buffer pH 8.2 and 500 mM NaCl, 0.2 mM PMSF, 50 μg/ml RNAse A and lyse using the French pressure cell press and batch binding on Nickel Sepharose beads (Amersham) as per manufacturer's recommendations. Beads were serially washed with lysis buffer containing increasing concentrations (1 mM and 30 mM) of Imidazole, and eluted in lysis buffer containing 500 mM of Imidazole. Alternatively, washed beads with bound AID-V5-His were stored in AID storage buffer as a source of bead-bound AID. All purifications were subjected to analysis by coomassie staining and Western blot, to verify the relative yield and purity of AID. Western blots were probed with anti-V5 and rabbit polyclonal AID (Abcam) or anti-GST (SantaCruz) antibodies, followed by the

secondary detection by Goat anti-Rabbit IgG (SantaCruz) same as above protocol. Three independently purified and cell lysate for untagged preparations were used in this study.

2.6 preparation of substrates

Single-stranded bubble substrates for AID enzyme assays were prepared as described (Larijani et al., 2007; Larijani et al., 2007). Briefly, DNA oligonucleotides were synthesized and subject to HPLC purification (IDT) and 2.5 nM of the target oligonucleotide was 5'-labeled with [γ-32P] dATP using polynucleotide kinase (New England Biolabs). The labelled oligonucleotide containing the target dC residue was purified through mini-Quick spin DNA columns (Roche) and annealed with 3-fold excess cold strand in a total volume of 50 µl, with incubation starting at 96 °C and slow cooling at 1 °C/min to 6 °C. Triplex bubble structures were prepared by mixing the labeled strand, the cold strand of the same length and shorter complimentary strand at a ratio of 1:3:9. Incubation started at 96 °C and slow cooling at 1 °C/30 sec to 6 °C

2.7 Alkaline cleavage deamination assay for AID activity

This assay has already been described (Sohail et al., 2003; Larijani et al., 2007; Dancyger et al., 2012) and was used to measure and quantify the catalytic activity of AID. Briefly, to determine substrate specificity, each AID was incubated with 2.5 nM radioisotope labelled bubble substrates with the following target sequence: TGC, TAC, AGC, GGC, GAC, or GTC in 100 mM phosphate buffer pH 7.21 at 32 °C for 2 hours. To determine time-point kinetics, 2.5 nM radioisotope TGCbub7 as a labelled substrate was incubated with 12-20 ng AID for time periods 3, 5, 10, 15, 30, 60, 120, 180 min in 100 mM phosphate buffer pH 7.21 at 32 °C and 37 °C. Fmol deaminated

substrate/µg AID was plotted against time incubated. These experiments were used to determine the time of the initial exponential phase of AID activity, in order to proceed to Michaelis-Menten kinetics. Enzyme kinetics were then performed using a range of labelled TGCbub7 substrate concentrations including 7.5, 5, 4, 2.5, 1.25, 0.625, 0.315 and 0.15 nM in a total volume of 10 µL per reaction tube in 100 mM phosphate buffer pH 7.21 incubated with 12-20 ng AID, at 32 °C and 37 °C for 4 hr. Percentage deamination was used to calculate velocity (fmol product/min incubation/µg AID), which was then plotted against substrate concentration. 2–3 independent preparations of each AID were tested in 2–3 individual experiments.

Negative controls were reactions with substrate and no enzymes, substrate with mother vectors, untransformed yeast and GST-V5-His fragment. Positive controls were reactions containing enzymes previously purified and known to be active from bacterial (B-GST-AID) or mammalian systems (M-AID-V5-His). A master mix of 7.8 µL H2O, 2 µL 10X UDG buffer and 0.2 µL UDG enzymes (NEB, USA) was incubated with the previous reaction for 30 minutes to excise the uridine. The total reaction volume is now 20 µl. 2 µl of 2 M NaOH (fresh; 200 mM final) was then added and temperature was increased to 96°C for 10 minutes to cleave the alkali-labile abasic site. Immediately before loading the gel, 8 mL of formaldehyde loading dye were added to each reaction and incubated at 96°C for 5 minutes. Samples were loaded on a 14% polyacrylamide denaturing gel (1X TBE, 25% Formamide, 14% acrylamide:bisacrylamide, 7M Urea) and run at 200-300 V until the dye front travelled approximately halfway down the gel. The resolved gels were exposed to a blanked Kodak Storage Phosphor Screen GP (Bio-Rad) and imaged using a Phosphor Imager (Bio-Rad, Hercules, CA, USA). Quantitation was done using an Image Lab™ 4.1 and Quantity One 1-D Analysis Software (Bio-Rad, Hercules, CA).

2.8 Deamination specific PCR assay

The plasmid construct used as a substrate is pcDNA3.1 V5-6xHIS containing a random target sequence. The plasmid was isolated by maxiprep (Qiagen), followed by purification of the supercoiled fraction through Cesium chloride gradient centrifugation. 50 ng of plasmid was denatured at 98 °C in 100 mM phosphate buffer pH 7.21 for 10 min followed by snap-cooling in an ice bath. 4 µl concentrated supernatant (cell free medium) of AID-V5-His and untagged AID from pPICZα and 4 µl of GST-AID, AID-V5-His and AID (12 ng of GST-AID, 18 ng of AID-V5-His and 20 ng untagged AID from pPICZnonαB were added to the reactions and incubated for 4 h at 32 ˚C. To detect AID-mediated mutations, 1 µl of each reaction was amplified by deamination-specific nested PCR using Taq DNA polymerase and mutation-specific primers, as previously described (Larijani et al., 2005).

We used following primers for this assay. A3D sense fwd 4: 5'-GGG ATA TAG GGG TTT TTT GAG GTT TGG TAT TAT TTA AAT- 3' and A3D sense rev 2: 5'-ACA CAA CCA ACT TTC AAT CCA ACC ACA ACC ACA ACC AAC TTT CAA TA-3'. A3D sense fwd 5: 5'-TTT ATT TTG GTT TTG TGG TAA TTG ATT GTT TGT TAA TAG G-3', A3D sense rev 3: 5'-CCA ACT TTC ATT CCA ACC ACA AAC TTT CAA TAA ATT-3'.

Each reaction contained: 10 µM of the forward and reverse primers (IDT Inc), 10 mM dNTP (Invitrogen), 1 µl deamination reactions, Taq DNA polymerase and 10x PCR buffer (100 mM Tris pH 8.3, 500 mM KCl, 15 mM MgCl2). The nested PCR reactions were 37 cycles of 94 ˚C for 30 seconds, 59 ˚C for 30 seconds, 72 ˚C for 1 min, with A3D sense fwd 4 and rev 3 primers, followed by a second round of PCR using 1 µl of sample from the first round, with primers A3D sense fwd 5 and rev 2 and an annealing temperature 58 ˚C. Some plasmid substrates are mutated by AID at different levels, while others are unmutated. The first type of PCR in the nested PCR has

deamination-specific primers and therefore selectively amplifies AID-mutated DNA. The inner primers used in the nested PCR reaction generate a 451 nt-product. PCR products were analysed on an 1% agarose gel and the 451 nt-long PCR amplicons was subsequently TA-cloned, and 10 amplicons from each reaction were sequenced to confirm AID-mutated mutations C to T, or G to A on the sense and non-sense strands, respectively. This experiment has been done once.

2.9 AID activity data quantification and statistical analysis

Band densitometry was performed as previously described (Abdouni et al., 2013). Briefly, Image Lab™ 4.1 analysis Software (Bio-Rad) was used to measure and compare the densitometry of the deaminated product band and the substrate band. Using these values, the relative amount of deaminated product (as a percentage of the total non-deaminated and deaminated bands) was derived. Background values are automatically determined by Image Lab, and were also manually determined from negative control lanes and subtracted from reaction lanes. For independent confirmation, each individual lane of an alkaline cleavage was also quantified three times using Quantity One 1-D Analysis software (Bio-Rad, Hercules, CA) to ensure the accuracy of data analysis. Typically, each experiment was repeated 2-5 times using at least 3 independently purified preparations of each type of purified AID. The GraphPad Prism software (GraphPad, San Diego, CA, USA) was used to graph and analyze the obtained data. Error bars represent standard deviation. Paired T-test in GraphPad Prism was used to analyze data to calculate the significance of differences in the activities of different versions of AID at different temperatures and time points.

Results

3.1 Demonstrating feasibility of expression and subsequent purification of AID in the yeast system

Fig. 3 shows our overall experimental strategy. Briefly, the expression of any foreign gene in yeast entails 3 principal steps: first, the insertion of the gene of interest into an expression vector and transformation into yeast cells (Fig. 3, top left), second, the selection of strains that express the foreign protein and third, scaling up the culture to obtain enough protein either in cell pellets for intracellular AID (Fig. 3 bottom left), or in media to obtain secreted AID (Fig. 3 bottom right). The sample then proceed to the purification step.

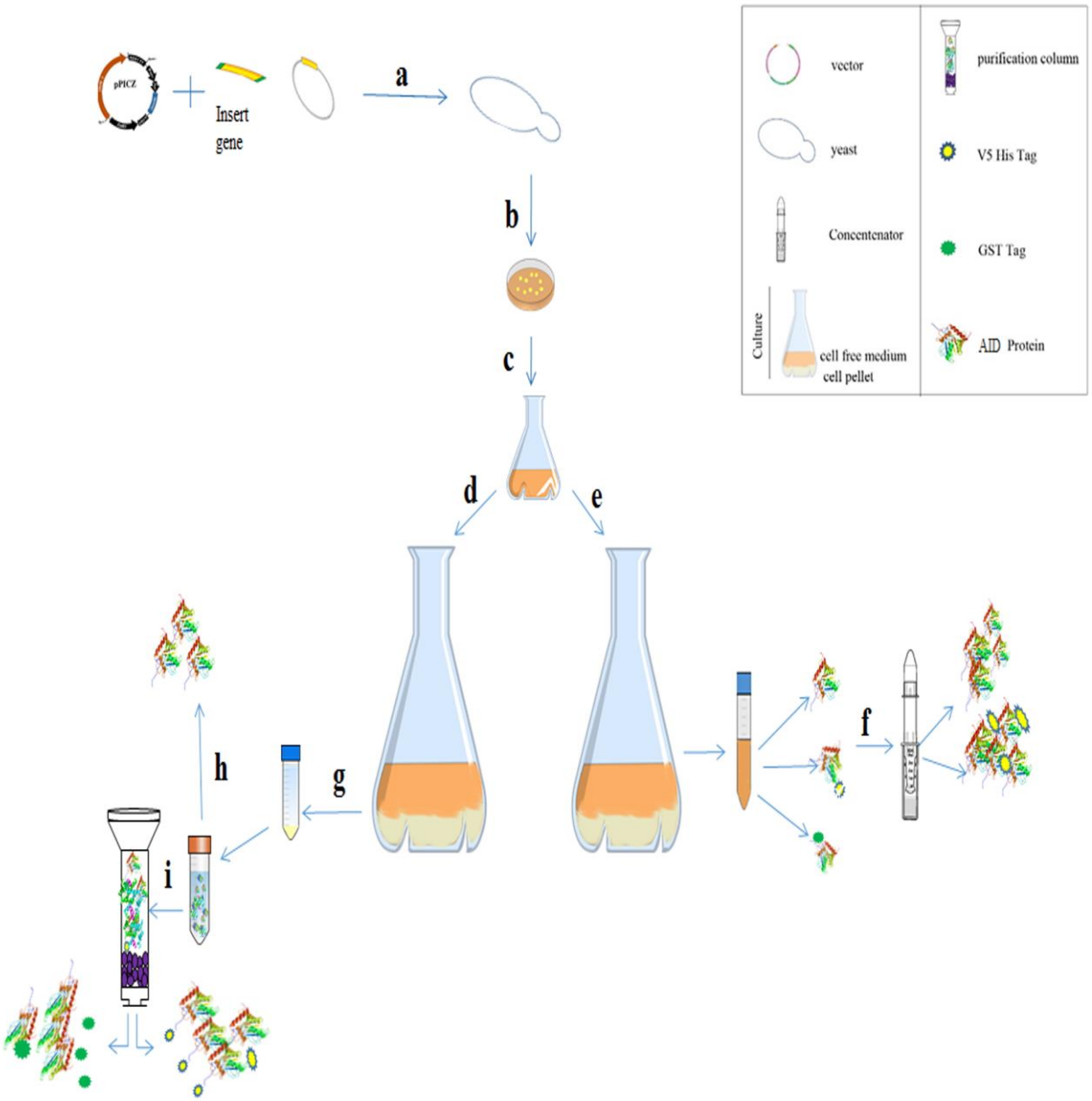
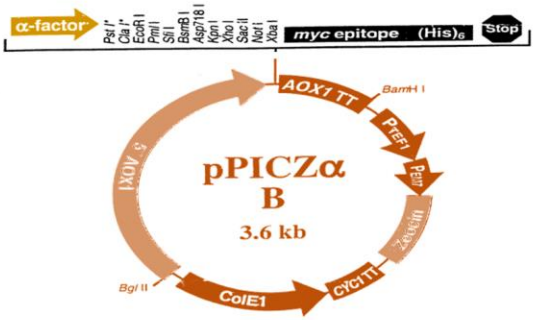

pPICZ
Insert gene
a
b
c
d
e
f
g
h
i
vector
yeast
Concentenator
Culture
cell free medium
cell pellet
purification column
V5 His Tag
GST Tag
AID Protein

Figure 3. Schematic of the key steps in expression and purification of AID in yeast. Recombinant proteins including human GST-AID, AID-V5-His and untagged AID were expressed, with and without the α-factor secretion signal at the N-terminus. The pPICZ vector contains a Zeocin resistance gene for selection in both *E. coli* and *Pichia* (**a** and **b**). Initial colony selection was carried out using Zeocin. Small scale expression trials with methanol induction carried out in 25 mL culture volume (**c**). Expression of GST-AID, AID-V5-His and untagged AID in yeast was scaled up and proteins were purified intra- (**d**) in pPICZnonαB vector and extra-cellular (secreted) in pPICZαB vector (**e**). Cell free medium was concentrated using a Vivaspin-2 column to collect secreted heterologous protein (**f**). Cell pellets were harvested and lysed to detect intracellular GST-AID, AID-V5-His and untagged AID (**g**). Cell pellets were lysed in a French pressure cell press and the supernatant was directly subjected to an activity assay (**i**) or applied to a column of both Glutathione-Sepharose high-performance and Nickel beads (**h**).

3.2. Construction of expression plasmids and selection of the recombinant *P. pastoris* clones

We obtained the expression construct pPICZα vector which is designed for the production of secreted proteins. It contains the α-factor signal which is the most widely used secretion signal in yeast and is in some cases a more efficient signal for secretion the leader sequence of the native heterologous proteins. This 89 amino acid long signal peptide is cleaved in two stages, first in the endoplasmic reticulum and then in the Golgi apparatus, resulting ultimately in secretion of the protein to which it was attached. To be able to generate both secreted as well as intracellular AID, we first modified the vector by removing the secretion signal sequence α-factor. This was achieved using restriction digestion sites flanking this region. To be able to excise the region encoding the α-factor, we introduced these restriction site for *EcoR I* through routine site-directed mutagenesis (Figs. 4a, b and c). We generated the non-secreted pPICZnonαB vector from parental vector pPICZαB. GST-His-V5, AID-V5-His and untagged AID open reading frames were then cloned into both versions of the expression vector, pPICZαB or in pPICZnonαB by routine DNA manipulation (Table 2 and appendix). Thus, we constructed intracellular Y-GST-AID, intracellular Y-AID-V5-His and intracellular Y-AID, which are in the pPICZnonαB vector, and secreted Y-AID and secreted AID-V5-His which are in the pPICZαB vector (Y letter indicates that preps are from yeast).

a



b

α-factor sequence

```
                5' end of AOX1 mRNA                                                    5' AOX1 priming site
811  AACCTTTTTT TTTATCATCA TTATTAGCTT ACTTTCATAA TTGCGACTGG TTCCAATTGA

871  CAAGCTTTTG ATTTTAACGA CTTTTAACGA CAACTTGAGA AGATCAAAAA ACAACTAATT

931  ATTCGAAACG ATG AGA TTT CCT TCA ATT TTT ACT GCT GTT TTA TTC GCA GCA
                Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala

983  TCC TCC GCA TTA GCT GCT CCA GTC AAC ACT ACA ACA GAA GAT GAA ACG GCA
     Ser Ser Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala
                                  α-factor signal sequence
1034 CAA ATT CCG GCT GAA GCT GTC ATC GGT TAC TCA GAT TTA GAA GGG GAT TTC
     Gln Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe

1085 GAT GTT GCT GTT TTG CCA TTT TCC AAC AGC ACA AAT AAG GGG TTA TTG TTT
     Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe
                                                                    Xho I*
1136 ATA AAT ACT ACT ATT GCC AGC ATT GCT GCT AAA GAA GAA GGG GTA TCT CTC
     Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu
          Kex2 signal cleavage          Pst I  EcoR I   Pml I          Sfi I          BsmB I
1187 GAG AAA AGA GAG GCT GAA GC TGCAG GAATTCAC GTGGCCCAG CCGGCCGTC TCGGA
     Glu Lys Arg Glu Ala Glu Ala
                      Ste13 signal cleavage
```

Site-Directed Mutagenesis

c

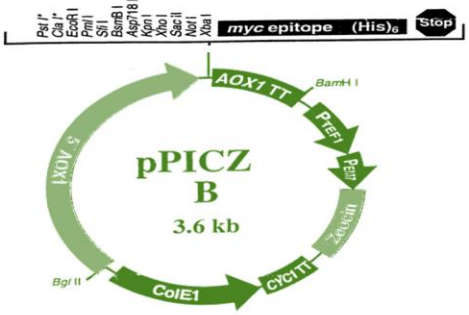
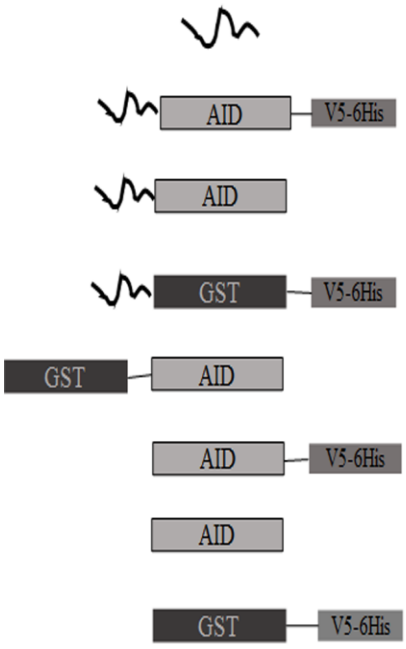


Figure 4. Schematic of representation of yeast secretion vector pPICZαB and yeast intracellular expression vector pPICZnonαB with 5' AOX1 promotor.

(**a**) Schematic of the pPICZαB vector which contains a secretion signal (α-factor), adapted from Invitrogen, USA. (**b**) The α-factor secretion signal sequence for directing secreted expression of the recombinant protein in the pPICZαB vector. The α-factor in this vector is flanked by a single *EcoR* I digestion site. To remove α-factor sequence we created a second *EcoR* I digestion site by standard site directed mutagenesis on the other side of the α-factor-encoding sequencing and were able to generate a new non- α-factor vector through *EcoR* I digestion. (**c**) Schematic of the pPICZnon α-factorB vector.



