

**Characterizing the DNA binding sites of activation-  
induced cytidine deaminase: A mutagenesis  
approach**

by

© Heather M. Lucas (B.Sc., B.Ed.)

A Thesis submitted to the

School of Graduate Studies

In partial fulfillment of the requirements of

**Master of Science in Medicine**

Immunology and Infectious Diseases

Faculty of Medicine

Memorial University of Newfoundland

**July 2013**

St. John's

Newfoundland

## **Abstract**

Activation induced cytidine deaminase (AID) is a 198 amino acid DNA mutating enzyme that converts deoxycytidine (dC) to deoxyuridine (dU) in single-stranded DNA (ssDNA). AID activity