Plasmid Construct In Yeast	Nucleotide bp	Protein molecular weight kDa	Vector
α-factor-	267	9.4	pPICZαB
α-factor-AID-His	1572	61	pPICZαB
α-factor-Untagged- AID	1438	57	pPICZαB
α-factor-GST-His	1110	41	pPICZαB
GST-V5-AID	1332	52	pPICZnonαB
AID-His	757	29	pPICZnonαB
Untagged- AID	620	24	pPICZnonαB
GST-V5-His	870	32	pPICZnonαB

Table 2. Constructs in pPIC*Zα*/nonαB expression *Pichia* Vector. z

The α-factor-AID-V5-His, α-factor-GST-AID, α-factor-untagged AID- and α-factor-GST-V5-His expression vectors were constructed by cloning cDNA into the pPICZαB vector using the *EcoR* I digestion site. AID-V5-His, GST-AID, untagged AID and GST-V5-His were cloned in pPICZnonαB in the same method. Mother vector pPICZαB expressing α-factor and pPICZnonαB were used as controls.

3.2 Expression and purification of the recombinant AID enzymes

The recombinant plasmids pPICZα/nonαB-AIDs were transformed into *P. pastoris*. Linear DNA can generate stable transformants via homologous recombination. Methanol induced cultures were collected every 6 h. Cell free supernatants were 5-fold concentrated using a Vivaspin-2 column. Cell pellets were extracted into SDS sample buffer using glass beads (425-600 μm, Sigma) by vortexing and interval incubating on ice. All samples were subjected to Western blot using anti-V5 epitope, anti-GST and anti-AID antibodies. We detected a main protein band of 24 kDa for untagged AID and 52 and 29 kDa for AID-V5-His and GST-AID, respectively (Figs. 5a, b and c), which are the expected sizes of each protein.

The expression of the recombinant AID enzymes reached a detectable level at around 18 h. We successfully expressed intracellular Y-GST-AID, intracellular Y-AID-V5-His and intracellular Y-AID also secreted Y-AID and secreted Y-AID-V5-His in *P. pastoris.* Direct quantification of expressed proteins on SDS- PAGEs was impossible because of their low concentration; therefore, we developed a semi-quantified Western blot-based method. In this method, quantification of protein expression was carried out based on a GST-V5-His standard curve for purified Y-GST-AID and Y-AID-V5-His and indirectly based on quantified AID-V5-His for untagged AID (Y-AID). Different concentration of proteins were detected with anti-AID antibody. The GST-V5-His concentration and size standard were first quantified based on a BSA standard curve on coommassie-stained SDS-PAGE (Fig. 5c) . Subsequently, known amounts of GST-V5-His were used in western blots and utilized to quantitate the yeast-expressed AIDs. We used this standard protein, since it can be probed with both anti-GST and anti-V5 antibodies and it is thus suitable for quantification of any GST- or V5-tagged proteins. For measurement of untagged AID (Y-AID), we first quantified Y-AID-V5-His or Y-GST-AID based on GST-V5-His on another gel with either

anti-V5 or GST antibodies. Then untagged AID (Y-AID) was measured based on the known AID-V5-His or GST-AID concentration using anti-AID antibody.The Band Analysis tools of the ImageLab software version 4.1 (Bio-Rad) were used to select and determine the background-subtracted density of the bands in all the gels and blots. We obtained 3, 4.5 and5 ng for AID-V5-His, GST-AID and untagged AID, respectively. Although we observed the expression of untagged intracellular AID in pPICZnonαB using Western blot (Fig. 5d), we could not identify untagged secreted AID from the secretion vector pPICZαB due to its low concentration; however, we are confident that the secreted AID is present in the medium because we were able to detect activity using a deamination PCR assay and an alkaline cleavage assay. This activity was neither observed on extracts of yeast transformed with empty expression vector, nor in extracts of untransformed yeast.

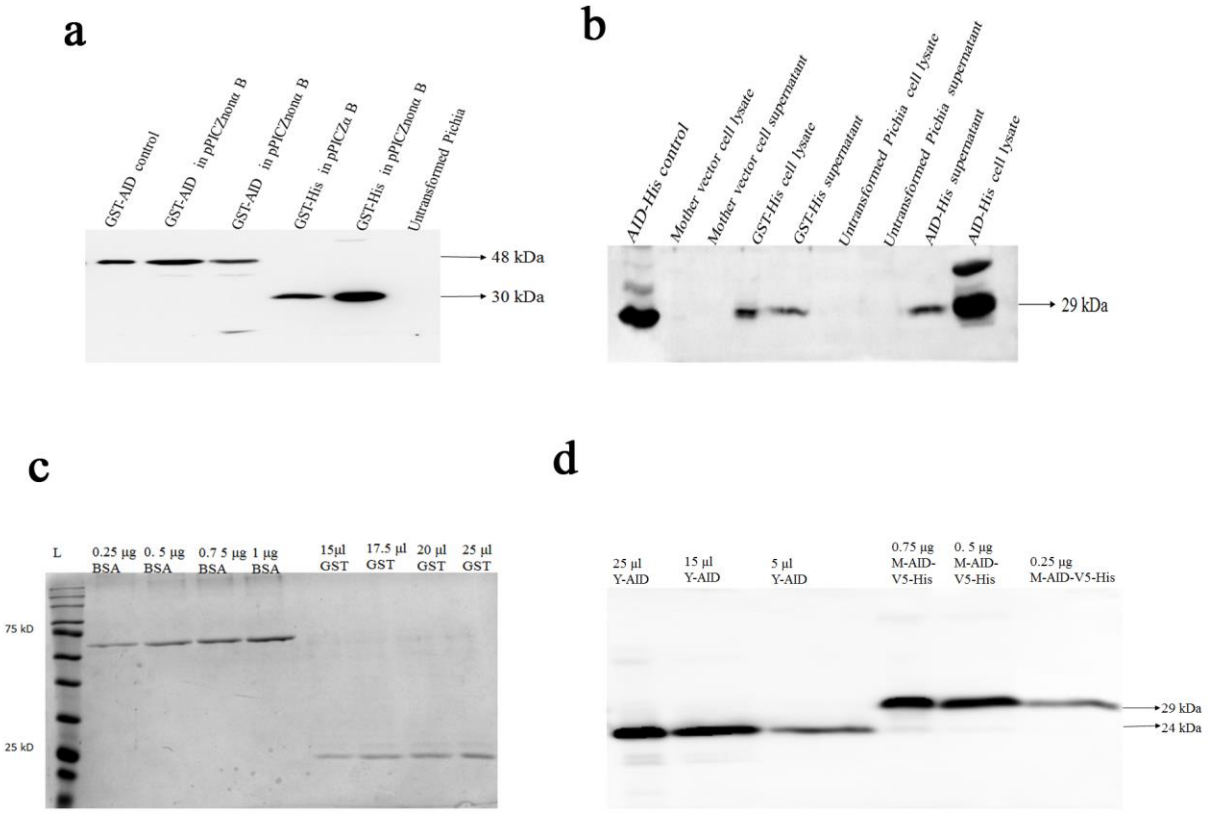

a
GST-AID control
GST-AID in pPICZnonu B
GST-AID in pPICZnonu B
GST-His in pPICZa B
GST-His in pPICZnonu B
Untransformed Pichia
48 kDa
30 kDa
b
AID-His control
Mother vector cell lysate
Mother vector cell supernatant
GST-His cell lysate
GST-His supernatant
Untransformed Pichia cell lysate
Untransformed Pichia supernatant
AID-His supernatant
AID-His cell lysate
29 kDa
c
L
0.25 μg BSA
0. 5 μg BSA
0.7 5 μg BSA
1 μg BSA
15μl GST
17.5 μl GST
20 μl GST
25 μl GST
75 kD
25 kD
d
25 μl Y-AID
15 μl Y-AID
5 μl Y-AID
0.75 μg M-AID-V5-His
0. 5 μg M-AID-V5-His
0.25 μg M-AID-V5-His
29 kDa
24 kDa

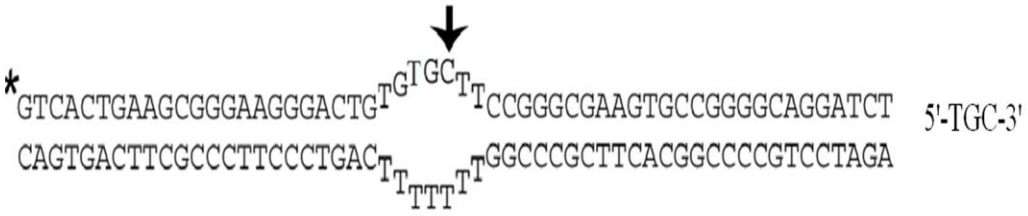
Figure 5. Successful expression and purification of AIDs in yeast detected by Western blot. Yeast cultures were grown with methanol in the media to induce AID expression. Cultures were centrifuged and yeast cell pellets were lysed using glass beads in standard protein lysis buffer and supernatant concentrated by a Vivaspin-2 column, electrophoresed on SDS-PAGE protein gels. Western blotting was carried out and blots were probed with anti-V5 antibodies. (**a**) Representative gel shows successful expression and secretion of AID-V5-His and GST-V5-His at the expected sizes of 48 kDa and 32 kDa respectively. The size is confirmed using a mammalian-purified M-AID-V5-His as size control. Cell lysate and supernatant of untransformed yeast as well as *Pichia* expressed mother (in pPICZα/nonαB) vectors were used as negatives. (**b**) successful purification of intracellular Y-GST-AID and intracellular Y-GST-V5-His as an example, the size was confirmed by running a purified B-GST-AID (bacterial) on the same gel.The negative control a culture that was not transformed with the AIDs expression vector. (**c**) Coomassie stained gel showing purification and expression of GST-V5-His. Coomassie gel also contains varying BSA for measurement of GST-V5-His. (**d**) An example of an indirectly semi-quantification of intracellular Y-AID-V5-His and intracellular untagged AID(Y-AID) by Western blotting. Untagged AID (Y-AID) was indirectly semi-quantified based on Y-AID-V5-His or Y-GST-AID standard curves. Briefly, since GST-V5-His contains 3 different epitopes (GST, V5 and His) it easily can be quantified using anti GST, V5 or His on Western blot. First the concentration of purified GST-V5-His was determined through visualization on SDS-PAGE based on the BSA standard curve. Then, known amounts of GST-V5-His (0.02 μg/μl) were used in western blots to verify the relative concentration of either GST or V5- His AID enzymes using anti V5-epitope or GST antibodies. Next, based on quantified GST-AID or AID-V5-His we estimated the concentration of untagged AID (Y-AID) using anti-AID antibody.

3.3 All versions of AID from yeast are active

Having determined that all versions of yeast-expressed AID were successfully expressed, we tested their cytidine deaminase activity. We applied the standard alkaline cleavage assay for cytidine deamination, which we and others have previously established to measure AID activity on ssDNA (Petersen-Mahrt et al., 2003; Dnacyger et al., 2012; Abdouni et al., 2013; Larijani et al., 2003). We incubated each AID at 37°C with a partially single-stranded bubble substrate containing a 7 nucleotide long ssDNA region with the AID-preferred WRC motif TGC (Fig. 6a). This substrate was chosen since we have previously shown that it is the most preferred substrate of purified AID amongst a panel of hundreds of different ssDNA shapes and sequences.

We found that all versions of our yeast-expressed AID can deaminate dC efficiently (from 5 % to 54%) on TGCbub7, including fusion and non-fusion tags in secretion vector pPICZαB and intracellular expression vector pPICZnonαB from yeast (intracellular Y-AID, intracellular Y-AID-V5-His, intracellular Y-GST-AID, secreted Y-AID and secreted -Y-AID-V5-His). This activity was not observed on any of the negative controls including yeast extracts transformed with empty expression vectors, untransformed yeast or a GST-V5-His fragment (Fig. 6b), providing confidence that the activity observed in the alkaline cleavage assay represents bona fide AID-mediated cytidine deamination. Not only did we show that all AID versions expressed in yeast are active, but also this is the first time that deaminase activity has been detected from untagged AID.

a


*GTCACTGAAGCGGGAAGGGACTGTGTGCTTCCGGGCGAAGTGCCGGGGCAGGATCT
CAGTGACTTCGCCCTTCCCTGACTTTTTTTGGCCCGCTTCACGGCCCCGTCCTAGA
5'-TGC-3'

b

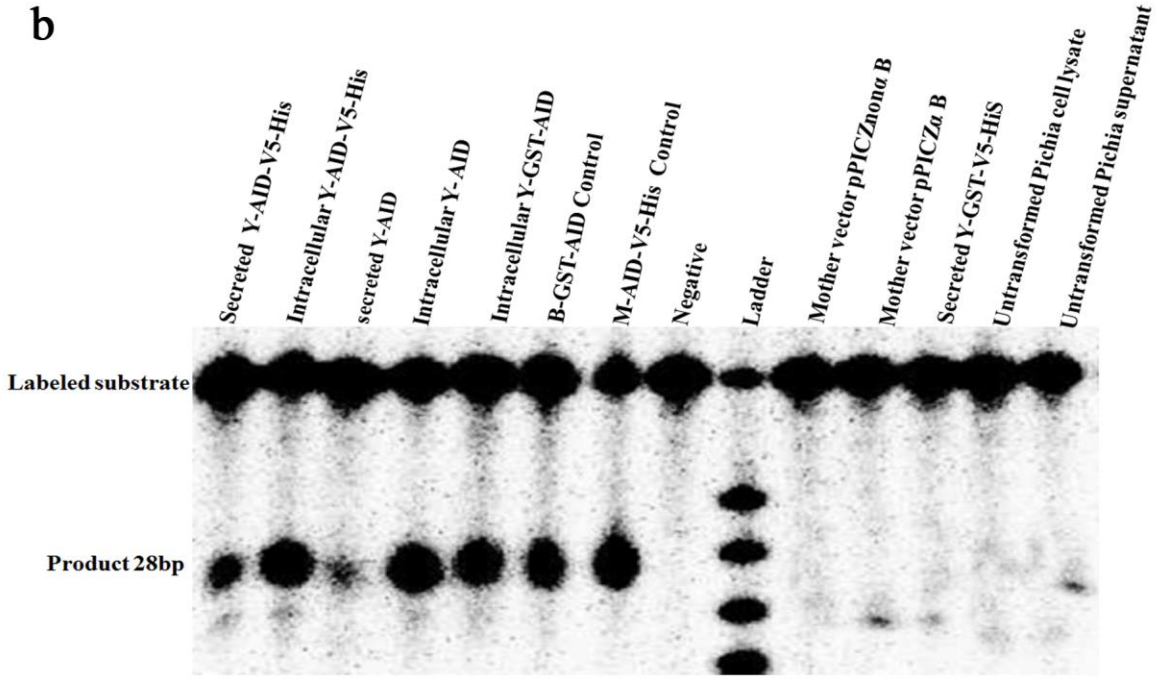

Secreted Y-AID-V5-His
Intracellular Y-AID-V5-His
secreted Y-AID
Intracellular Y- AID
Intracellular Y-GST-AID
B-GST-AID Control
M-AID-V5-His Control
Negative
Ladder
Mother vector pPICZnonα B
Mother vector pPICZα B
Secreted Y-GST-V5-HiS
Untransformed Pichia cell lysate
Untransformed Pichia supernatant
Labeled substrate
Product 28bp

Figure 6. All AID versions are active.

(**a**) Partially single-stranded bubble substrate used in alkaline cleavage deamination assay TGCbub7. This substrate is 56 nt in length and comprises a 7 nt long single-stranded bubble region that has cytidine, the arrow indicates the target cytidine; asterisk indicates the radioactively labeled strand. (**b**) 25 fmol TGCbub7 was incubated with 12-20 ng of intracellular AID and 4 µL concentrated cell free media for secreted AID in 100 mM activity buffer (final volume 10 µL) for 2 hrs at 37 °C. A representative alkaline cleavage gel is shown. Different forms of AID from yeast can efficiently deaminate dC in context of TGC sequence. The presence of a band in all expressing yeast extracts (intracellular Y-AID, intracellular Y-AID-V5-His, intracellular Y-GST-AID, secreted Y-AID and secreted Y-AID-V5-His) indicates enzyme activity. Previously purified mammalian (M-AID-V5-His) and bacterial produced AID (B-GST-AID) were used as positive controls, negative controls included a reaction with no enzyme added to a reaction containing yeast extract from a culture of yeast that was not transformed with AID expressing vectors and cell extractions of yeast expressing empty pPICZα/nonα B vectors as well as the GST-V5-His fragment.

3.4 All version of AID expressed in yeast maintain AID's bona fide sequence preference

AID preferentially deaminates cytidines within WRC (W = A or T, R = A or G) hot-spot motifs (Sohail et al., 2003; Larijani et al., 2005; Dancyger et al., 2012; Larijani et al., 2007; Greisman et al., 2012; Lu et al., 2013). Although AID/APOBEC can mutate any given dC in ssDNA, each AID/APOBEC enzyme has a unique three or four nucleotide-long preferred motif, or "hotspot", based on the -2 and -1 dinucleotide immediately upstream of the target dC. Since these signature trinucleotide hotspots are unique, they serve as the main barcode used to determine which AID/APOBEC family enzyme(s) have acted upon a given cellular or viral genome. For instance, AID's favoured target is the trinucleotide WRC (W=A/T, R=A/G), and these hotspots are the sequence motifs targeted by purified AID *in vitro*, AID exogenously expressed in bacteria or cell lines, as well as *bona fide* AID activity *in vivo* in antibody diversification as well as its off-target genome-wide mutagenesis.

Given the importance of their sequence specificity in their genomic targeting and off-targeting, it is a topic that has been intensely studied, using mutagenesis and chimeric approaches at the structural level. We and others have previously shown that the core catalytic pocket architecture responsible for accommodating and deaminating the dC nucleotide is well conserved within the AID/APBOEC family. The trinucleotide specificity is mediated by specific regions on the surface, proximal to the catalytic pocket, as mediated by a substrate specificity loop, which is highly divergent amongst family members. It has been shown that swapping this loop between various family members can transfer the sequence specificity of the donor, and that differences in the length and amino acid sequence of this substrate specificity loop yield varying surface topologies next to the catalytic pocket, which have been suggested to result in differing preferences for

accommodation of pyrimidine or purine bases at the -2 and -1 positions relative to the target cytidine.

To test whether AID expressed from yeast in different types has a similar sequence specificity to AID purified in mammalian and bacterial systems using different fusion tags, we tested its WRC specificity using our standard alkaline cleavage deamination and deamination specific PCR assays. We tested AID on 6 different oligonucleotide substrates that are identical to TGCbub7 substrate (Fig. 7a) except for the -2 and -1 position nucleotides immediately upstream of the target dc forming either WRC motifs (hot spots; -TGC-, -AGC- and –TAC-) or non-WRC motifs(cold spots; -GGC-, -GTC- and –GAC-) (Fig. 7a).

We found that intracellular Y-AID was almost 3.5-fold more active on WRC motifs (TGC: 54%, TAC: 41%, AGC: 34% than on non-WRC motifs (GGC: 17%, GTC: 12.5%, GAC: 8%). We found that intracellular Y-AID-V5-His was almost 2.7-fold more active on WRC motifs (TGC: 23%, TAC: 32%, AGC: 21%) than on non-WRC sequences (GGC: 15%, GTC: 5.3, GAC: 7.5). Intracellular Y-GST-AID showed on average 28% activity on WRC sequences, while it did not exhibit any activity on non-WRC sequences. Since the secreted AID enzymes including Y-AID and Y-AID-V5-His only deaminated TGC and TAC (5% deamination), we failed to calculate deamination efficiency ratio for them (Fig. 7b). Since WRC specificity is highly sensitive to the conformation of the enzyme's catalytic pocket and surrounding surface regions, the substrate specify confirmed our notion that yeast can properly express and fold the AID protein, and accurately folded near to the pocket and on the surface AID.

a


*GTCACTGAAGCGGGAAGGGACTGTGTGCTTCCGGGCGAAGTGCCGGGGCAGGATCT
CAGTGACTTCGCCCTTCCCTGACTTTTTTTGGCCCGCTTCACGGCCCCGTCCTAGA
5'-TGC-3'
*GTCACTGAAGCGGGAAGGGACTGTGAGCTTCCGGGCGAAGTGCCGGGGCAGGATCT
CAGTGACTTCGCCCTTCCCTGACTTTTTTTGGCCCGCTTCACGGCCCCGTCCTAGA
5'-AGC-3'
*GTCACTGAAGCGGGAAGGGACTGTGTACTTCCGGGCGAAGTGCCGGGGCAGGATCT
CAGTGACTTCGCCCTTCCCTGACTTTTTTTGGCCCGCTTCACGGCCCCGTCCTAGA
5'-TAC-3'
*GTCACTGAAGCGGGAAGGGACTGTGGGCTTCCGGGCGAAGTGCCGGGGCAGGATCT
CAGTGACTTCGCCCTTCCCTGACTTTTTTTGGCCCGCTTCACGGCCCCGTCCTAGA
5'-GGC-3'
*GTCACTGAAGCGGGAAGGGACTGTGGTCTTCCGGGCGAAGTGCCGGGGCAGGATCT
CAGTGACTTCGCCCTTCCCTGACTTTTTTTGGCCCGCTTCACGGCCCCGTCCTAGA
5'-GTC-3'
*GTCACTGAAGCGGGAAGGGACTGTGGACTTCCGGGCGAAGTGCCGGGGCAGGATCT
CAGTGACTTCGCCCTTCCCTGACTTTTTTTGGCCCGCTTCACGGCCCCGTCCTAGA
5'-GAC-3'

b

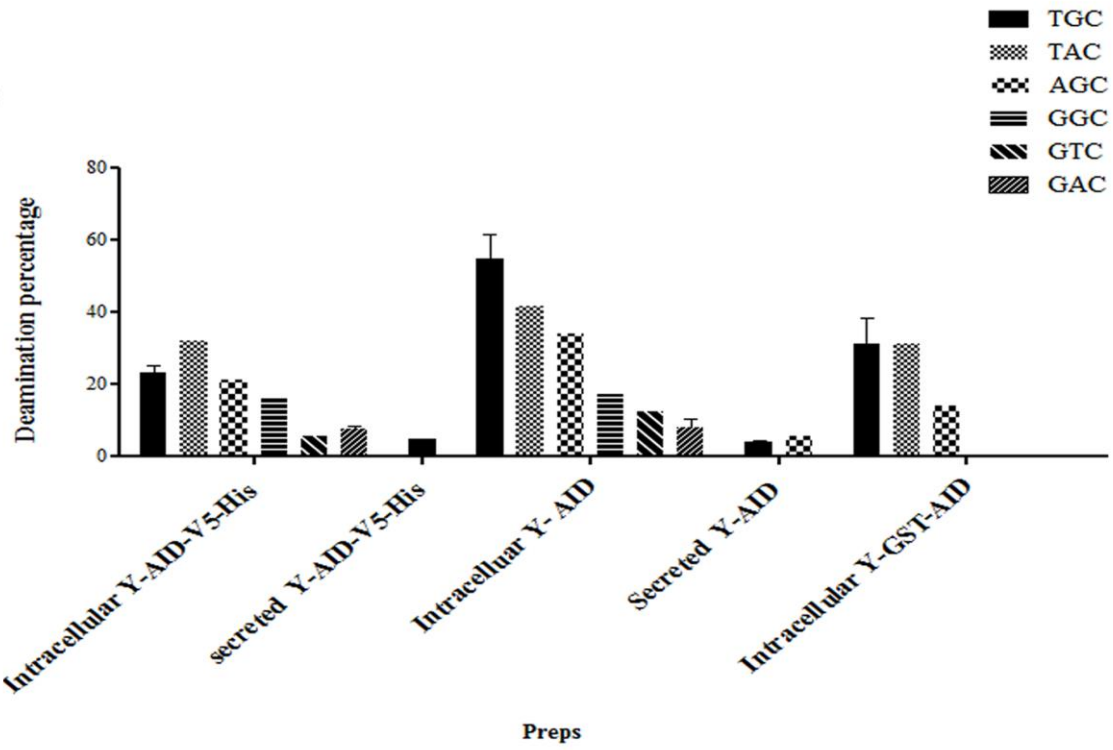

TGC
TAC
AGC
GGC
GTC
GAC
Deamination percentage
80
60
40
20
0
Intracellular Y-AID-V5-His
secreted Y-AID-V5-His
Intracellular Y- AID
Secreted Y-AID
Intracellular Y-GST-AID
Preps

Figure 7. **Enzyme assay and substrate specificity demonstrating AID enzymes produced in yeast exhibit WRC preference.**

We tested the activity of yeast concentrated supernatant Y-AID and Y-AID-V5-His, intracellular Y-GST-AID, intracellular Y-AID-V5-His and intracellular Y-AID using the standard alkaline cleavage assay for measuring cytidine deaminase activity and substrate preference of AID/APOBEC enzymes. (**a**) All substrates used in activity assays contain a single cytidine within the bubble and differ only in the dinucleotide in the -2 and -1 positions immediately upstream of the target cytidine constituting either WRC motifs (hot spots; -TGC-, -AGC- and –TAC-) or non-WRC motifs (cold spots; -GGC-, -GTC- and –GAC-), the arrow indicates the target cytidine; asterisk indicates the radioactively labeled strand. (**b**) The histogram shows substrate specificity of each AID on six bubble substrates (Deamination percentage) with the following target sequence: TGC, TAC, AGC, GGC, GTC, and GAC. Error bars show s.d. from three independent experiments.

3.5 Complete trinucleotide mutability index of different types of AID from yeast

To confirm deaminase activity of AID enzymes we applied the deamination-specific PCR assay which can be used to reliably compare relative amounts of plasmid DNA mutated by purified AID (Larijani et al., 2005; Larijani et al., 2012). In this assay, a target plasmid is made single-stranded by heat denaturation and incubated with AID. Following this incubation, AID-mutated sequences are specifically amplified by PCR using primers that preferentially bind to deaminated plasmid DNA. The target plasmid was boiled, snap-cooled then incubated with AID enzymes. When the target plasmid was incubated with the extract of yeast transformed with the mother vector or untransformed yeast negative controls, no amplification using the deamination-specific primers was observed (Fig. 8a, top panel); therefore, the PCR-amplified bands from the AID incubations can be attributed to specific AID-mediated dC to dU conversion. To determine the preferred sequence context mutated by AID enzymes, AID-treated PCR products were cloned and sequenced. Our analysis specifically focused on the mutability index of trinucleotide motifs (Larijani et al., 2005). Secreted Y-AID-V5-His deaminated cytidine in AAC and GTC motifs at a higher levels than other motifs and while no activity on GCC and TCC was detected (Fig 8b). Intracellular Y-AID-V5-His showed deamination activity on all motifs except for CTC and GTC and less deamination on GCC and GAC, which was almost the same as the secreted Y-AID except it deaminated CTC at a very low level (Figs 8c and d). Intracellular Y-AID preferred TTC, TGC, GAC, TAC, TCC, CGC, GGC and AAC motifs and no activity on GTC, respectively low activity on GCC and CCC was observed (Fig 8e). Intracellular Y-GST-AID did not mutate GGC, GCC, GTC, AGC and CGC, and it deaminated CCC less than others (Fig. 8f).

Overall, cytidine deaminase activity in the alkaline cleavage assay as discussed in the preceding sections is apparent from the presence of a specific cleaved product; however, direct sequencing

of C to T mutations in amplicons obtained in the PCR assay provides undisputable and direct proof that all AID enzymes from yeast are active and can deaminate dC. Moreover, we confirmed that in general, the yeast AID preparations maintained WRC specificity, despite variations in the degree of this preference amongst the different AID forms (Fig. 8b).

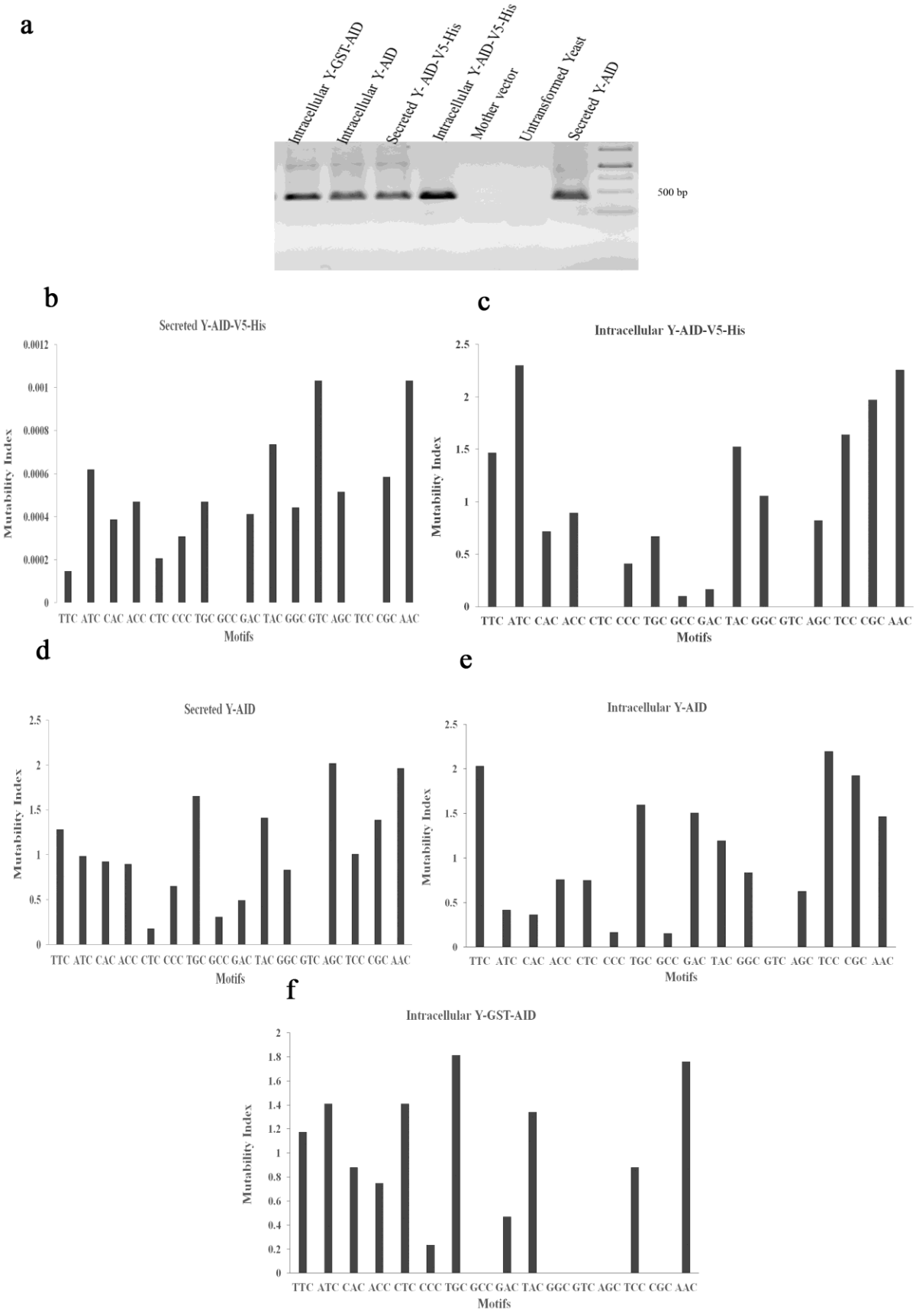

a
Intracellular Y-GST-AID
Intracellular Y-AID
Secreted Y-AID-V5-His
Intracellular Y-AID-V5-His
Mother vector
Untransformed Yeast
Secreted Y-AID
500 bp
b
Secreted Y-AID-V5-His
Mutability Index
Motifs
TTC ATC CAC ACC CTC CCC TGC GCC GAC TAC GGC GTC AGC TCC CGC AAC
c
Intracellular Y-AID-V5-His
Mutability Index
Motifs
TTC ATC CAC ACC CTC CCC TGC GCC GAC TAC GGC GTC AGC TCC CGC AAC
d
Secreted Y-AID
Mutability Index
Motifs
TTC ATC CAC ACC CTC CCC TGC GCC GAC TAC GGC GTC AGC TCC CGC AAC
e
Intracellular Y-AID
Mutability Index
Motifs
TTC ATC CAC ACC CTC CCC TGC GCC GAC TAC GGC GTC AGC TCC CGC AAC
f
Intracellular Y-GST-AID
Mutability Index
Motifs
TTC ATC CAC ACC CTC CCC TGC GCC GAC TAC GGC GTC AGC TCC CGC AAC

Figure 8. Different versions of AID show activity in deamination specific PCR assay.

(**a**) A representative AID deamination specific PCR assay gel for all AID enzymes expressed in yeast. The target denatured plasmid was incubated with AID. The reactions were then subjected to nested PCR with primers specific to the target plasmid with either T or A substituted for C or G, respectively, which selectively amplify AID-mutated DNA. The inner primers used in the nested PCR reaction generate a 451 bp-product. The agarose gels were stained with ethidium bromide and photographed with colors inverted. Trinucleotide mutability indices of purified and concentrated AID from yeast. The y axis shows the mutability index for all 16 NNC motifs. Relative preference for each individual NNC sequence was obtained by dividing its mutation rate by the average value for all 16 NNCs. The x axis shows NNC DNA motifs. (**b**) Secreted Y-AID-V5-His (**c**) Intracellular Y-AID-V5-His (**d**) secreted Y-AID (**e**) Intracellular Y-AID (**f**) Intracellular Y-GST-AID. This experiment has been done only once.

3.6 Comparison of the catalytic kinetics of different types of AID from yeast

We investigated the catalytic rates of AID from yeast by comparing Michaelis-Menten enzyme velocities. In order to determine the optimal duration of incubation with substrate for each AID enzyme to measure initial catalytic velocities, we first performed time course enzyme kinetics comparisons. TGCbub7 was incubated for various time points (3, 5, 10, 15, 30, 60, 120 and 180 min) with 12-20 ng of AID at 32 ˚C and 37 ˚C (Fig. 9a). We chose these temperatures because we previously showed that the optimal temperature for human AID is 32 ˚C (Quinlan et al., 2017). We used TGCbub7 as a substrate because it has been found to be AID's most favored substrate amongst an extensive panel of substrates of varying structures and sequences (Larijani et al., 2007).

Deamination kinetics analysis revealed that all versions of AID from yeast exhibited enzymatic activity with detectable product at 3 min at both temperatures. The product formation was increased to 5.8-fold for intracellular Y-AID (Fig. 9b), 3.8-fold for intracellular Y-AID-V5-His (Fig. 9c) and 3.3-fold for intracellular Y-GST-AID (Fig. 9d) at 60 min. The product formation was saturated after 60 min for all samples at 32 ˚C and 37 ˚C.

Next we evaluated the initial deamination enzyme velocity of yeast-expressed intracellular AID enzymes. 12 ng equivalent of Y-GST-AID, 20 ng equivalent of Y-AID and 12 ng of Y-AID-V5-His were incubated with the substrate (TGCbub7) concentrations ranging from 0.15 to 7.5 nM at 32 ˚C and 37 ˚C (Fig. 10a). The velocity of product formation was determined by calculating the amount of product formed in a unit of time by a given amount of AID and plotted against substrate concentration. As shown in (Figs. 10b, c and d), the maximal activity for all AIDs was higher at 32 ˚C than at 37 ˚C. The velocity of deamination for intracellular Y-GST-AID and Y-AID were almost 2.5-fold, and for Y- AID was 1.5-fold at 32 ˚C which were higher than 37 ˚C (Figs. 10b, c and d). Y-AID shows significantly more activity at 32 ˚C and 37 ˚C compared with Y-GST-AID

and Y-AID-V5-His (Figs. 10e and f). It was not possible to estimate the Km since substrate saturation kinetics were not observed. However, as the velocity continues to increase beyond the highest substrate concentration tested, the Km is likely to be higher. In conclusion, measuring the deamination kinetics of GST-AID, AID-V5-His and untagged AID from the yeast expression system revealed that the optimal temperature for maximal enzyme velocity is 32 °C which is consistent with AID purified from bacteria and human cell lines. Moreover, untagged AID (Y-AID) exhibited the highest initial velocity compared with Y-GST-AID and Y-AID-His. The results confirmed our hypothesis that folded tags (GST) or small tags (V5-His) placed at either the N- or C- termini impact AID activity.

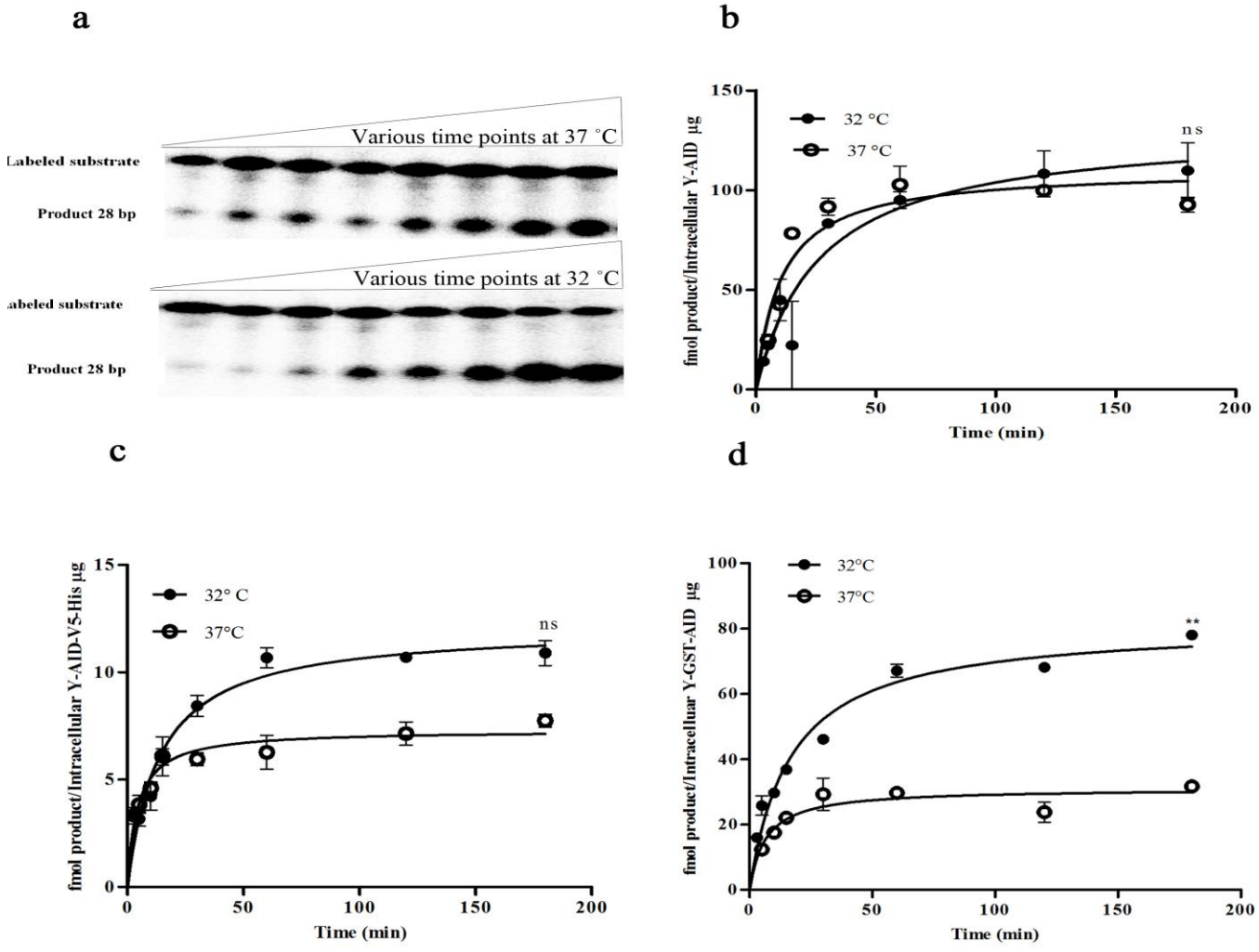

a
Various time points at 37 °C
Labeled substrate
Product 28 bp
Various time points at 32 °C
abeled substrate
Product 28 bp
b
32 °C
37 °C
ns
fmol product/Intracellular Y-AID µg
Time (min)
c
32° C
37°C
ns
fmol product/Intracellular Y-AID-V5-His µg
Time (min)
d
32°C
37°C
**
fmol product/Intracelluar Y-GST-AID µg
Time (min)

Figure 9. Deamination kinetics comparing the activities of intracellular Y-GST-AID, Y-AID-V5-His and Y-AID. 25 fmol substrate was incubated with AID for various times at 32 ˚C and 37 ˚C. (**a**) A representative alkaline cleavage gel of product formation for 12-20 ng AID is shown as a function of incubation time. Comparison of activity kinetics on TGCbub7 for intracellular Y-AID (**b**), intracellular Y-AID-V5-His. P values for 32 ˚C *vs.* 37 ˚C were > 0.05 and not significant for Y-AID and Y-AID-V5-His (**c**) and intracellular Y- GST-AID. P value for 32 ˚C *vs.* 37 ˚C was 0.004 and difference was significant. (**d**). Error bars show s.d. from three independent experiments.

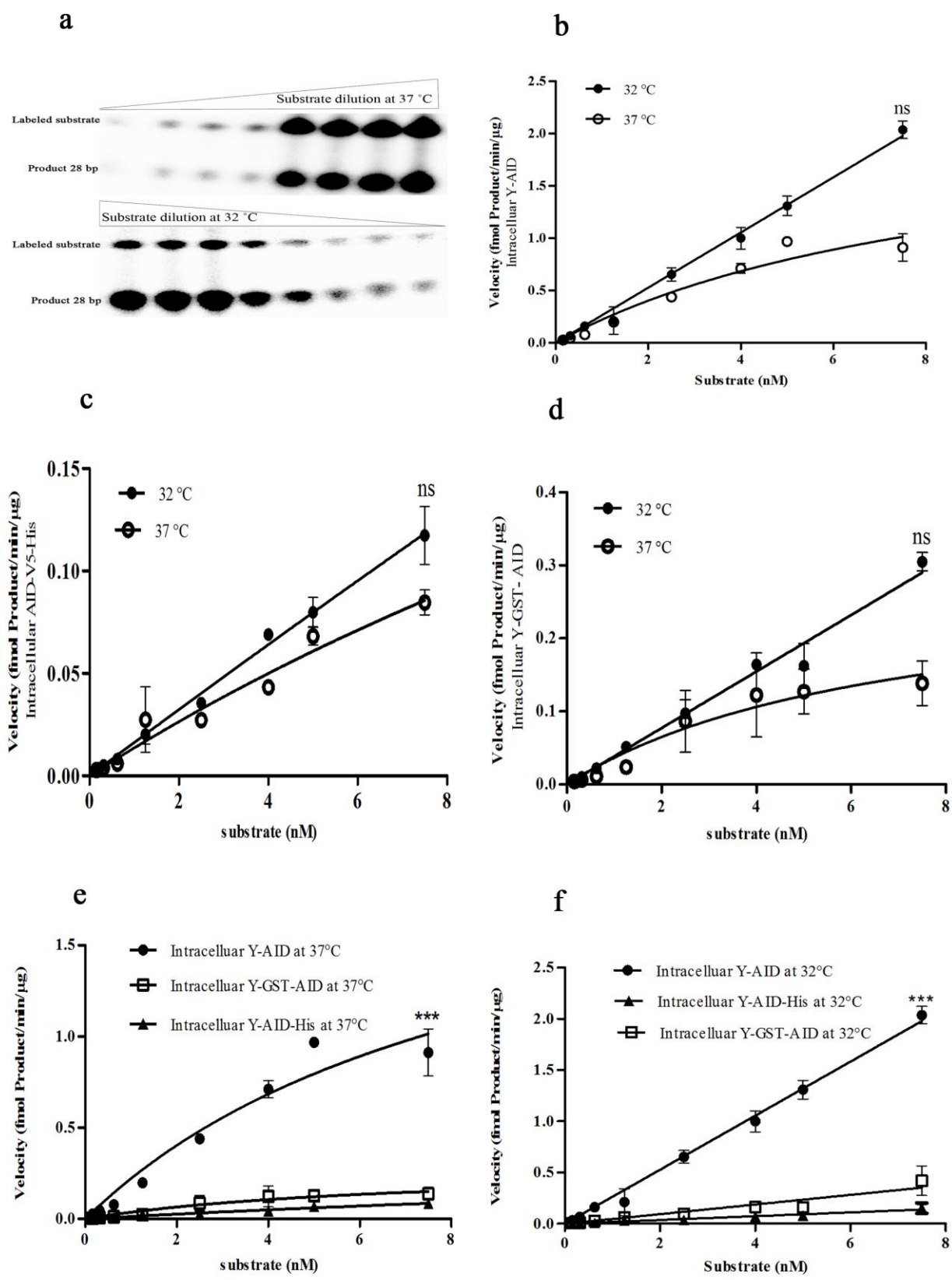

a
Substrate dilution at 37 °C
Labeled substrate
Product 28 bp
Substrate dilution at 32 °C
Labeled substrate
Product 28 bp
b
32 °C
37 °C
ns
Velocity (fmol Product/min/µg)
Intracelluar Y-AID
Substrate (nM)
c
32 °C
37 °C
ns
Velocity (fmol Product/min/µg)
Intracellular AID-V5-His
substrate (nM)
d
32 °C
37 °C
ns
Velocity (fmol Product/min/µg)
Intracelluar Y-GST- AID
substrate (nM)
e
Intracelluar Y-AID at 37°C
Intracelluar Y-GST-AID at 37°C
Intracelluar Y-AID-His at 37°C
***
Velocity (fmol Product/min/µg)
substrate (nM)
f
Intracelluar Y-AID at 32°C
Intracelluar Y-AID-His at 32°C
Intracelluar Y-GST-AID at 32°C
***
Velocity (fmol Product/min/µg)
Substrate (nM)

Figure 10. Deamination kinetics comparing the activities of Intracellular Y-GST-AID, Y-AID-V5-His and Y-AID at 32 ˚C and 37 ˚C. (**a**) Examples of the representative deamnation assay gels used to measure activity of AID from yeast on serial dilutions of TGCbub7. Assay was performed using defined concentrations of GST-AID, AID-V5-His and untagged AID or GST-AID and serial dilutions of the substrate starting with 0.15 nM. The product and substrate bands were quantified and used together with the incubation time and the amount of AID to calculate the reaction velocity at 32 ˚C and 37 ˚C (graph). Velocity is defined as the amount of deaminated product generated by a given amount of AID in a unit of time and plotted against substrate concentration. Deamination kinetics showing the activity of intracellular Y-AID (**b**), intracellular Y-AID-V5-His (**c**) and intracellular Y- GST-AID (**d**) from yeast on dilutions (0.15 to 7.5 nM) of TGCbub7 substrates at 32 ˚C and 37 ˚C. P values for all were > 0.05 and not significant. Deamination kinetics of Y-AID, Y-GST-AID and Y-AID-V5-His were compared each at 37 ˚C P value <0.0001 (**e**) Deamination kinetics of Y-AID, Y-GST-AID and Y-AID-V5-His were compared each at 32 ˚C P value <0.0001 (**f**). Error bars show s.d. from three independent experiments.

Discussion

AID/APOBECs are cytosine deaminase enzymes involved in boosting immune response and contributing to mutation in cancer. Yet, many aspects of their structure and regulation remained unknown. For instance, although each AID/APOBEC enzyme can mutate genome-wide, they exhibit varying levels of genotoxicity and seem to be targeted to different loci, and the mechanism of this differential regulation is as yet not understood. Over the last few decades, many studies have focused on finding ways to express these toxic enzymes in variety of organisms that are quite different from the source organism. This would minimize the potential damage of the toxic enzymes on the host cells during expression as well as decrease interactions with other cellular proteins that may hinder purification. To date the main hosts used for the production of recombinant AID/APOBECs proteins are bacteria, insect cells, and mammalian cells. However, the expression and purification performed using the systems are labour-intensive and time-consuming (Larijani et al., 2012; Robbiani et al., 2009; Cocker et al., 2016).

To simplify the expression, solubilisation, and purification of recombinant proteins, the addition of a fusion tag at either the N- or C-terminus of recombinant proteins has been a commonly-used strategy. Due to the aforementioned inherent difficulties with AID/APOBEC purification, these fusion tags serve the dual purpose of stabilizing protein folding, as well as allowing for purification. On the other hand, it has been reported that both N and C termini of AID and other APOBECs are required for specific aspects of its enzymatic function such as homodimer formation and cellular localization; thus any additional amino acids from a fusion tag in either C or N-terminal might exert an artificial influence on AID function and structure (Bransteitter et al., 2009; King et al., 2015 and 2017; Zahn et al., 2014 ; Mondal et al., 2016).

While it is possible to cleave fusion tags from the native protein during or after column purification and a number of studies used the cleavage of the large tags to achieve untagged A3A, AID and A3G (Pham et al., 2016; Chelico et al., 2006; Pham et al., 2013), the cleavage may impose additional time, causing lower yield and enzyme inactivation. Additionally, tag cleaving is performed after expression and purification which is relatively after protein folding meaning tag might have imposed its effects on folding specially for big tags such as GST tag. Due to the tendency of AID/APOBECs to aggregate (Chelico et al., 2006), this cleavage is expected to lower solubility. In our experience, cleavage of the GST tag from the bacterially expressed GST-AID results in AID aggregation and inactivation (Larijani et al., 2012). Thus, the use of expression tags followed by tag cleavage is not an optimal method of obtaining a truly native protein, and a much better approach would be to express a protein that is native from the beginning.

Since yeast can be made to secrete the protein of interest into a medium which is relatively clean and devoid of other cellular debris, this approach can be used to generate untagged proteins, rather than necessitating the use of purification tag followed by tag removal.

The Methylotrophic yeast, *P. pastoris*, has been developed as a useful host for the high level expression of heterologous enzymes and proteins under the AOX1 gene promoter which is tightly regulated by methanol. *P. pastoris* has some advantages over prokaryotic cells, especially in regards to post-transcriptional and post-translational modifications; also the *P. pastoris* allows secretion of the recombinant protein. Secreting proteins is a desirable method for production of especially toxic enzymes, as it has less effect on cells and requires simpler purification from the secreted media, rather than separating the protein from the intracellular matrix that contains a large number of native proteins. *P. pastoris* is an attractive production host as it can be cultivated cheaply and rapidly on simple protein free medium, unlike mammalian cell lines. Over 500 proteins, from

industrial to biopharmaceuticals, have been expressed in *P. pastoris* (Macauley-Patrick et al., 2005).

Despite the above-mentioned yeast advantages, the expression of heterologous proteins in *P. pastoris*, posed a number of difficulties. Expression of a protein in a yeast system could result in a product with slightly different posttranslational modifications compared with the native enzyme (Hamilton et al., 2015); however, this could also be the case when enzymes are expressed in other eukaryotic cells such as insect cells. Another point which has to be taken into consideration is low yield of secreted protein since inefficient removal of the signal peptide may result in protein aggregation and retention within incorrect compartments, such as the endoplasmic reticulum (Lin-Cereghino et al., 2013; Kurjan et al., 1982). Moreover, this might initiate other processes such as a type of feedback mechanism called repression under secretion stress (RESS) that is activated in response to an impairment of protein folding and secretion (Lambertz et al., 2013). Furthermore, it is possible that secreted proteins are susceptible to cleavage by secreted proteases in the media, and it is also possible that other metabolic by-products or factors present in the media, or even the changing pH of the media due to natural metabolic processes can all influence the secreted protein (Macauley-Patrick et al., 2005). These can be mitigated by adjusting media pH, or addition of protease inhibitors to the media. Additionally, The AOX1 promoter which is highly regulated by methanol can lead to several disadvantages such as increased need of oxygen (Krainer et al., 2011) or production of hydrogen peroxide (H_2O_2) which can become a damaging reactive oxygen species (Kern et al., 2007). Hydrogen peroxide is an oxidizer and can degrade secreted AID/APOBECs carrying Zn.

Given their relatively small genome and eukaryotic protein translation machinery;yeast make a versatile eukaryotic model to probe the diverse effects of AID/APOBECs on the human the

genome. Previously, AID/APOBEC-expressing yeast has been used to model transcription-dependent cytosine deamination, various aspects of genome targeting in cancer or retro-element transposition inhibition by AID/APOBEC proteins (Lada et al., 2015; Poltoratsky et al., 2004; Schumacher et al., 2005); Here we described,) and the first expression and purification scheme of AID/APOBECs from yeast.;. Using deamination-specific PCR, we were able to measure the number of deaminations in WRC and non-WRC motifs. In contrast to our previously established alkaline cleavage assay, which shows a clear preference for WRC motifs, we found the deamination-specific PCR illustrated a broadened targeting specificity. This difference is likely due to the fact that the two assays are vastly different; whilst the former examines AID activity on a long stretch of plasmid DNA with any type of secondary structure, the latter measures AID activity on a well-defined bubble substrate with a single target cytidine.

Conclusion

In this study we demonstrated that GST-AID and AID-V5-His can be produced in yeast and will be as robust as those produced in mammalian cells. The secretion system will allow us to gain active AID/APOBEC proteins by simple concentration of the media without using purification tags. Also construction and expression of untagged AID/APOBECs provides a novel platform to study the true native molecular properties of these proteins in the absence of added fusion tags

Future directions

This system needs to be better optimized in terms of gaining higher yields and less degraded proteins especially for the secretion system. To overcome protein degradation it may prove advantageous to use protease-deficient strains. It is also possible to decrease proteolytic degradation by optimizing pH and adding amino acid supplements like casamino acids to the culture medium. These supplements can enhance protein stability by acting as alternative and competing substrates for one or more proteases, and also can repress protease induction caused by nitrogen limitation. We will supplement the media with a protease inhibitor cocktail to ensure we obstruct as much of the potential proteolysis as possible. Also, we will fully purify the untagged secreted AID more efficiently using Fast Protein Liquid Chromatography (FPLC), rather than enrichment by simple concentration of media as we presented in this thesis. Given the encouraging data presented in this thesis, we propose that other APOBECs can also be successfully expressed and purified in yeast allowing for more accurate structure:function analysis.

The system can also be used to express candidate regulatory proteins in conjunction with AID/APOBECs and evaluate their interactions and/or effects on AID/APOBEC activity. For instance, yeast was the first model to study RNA exome functions and it is now clear that the RNA exosome plays a role in targeting AID to CSR. However, the mechanistic details of this association, are not yet understood. The study of AID and molecular interactions facilitated by yeast could make important contributions to paving a clearer and better path towards developing drugs and other strategies to manipulate the activity of AID/APOBECs.

References

1. Abdouni, H., King, J.J., Suliman, M., Quinlan, M., Fifield, H., Larijani, M. (2013) Zebrafish AID is capable of deaminating methylated deoxycytidines. *Nucleic Acids Res* **41(10)**:5457-5468.

2. Abdouni, H., King, J.J., Ghorbani, A., Fifield, F., Berghuis, L., Larijani, M. (2017) RNA enhances the activity of purified AID on DNA sequences of Immunoglobulin switch regions. *In press*

3. Ahmad, M., Hirz, M., Pichler, H., Schwab, H. (2014) Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol* **98**:5301–17.

4. Aricescu, A.R., Lu, W., Jones, E.Y. (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol* **62**:1243–1250.

5. Baron, M., Reynes, J.P., Stassi, D., and Tiraby, G. (1992) A Selectable Bifunctional b-Galactosidase: Phleomycinresistance Fusion Protein as a Potential Marker for Eukaryotic Cells. *Gene* **114**:239-243.

6. Barreto, V., Pan-Hammarstrom, Q., Zhao, Y., Hammarstrom, L., Misulovin, Z. and Nussenzweig, M. (2005) AID from bony fish catalyzes class switch recombination. *J Exp Med* **202(6)**:733-738.

7. Bransteitter, R., Prochnow, C., Chen, X.S. (2009) The current structural and functional understanding of APOBEC deaminases. *Cell Mol Life Sci* **66**:3137-3147.

8. Chelico, L., Pham, P., Calabrese, P. and Goodman, M.F. (2006) APOBEC3 G DNA deaminase acts processively 3' −> 5' on single-stranded DNA. *Nat Struct Mol Biol* **13**: 392–399.

9. Conticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K., Neuberger, M.S. (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. *Mol Biol Evol* **22:**367-377.

10. Conticello, S.G. (2008) The AID/APOBEC family of nucleic acid mutators. *Genome Biol* ***9***:229.

11. Cos, O., Ramon, R., Montesinos, J.L., Valero, F. (2006) Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast Pichia pastoris under different promoters: A review. *Microbial cell factories* **5:**17.

12. Cregg, J.M., Vedvick, T.S., Raschke, W.C. (1993) Recent Advances in the Expression of Foreign Genes in Pichia pastoris. *Bio/Technology* **11:** 905-910.

13. Dancyger, A., King, J., Quinlan, M., Fifield, H., Tucker, S., Saunders, H., Berru, M., Magor , B., Martin, A., Larijani, M. (2012) Differences in the enzymatic efficiency of bony fish and human AID are mediated by a single residue in the C-terminus that modulates singlestranded DNA binding. *FASEB J* **26(4):**1517-25.

14. Dooley, H., Stanfield, R. L., Brady, R. A., M. F. Flajnik. (2006) First molecular and biochemical analysis of in vivo affinity maturation in an ectothermic vertebrate. *Proc Natl Acad Sci USA* **103:**1846 –1851.

15. Drocourt, D., Calmels, T.P.G., Reynes, J. P., Baron, M., Tiraby, G. (1990) Cassettes of the Streptoalloteichus hindustanus ble Gene for Transformation of Lower and Higher Eukaryotes to Phleomycin Resistance. *Nucleic Acids Res* 1**8:**4009.

16. Durandy, A., Peron, S., Taubenheim, N. and Fischer, A. (2006) Activation-induced cytidine deaminase: structure-function relationship as based on the study of mutants. *Hum Mutat* **27:**1185- 1191.

17. Graf, A., Dragosits, M., Gasser, B., Mattanovich, D. (2009) Yeast systems biotechnology for the production of heterologous proteins. *FEMS Yeast Res* **9(3):**335–48.

18. Greisman, HA., Lu, Z., Tsai, AG., Greiner, TC., Yi, HS., Lieber MR. (2012) IgH partner breakpoint sequences provide evidence that AID initiates t(11;14) and t(8;14) chromosomal breaks in mantle cell and Burkitt lymphomas. *Blood* **120(14)**.

19. Hamilton, SR., Zha D. (2015) Progress in yeast Glycosylation Engineering. *Methods Mol Bio* **1321:**73-90.

20. Harris, RS. (2008) Enhancing immunity to HIV through APOBEC. *Nat Biotechnol* **26:** 1089-1090.

21. Harris, RS., Sheehy, AM., Craig, HM., Malim, MH., Neuberger, MS. (2003) DNA deamination: not just a trigger for antibody diversification but also a mechanism for defense against retroviruses. *Nat Immunol* **4:**641-643.

22. Honjo, T., Muramatsu, M., Fagarasan, S. (2004) AID: how does it aid antibody diversity? *Immunity* **20:**659-668.

23. Honjo, T., Kinoshita, K., Muramatsu, M. (2002) Molecular mechanisms of class switch recombination: Linkage with Somatic Hypermutation. *Annu Rev Immunol* **20:**165-196.

24. Jankovic, M., Robbiani, DF., Dorsett, Y., Eisenreich, T., and et al. (2010). Role of the translocation partner in protection against AID-dependent chromosomal translocations. *Proc Natl Acad Sci U S A* **107**:187-92

25. Jia, B., Jeon, CO. (2016) High-throughput recombinant protein expression in Escherichia coli: Current status and future perspectives. *Open Biol* **6:**160196.

26. Katakura, Y., Zhang, W., Zhuang, G., Omasa, T., Kishimoto, M., Goto, Y., Suga, K.I. (1998) Effect of methanol concentration on the production of human b2-glycoprotein I domain V by a recombinant *Pichia pastoris*: a simple system for the control of methano concentration using a semiconductor gas sensor. *J Ferm Bioeng* **86(5):**482–487.

27. Kern, A., Hartner, F.S., Freigassner, M., Spielhofer, J., Rumpf, C., Leitner, L., Frohlich, K.U., Glieder, A. (2007) Pichia pastoris "just in time" alternative respiration. *Microbiology* **153:**1250-1260.

28. Kobayashi, K., Kuwae, S., Ohya, T., Ohda, T., Ohyama, M., Ohi, H., Tomomitsu, K., Ohmura,T. (2000) High-level expression of recombinant human serum albumin from the methylotrophic yeast *Pichia pastoris* with minimal protease production and activation. *J Biosci Bioeng* **89:**55–61.

29. Koutz, PJ., Davis, GR., Stillman, C., Barringer, K., Cregg, JM., and Thill, G. (1989) Structural Comparison of the Pichia pastoris Alcohol Oxidase Genes. *Yeast* **5:**167-177.

30. Krainer, F.W., Dietzsch. C., Hajek, T., Herwig, C., Spadiut, O., Glieder, A. (2011) Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. *Microb Cell Fact*. **11:**22.

31. Kurjan, J., Herskowitz, I. (1982) Structure of a yeast pheromone gene (MF): a putative -factor precursor contains four tandem copies of mature -factor. *Cell* **30:**933–943 7.

32. Kurtzman, C. (2009) Biotechnological strains of *Komagataella (Pichia) pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. *J Ind Microbiol Biotechnol* **36:**1435–1438.

33. Lambertz, C., Garvey, M., Klinger, J., Heesel, D., Klose, H., Fischer, R., Commandeur, U. Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. *Biotechnol Biofuels*.**7:**1–15

34. Lada, A.G., Kliver, S.F., Dhar. A., Polev, D.E., Masharsky, A.E., Rogozin, I.B., et al. (2015) Disruption of Transcriptional Coactivator Sub1 Leads to Genome-Wide Re-

distribution of Clustered Mutations Induced by APOBEC in Active Yeast Genes. *PLoS Genet*. **11(5):**e1005217.

35. Larijani, M., Frieder, D., Basit, W., Martin, A. (2005) The mutation spectrum of purified AID is similar to the mutability index in Ramos cells and in ung(-/-)msh2(-/-) mice. *Immunogenetics* **56:**840-845

36. Larijani, M., Frieder, D., Sonbuchner, T.M., Bransteitter, R., Goodman, M.F., Bouhassira, E.E., Scharff, M.D., Martin, A. (2005) Methylation protects cytidines from AID-mediated deamination. *Mol Immunol* **42:**599-604.

37. Larijani, M., Petrov, A.P., Kolenchenko, O., Berru, M., Krylov, S.N., Martin, A. (2007) AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence- independent manner. *Mol Cell Biol* **27:**20-30.

38. Larijani, M. and Martin, A. (2007) Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. *Mol Cell Biol* **27:**8038-8048.

39. Larijani, M. and Martin, A. (2012) The biochemistry of activation-induced deaminase and its physiological functions. *Semin Immunol* **24:**255-263.

40. Lin-Cereghino, J., Wong, W.W., Xiong, S., Giang, W., Luong, L.T., Vu, J., Johnson, S.D., and Lin-Cereghino, GP. (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. *Biotechniques* **38:**44, 46, 48.

41. Liu, L., Haiquan, Yang, H., Shin, D. (2013) How to achieve high-level expression of microbial enzymes strategies and perspectives. *Bioengineered* **4(4):**212–223.

42. Longerich, S., Basu, U., Alt, F. and Storb, U. (2006) AID in somatic hypermutation and class switch recombination. *Curr Opin Immunol* **18(6):** 769.

43. Lu, Z. et al. (2013) BCL6 breaks occur at different AID sequence motifs in Ig-BCL6 and non-Ig-BCL6 rearrangements. *Blood* **121:**4551–4554.

44. Martin, A., Bardwell, P., Woo, C., Fan, M., Shulman, M. and Scharff, M. (2002) Activation- induced cytidine deaminase turns on somatic hypermutation in hybridomas. *Nature* **415(6873):**802-806.

45. Macauley-Patrick, S., Fazenda, M.L., McNeil, B. and Harvey, L.M. (2005) Heterologous protein production using the Pichia pastoris expression system. *Yeast* **22:**249-270.

46. Mechtcheriakova, D., Svoboda, M., Meshcheryakova, A., Jensen-Jarolim, E. (2012) Activation-induced cytidine deaminase (AID) linking immunity, chronic inflammation, and cancer. *Cancer Immunol Immunother* **61(9):**1591–8.

47. Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O, Honjo, T., (1999) 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:**18470-18476.

48. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinki, Y., Honjo, T. (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102:**553-563.

49. Neuberger, M.S., Di., Noia, J.M., Beale, R.C., Williams, G.T., Yang, Z., Rada, C. (2005) Somatichypermutation at A·T pairs: polymerase error versus dUTP incorporation. *Nat Rev Immunol* **5:**171-78.

50. Noseda, D.G., Blasco, M., Recupero, M., Galvagno, M.A. (2014) Bioprocess and downstream optimization of recombinant bovine chymosin B in Pichia (Komagataella) pastoris under methanol-inducible AOXI promoter. *Protein Expr Purif* **104C:** 85-91.

51. Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P., Muramatsu, M., Honjo, T., et al. (2008) AID is required for germinal center-derived lymphomagenesis. *Nat Genet* **40**:108–12.

52. Pham, P., Landolph, A., Mendez, N., Li, N. and Goodman, M.F. (2013) A Biochemical Analysis Linking APOBEC3A to Disparate HIV-1 Restriction and Skin Cancer. *J Biol Chem* **288:**29294–29304.

53. Pham, P., Afif, S.A., Shimoda, M., Maeda, K., Sakaguchi, N., Pedersen, LC., et al. (2016) Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification. *DNA Repair (Amst)* **43:**48–56.

54. Prochnow, C., Bransteitter, R., Chen X.S. (2016) APOBEC deaminases-mutases with defensive roles for immunity. *Sci China C Life Sci* **52(10):**893-902.

55. Prohaska, K.M., Bennett, R.P., Salter, J.D., Smith, HC. (2014) The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. *Wiley Interdiscip Rev RNA* **5:**493–508.

56. Poltoratsky, V.P., Wilson, S.H., Kunkel, T.A., Pavlov, Y.I. (2004) Recombinogenic phenotype of human activation-induced cytosine deaminase. *Immunol* **172:**4308–4313.

57. Quinlan, E.M., King, J.J., T.Amemiya C., Hsu, E., Larijani, M. (2017) Biochemical regulatory features of AID remain conserved from lamprey to humans. *Mol Cell Biol.*

58. Rada, C., Di Noia, J.M., Neuberger, M.S. (2004) Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell* **16**: 163-171.

59. Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., et al. (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102(5)**:565-75.

60. Rosano, G.L., Ceccarelli, E.A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5:**172.

61. Robbiani, D.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., et al. (2009) AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol Cell* **36:**631-641.

62. Sahdev, S., Khattar, S.K., Saini, K.S. (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* **307:**249–64.

63. Schumacher, A.J., Nissley, D.V., Harris, R.S. (2005) APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. *Proc Natl Acad Sci* **102:**9854–9859.

64. Scorer, C.A., Buckholz, R.G., Clare, J.J., and Romanos, M.A. (1993) The Intracellular Production and Secretion of HIV-1 Envelope Protein in the Methylotrophic Yeast Pichia pastoris. *Gene* **136:**111-119.

65. Shen, H. M., N. Michael, N. Kim, and U. Storb. (2000). The TATA binding protein, c-Myc and survivin genes are not somatically hypermutated, while Ig and BCL6 genes are hypermutated in human memory B cells. *Int. Immunol* **12**: 1085–1093

66. Shobayashi, M., Mukai, N., Iwashita, K., Hiraga, Y., Iefuji, H. (2006) A new method for isolation of S-adenosylmethionine (SAM)-accumulating yeast. *Appl Microbio Biotechnol* **69:**704-710.

67. Sousa, M.M., Krokan, H.E., Slupphaug, G. (2007) DNA-uracil and human pathology. *Mol Aspects Med* **28(3-4):**276-306.

68. Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **72:**211-22.

69. Tschopp, J.F., Brust, P.F., Cregg, J.M., Stillman, C.A., Gingeras, T.R. (1987) Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris. *Nucleic Acids Res* **15:**3859–3876.

70. Wang, M., Jiang, S., Wang, Y. (2016) Recent advances in the production of recombinant subunit vaccines in Pichia pastoris. *Bioengineered* **7(3):**155-165.

71. Wingfield, PT. (2015) Overview of the purification of recombinant proteins. Current protocols in protein. *scienc* **80:**6.1–.3.

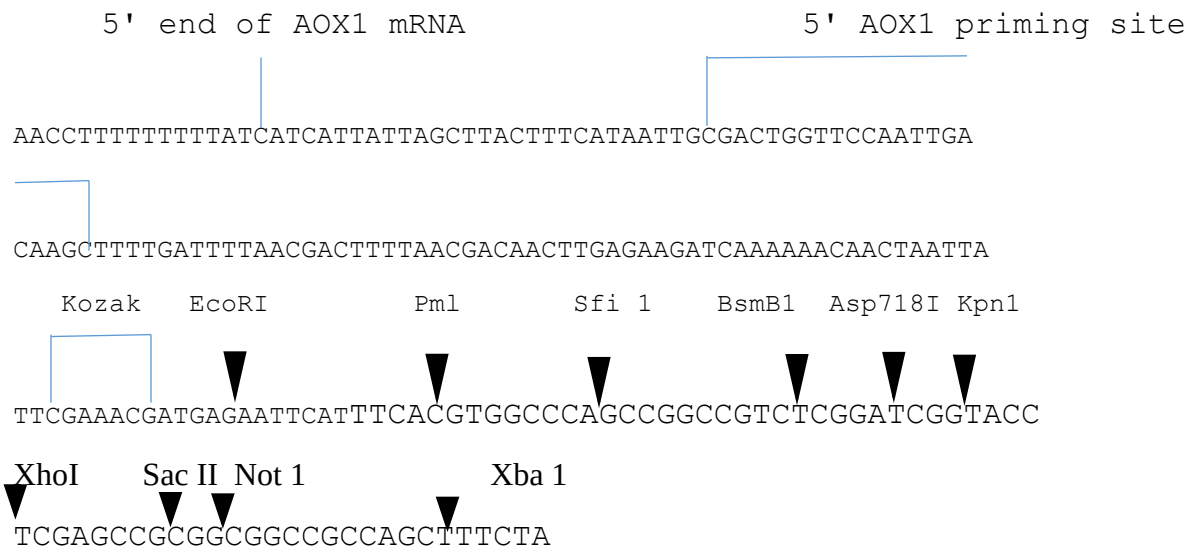
72. Zahn, A., Eranki, A.K., Patenaude, A.M., Methot, S.P., Fifield, H., Cortizas, E.M., Foster, P., Imai, K., Durandy, A., Larijani, M., Verdun R.E., Di Noia JM. (2014) Activation induced deaminase C-terminal domain links DNA breaks to end protection and repair during class switch recombination. *Proc Natl Acad Sci USA* **111:E** 988–E997

73. Zhou, Q., Li, M., Xi, T. (2009) Cloning and expression of a clamworm antimicrobial peptide perinerin in Pichia pastoris. *Current Microbiology* **58:**384–388.

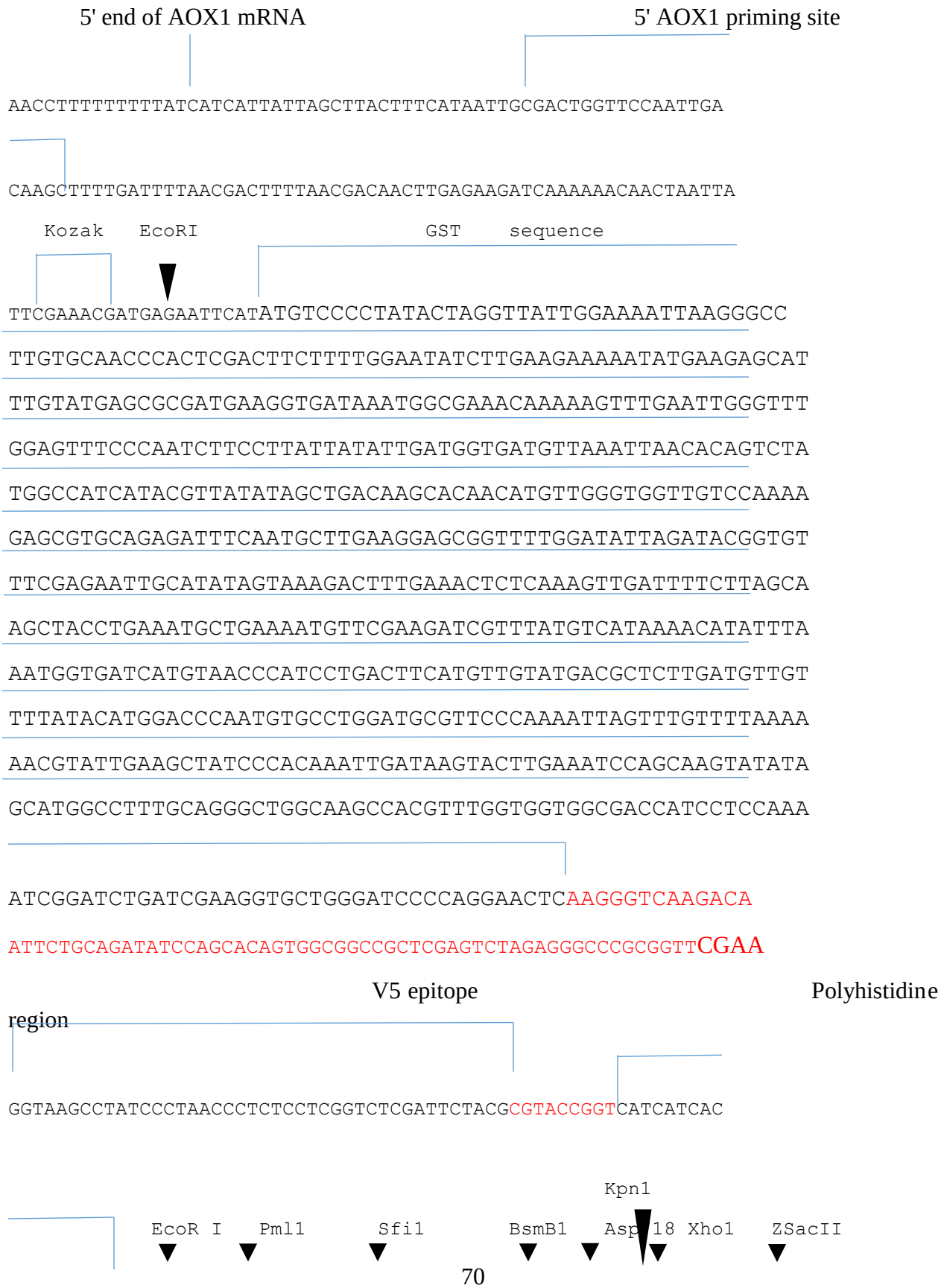
Appendix

Below are the sequences of the integrated oligonucleotide fragments with features common to all pPICZnonαB and PICZαB vectors. Sequences in the red are linkers after GST tag and AID or AID and V5-His tag.

1. **Mother vector pPICZnonαB**



2. **GST-V5-His in pPIZCnonαB**


5' end of AOX1 mRNA
5' AOX1 priming site
AACCTTTTTTTTTATCATCATTATTAGCTTACTTTCATAATTGCGACTGGTTCCAATTGA
CAAGCTTTTGATTTTAACGACTTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTA
Kozak
EcoRI
GST
sequence
TTCGAAACGATGAGAATTCATATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCC
TTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCAT
TTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTGAATTGGGTTT
GGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTA
TGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAA
GAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGT
TTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCA
AGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTA
AATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGT
TTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAA
AACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATA
GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAA
ATCGGATCTGATCGAAGGTGCTGGGATCCCCAGGAACTCAAGGGTCAAGACA
ATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCGAA
V5 epitope
Polyhistidine
region
GGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCAC
Kpn1
EcoR I
Pml1
Sfi1
BsmB1
Asp 18
Xho1
ZSacII

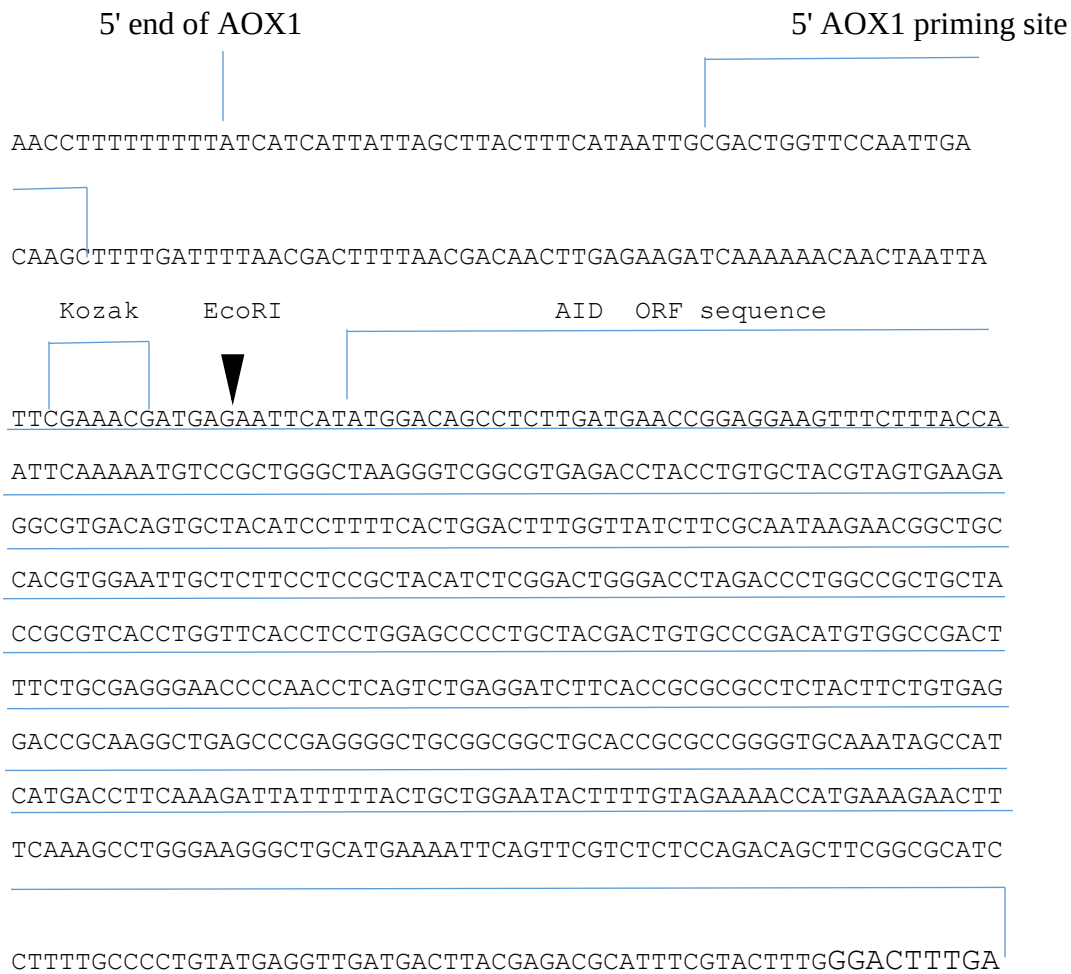
CATCACCATTGAGAATTCACGTGGCCCAGCCGGCCGTCTCGGATCGGTACCTCGAGCCGCG

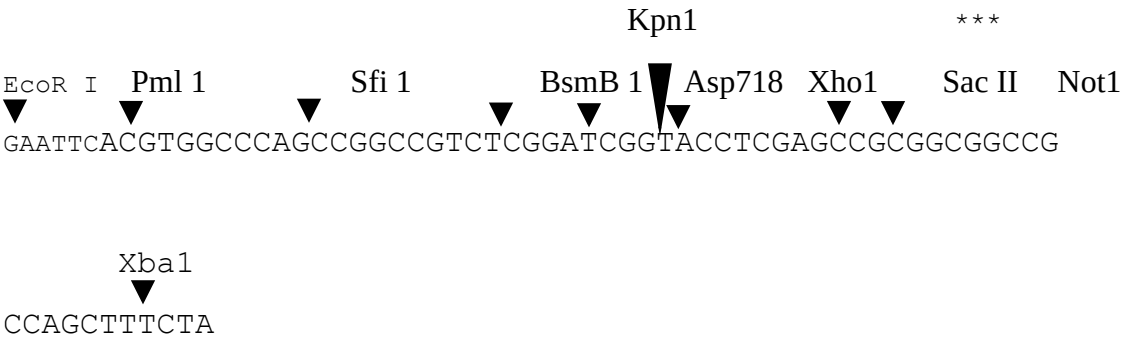
***

Not 1 Xba 1

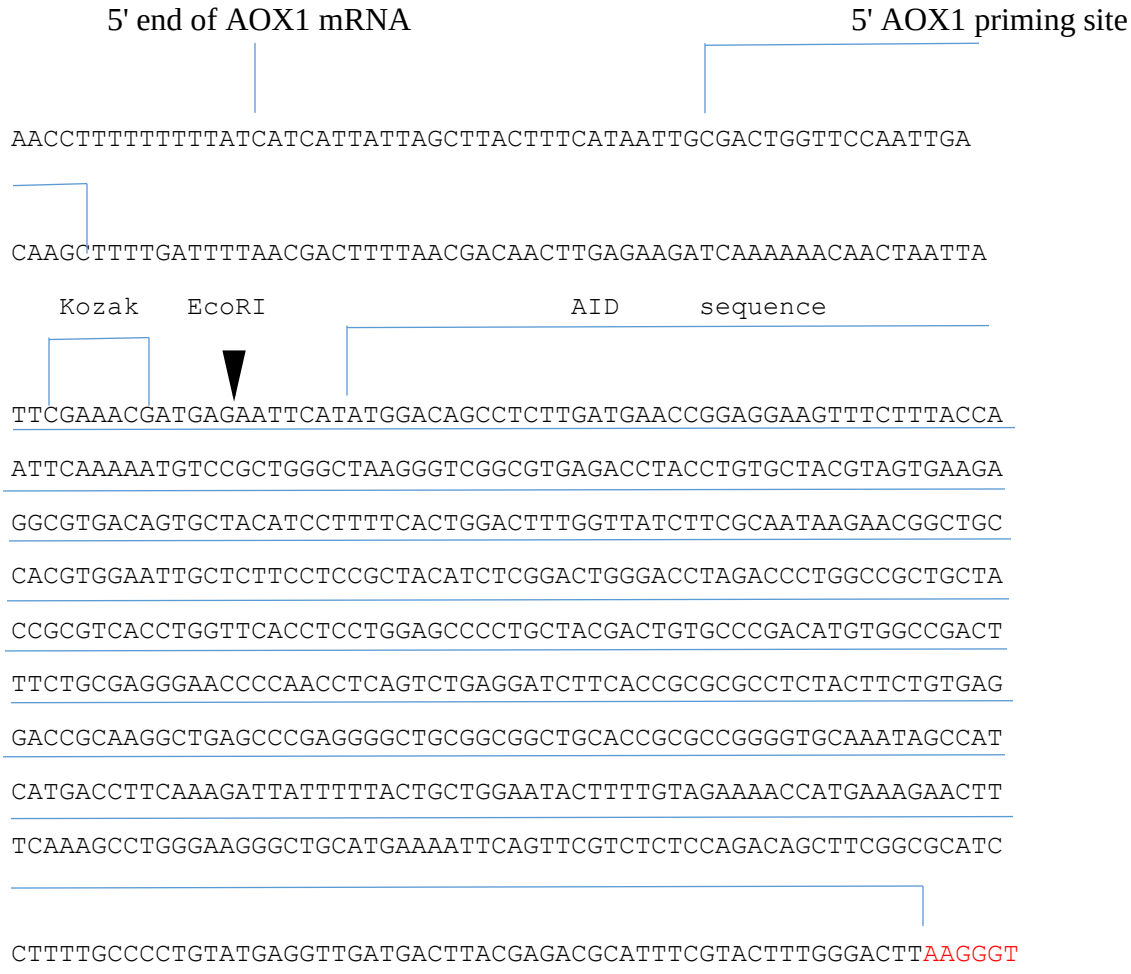
GCGGCCGCCAGCTTTCTA

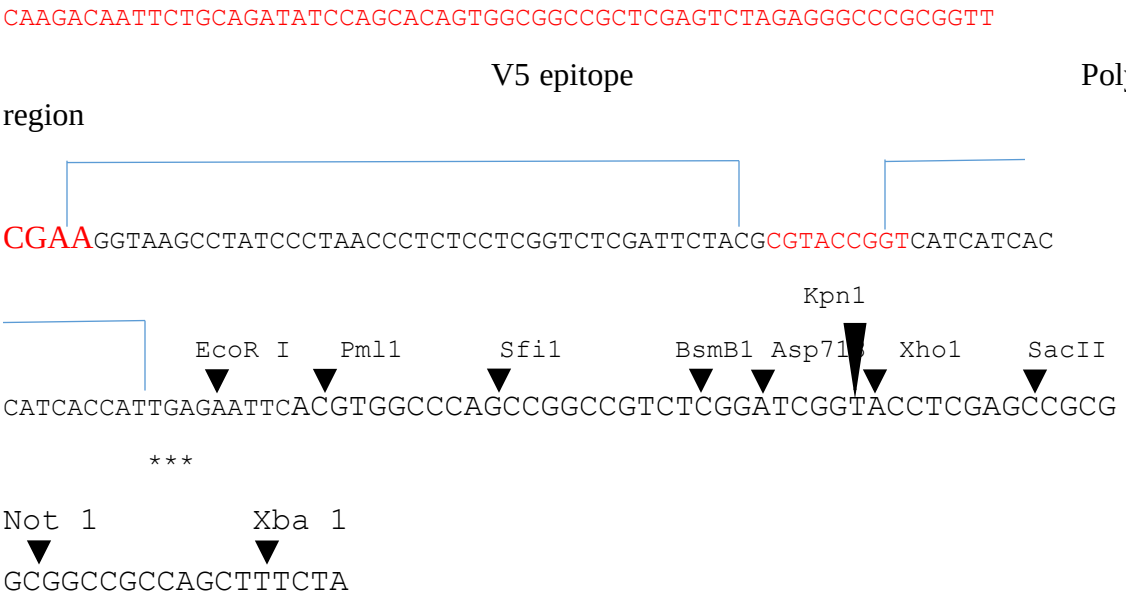
3. Untagged AID in pPICZnonαB



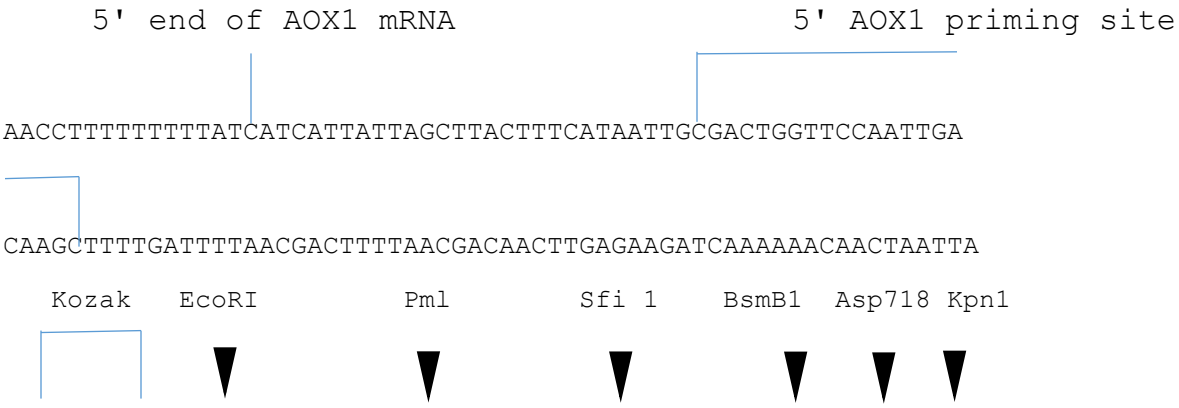


4. AID-V5-His in pPIZCnonαB





1. **Mother vector pPICZnonαB**



TTCGAAACGATGAGAATTCATTTCACGTGGCCCAGCCGGCCGTCTCGGATCGGTACC

Xho I Sac II Not 1 Xba 1

TCGAGCCGCGGCGGCCGCCAGCTTTCTA

2. GST-V5-His in pPIZCnonαB

5' end of AOX1 mRNA 5' AOX1 priming site

AACCTTTTTTTTTATCATCATTATTAGCTTACTTTCATAATTGCGACTGGTTCCAATTGA

CAAGCTTTTGATTTTAACGACTTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTA

Kozak EcoRI GST sequence

TTCGAAACGATGAGAATTCATATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCC

TTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCAT

TTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTGAATTGGGTTT

GGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTA

TGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAA

GAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGT

TTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCA

AGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTA

AATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGT

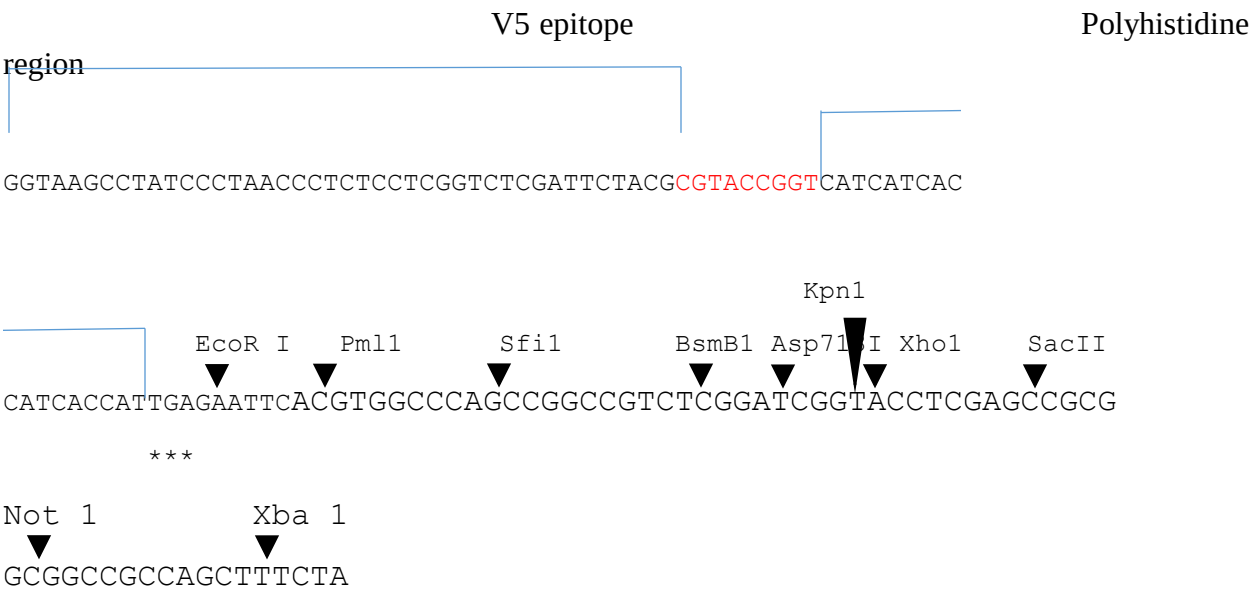
TTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAA

AACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATA

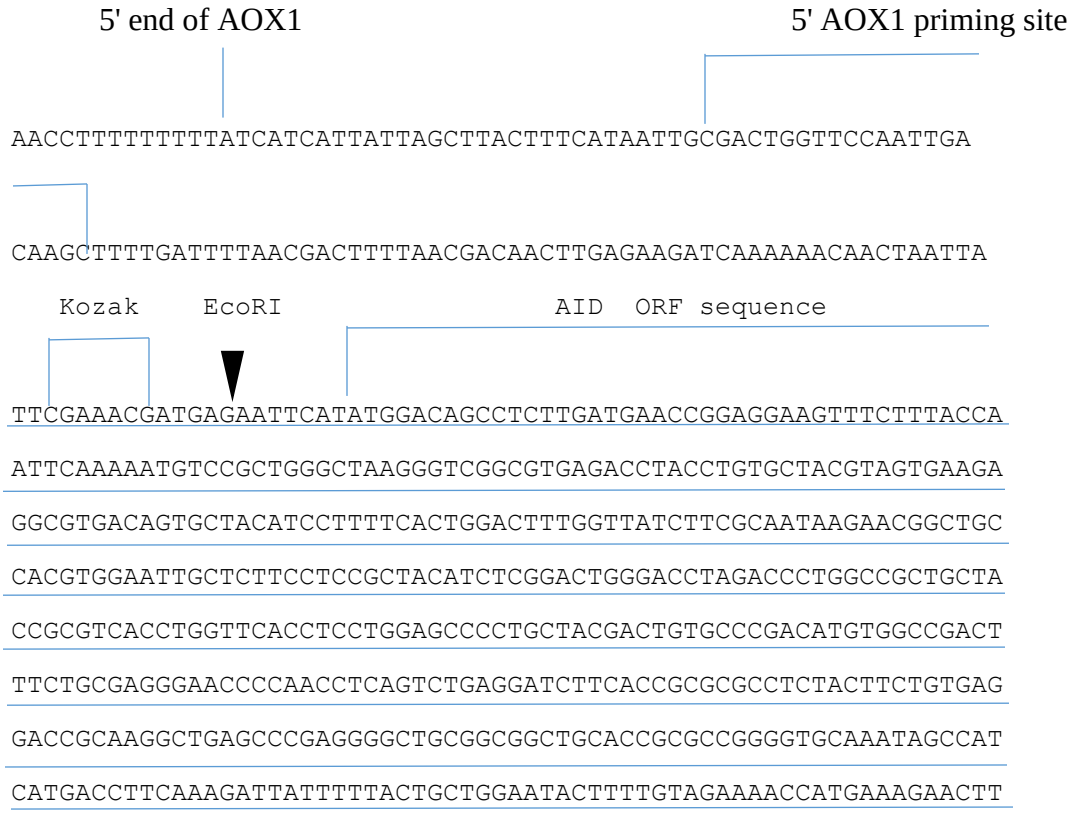
GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAA

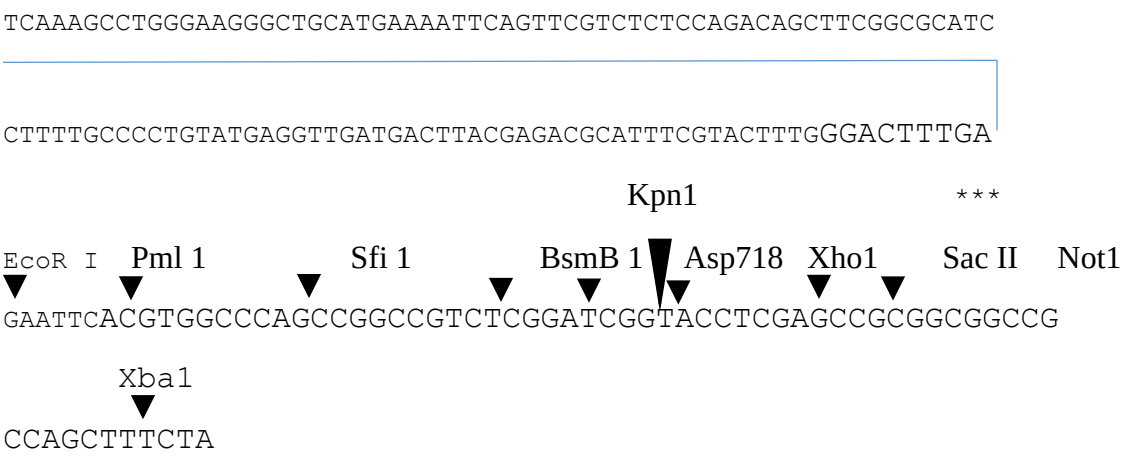
ATCGGATCTGATCGAAGGTGCTGGGATCCCCAGGAACTCAAGGGTCAAGACA

ATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCGAA

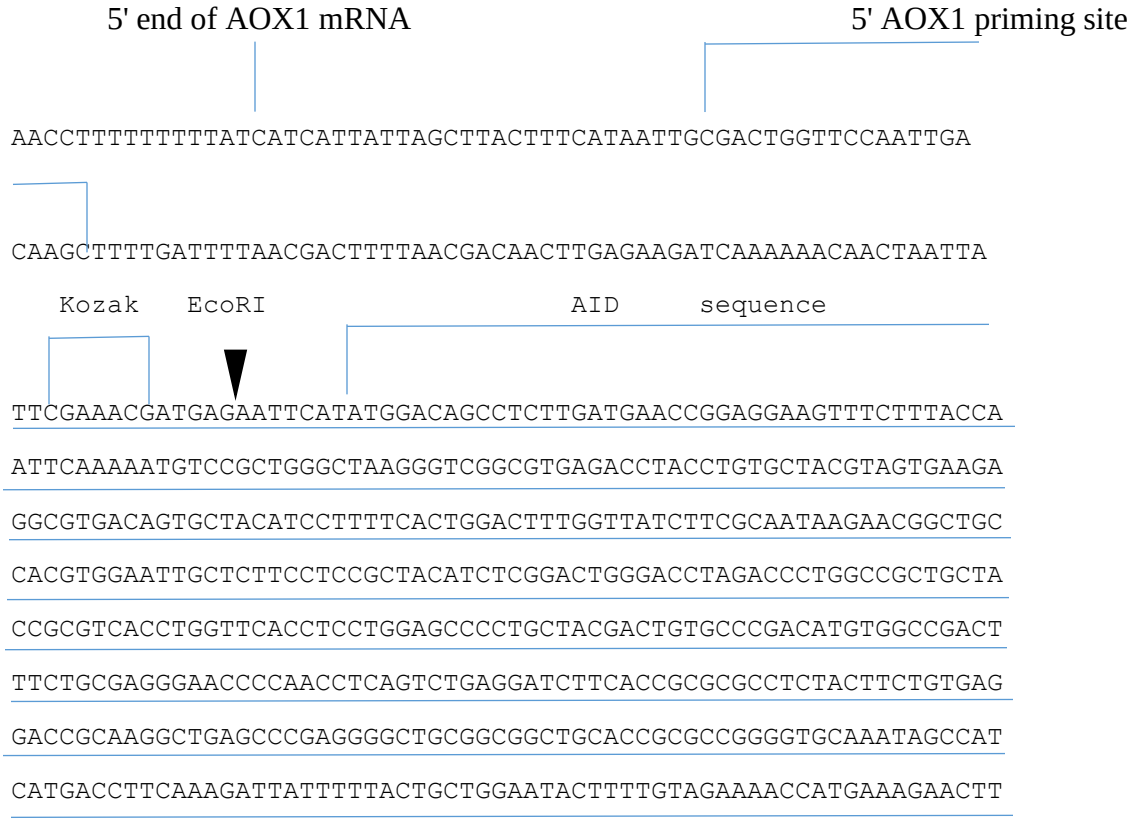


3. Untagged AID in pPICZnonαB

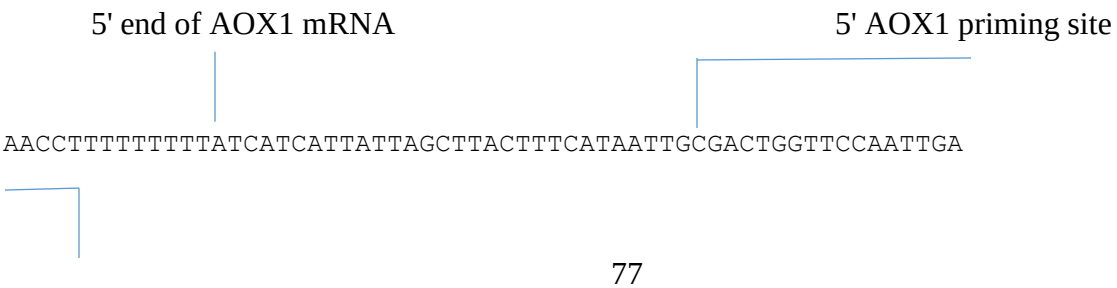


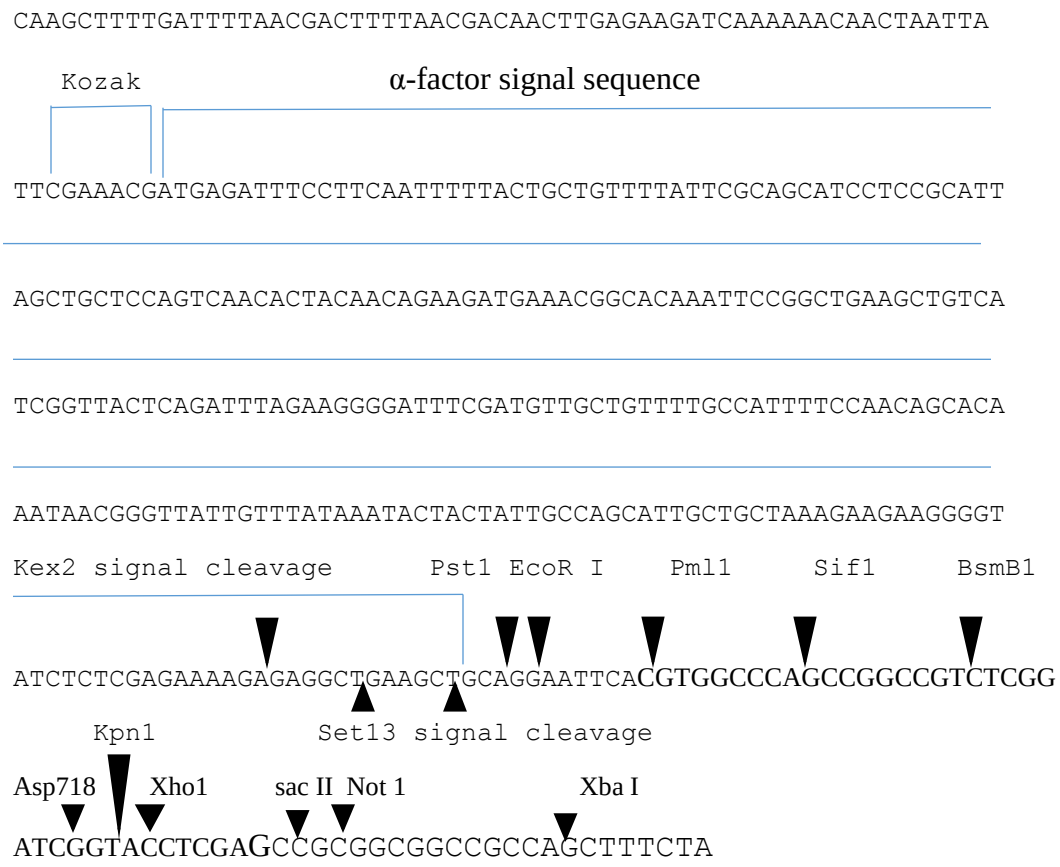


4. AID-V5-His in pPIZCnonαB

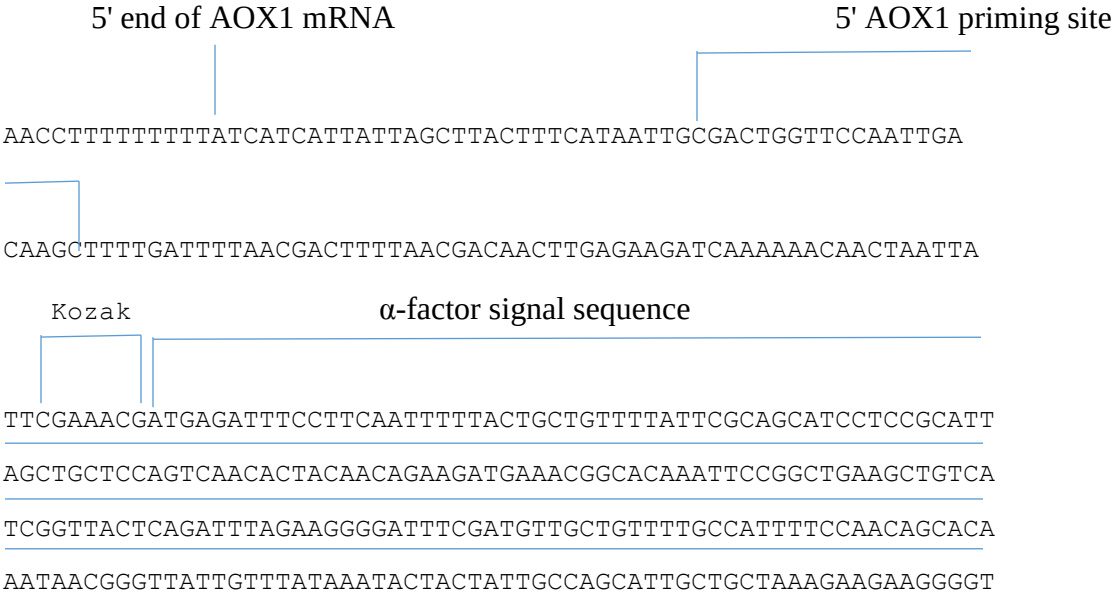


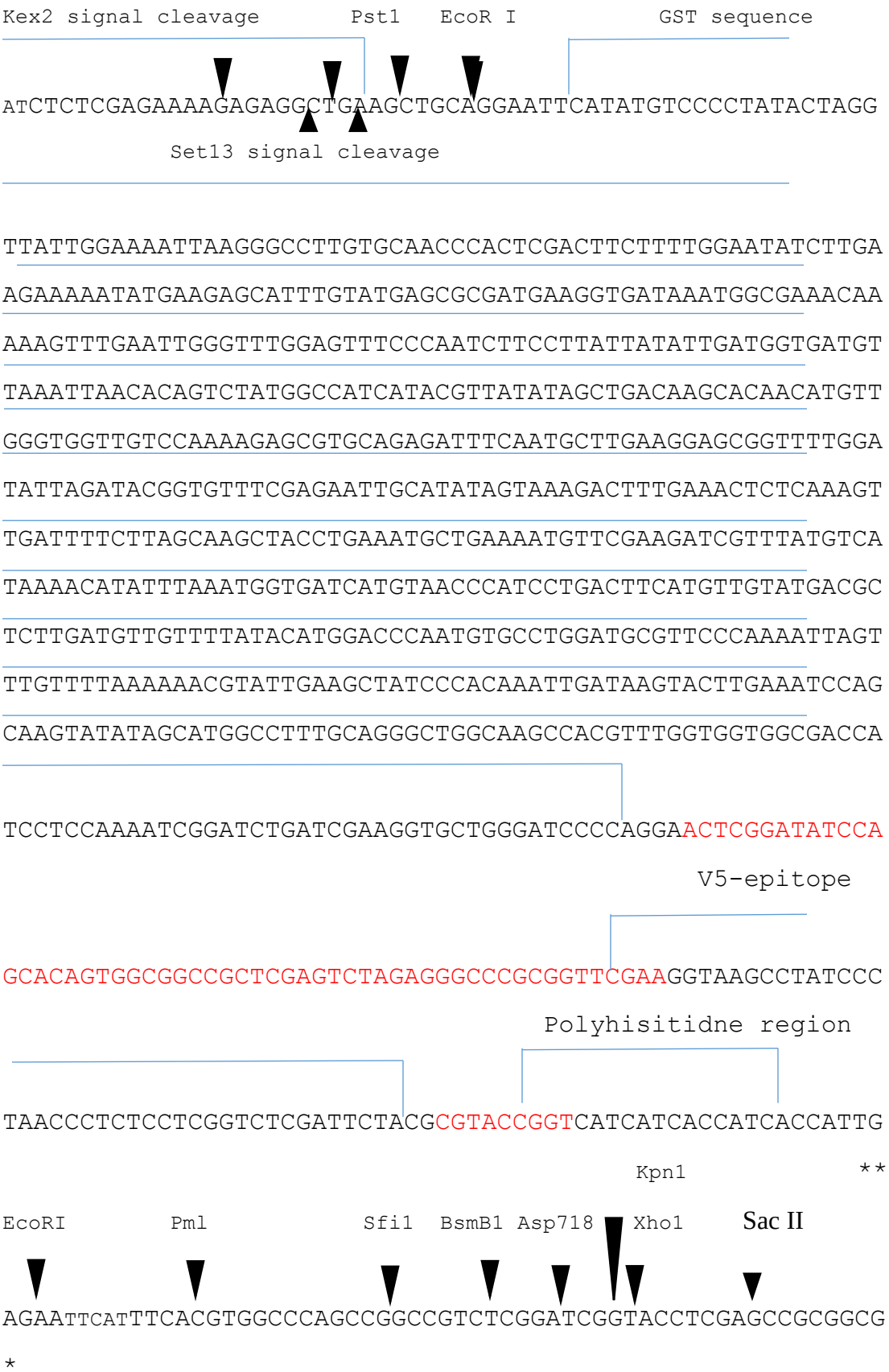
5. Mother vector in pPIZCαB

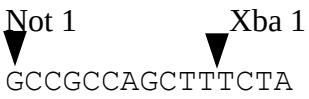




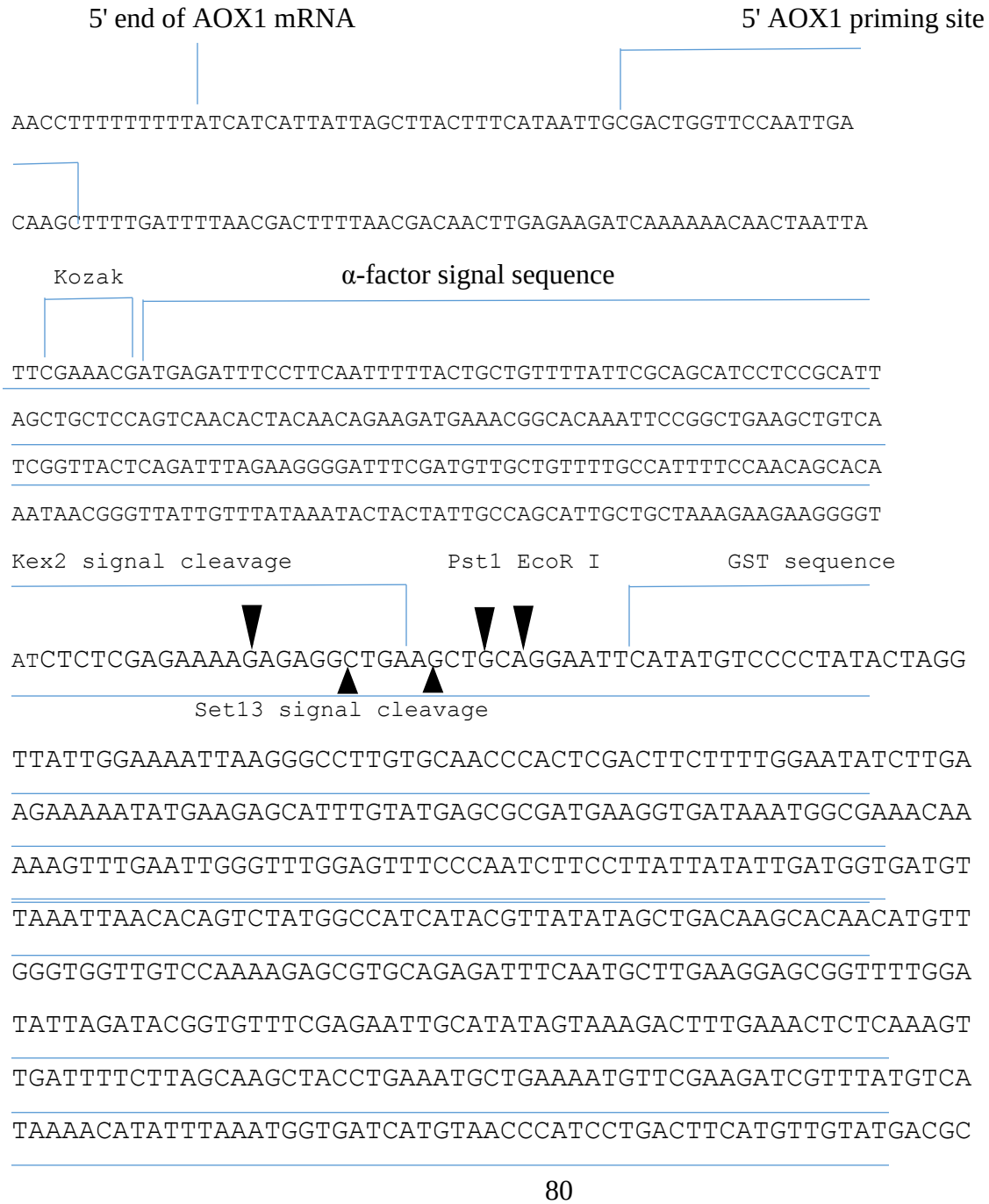
6. GST-V5-His in pPIZCαB

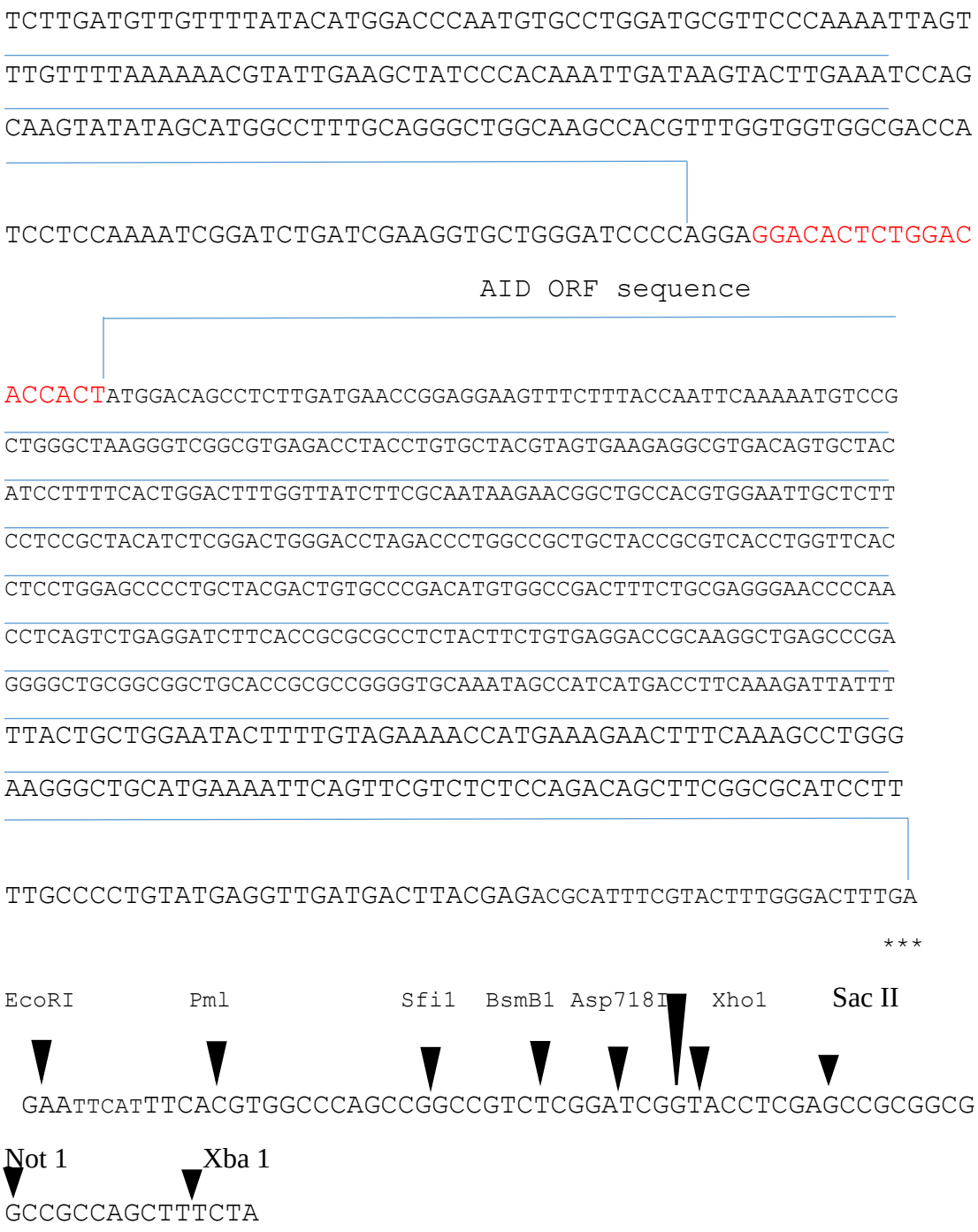



Kex2 signal cleavage
Pst1
EcoR I
GST sequence
ATCTCTCGAGAAAAGAGAGGCTGAAGCTGCAGGAATTCATATGTCCCCTATACTAGG
Set13 signal cleavage
TTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGA
AGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAA
AAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGT
TAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTT
GGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGA
TATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGT
TGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCA
TAAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGC
TCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGT
TTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAG
CAAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCA
TCCTCCAAAATCGGATCTGATCGAAGGTGCTGGGATCCCCAGGAACTCGGATATCCA
V5-epitope
GCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCC
Polyhisitidne region
TAACCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTG
Kpn1
**
EcoRI
Pml
Sfi1
BsmB1
Asp718
Xho1
Sac II
AGAATTCATTTCACGTGGCCCAGCCGGCCGTCTCGGATCGGTACCTCGAGCCGCGGCG
*

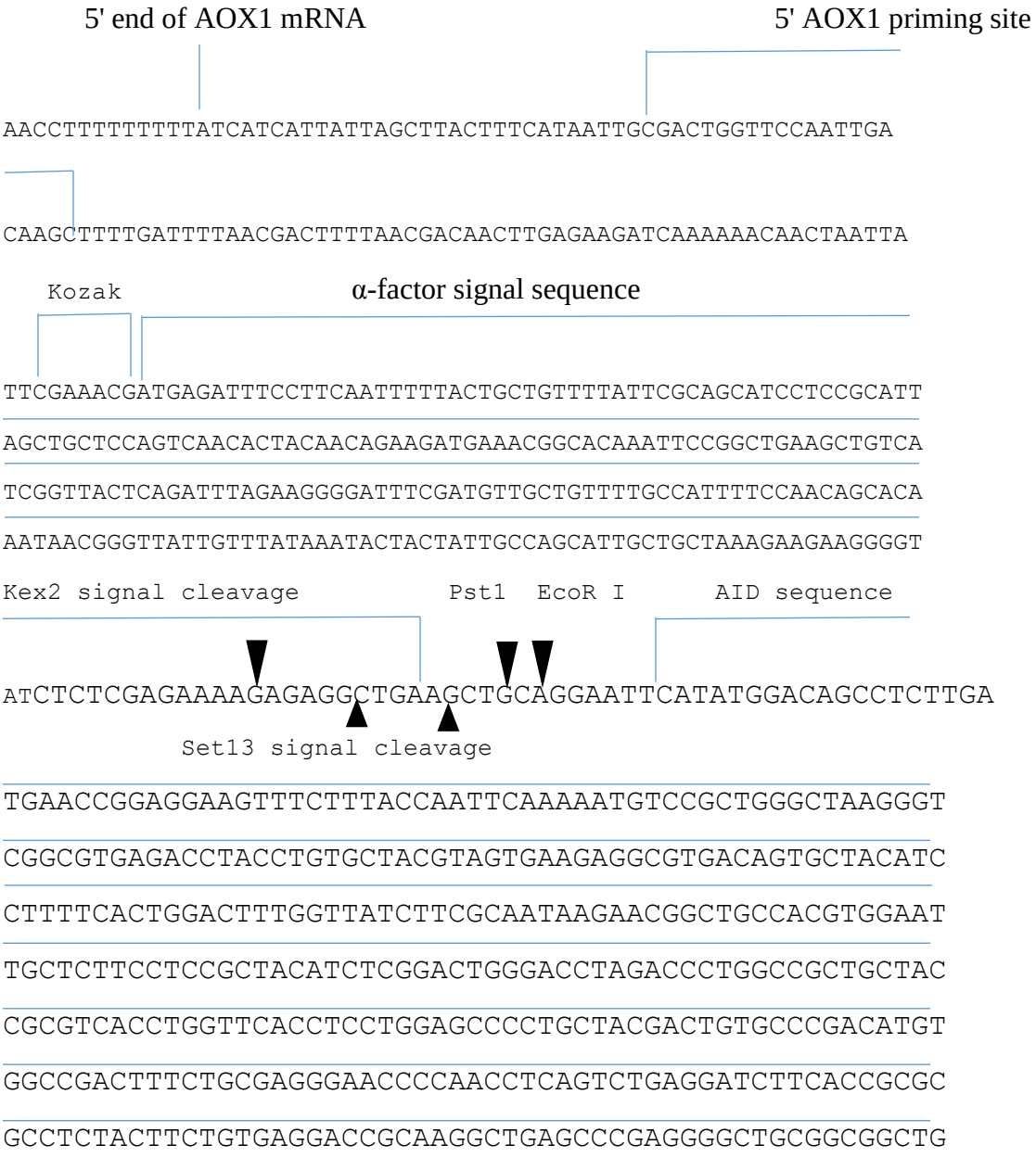


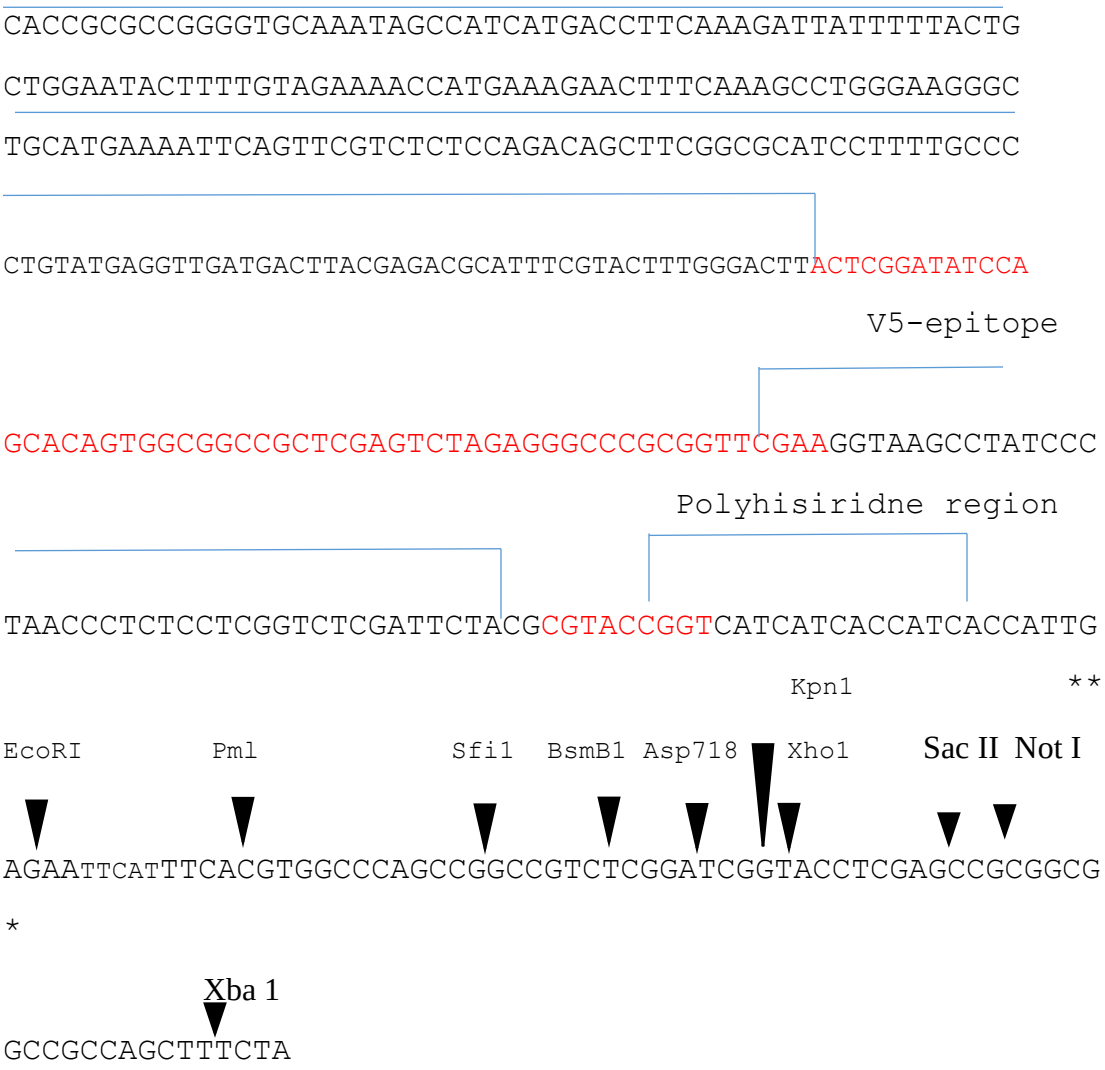
7. GST-AID in pPIZCαB



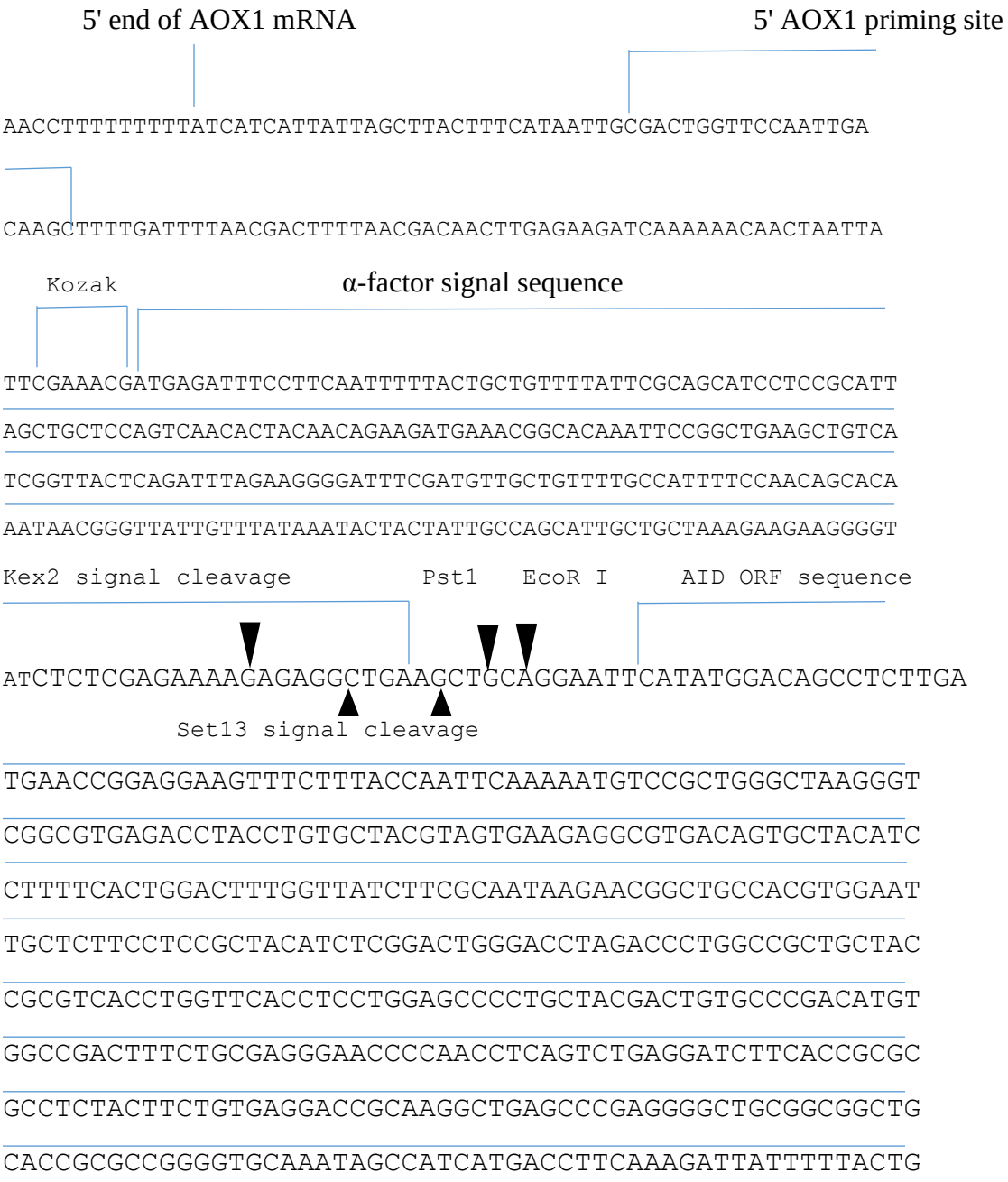

TCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGT
TTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAG
CAAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCA
TCCTCCAAAATCGGATCTGATCGAAGGTGCTGGGATCCCCAGGAGGACACTCTGGAC
AID ORF sequence
ACCACTATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCG
CTGGGCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTAC
ATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTCTT
CCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCCGCTGCTACCGCGTCACCTGGTTCAC
CTCCTGGAGCCCCTGCTACGACTGTGCCCGACATGTGGCCGACTTTCTGCGAGGGAACCCCAA
CCTCAGTCTGAGGATCTTCACCGCGCGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGA
GGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGCCATCATGACCTTCAAAGATTATTT
TTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAAGCCTGGG
AAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT
TTGCCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTTTGA
***
EcoRI
Pml
Sfi1
BsmB1
Asp718I
Xho1
Sac II
GAATTCATTTCACGTGGCCCAGCCGGCCGTCTCGGATCGGTACCTCGAGCCGCGGCG
Not 1
Xba 1
GCCGCCAGCTTTCTA

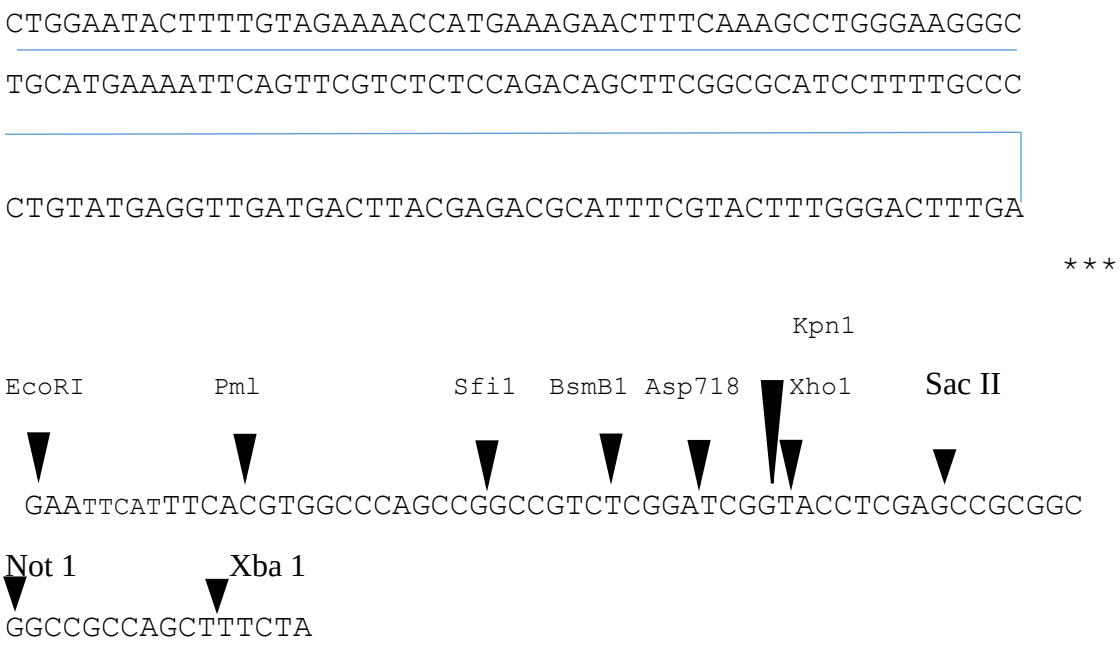
8. AID-V5-His-pPIZCαB




CACCGCGCCGGGGTGCAAATAGCCATCATGACCTTCAAAGATTATTTTTACTG
CTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAAGCCTGGGAAGGGC
TGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGCCC
CTGTATGAGGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTTACTCGGATATCCA
V5-epitope
GCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCGAAGGTAAGCCTATCCC
Polyhisiridne region
TAACCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTG
Kpn1
**
EcoRI
Pml
Sfi1
BsmB1
Asp718
Xho1
Sac II
Not I
AGAATTCATTTCACGTGGCCCAGCCGGCCGTCTCGGATCGGTACCTCGAGCCGCGGCG
*
Xba 1
GCCGCCAGCTTTCTA

8. Untagged AID in pPIZCαB




CTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAAGCCTGGGAAGGGC
TGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGCCC
CTGTATGAGGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTTTGA
***
Kpn1
EcoRI
Pml
Sfi1
BsmB1
Asp718
Xho1
Sac II
GAATTCATTTCACGTGGCCCAGCCGGCCGTCTCGGATCGGTACCTCGAGCCGCGGC
Not 1
Xba 1
GGCCGCCAGCTTTCTA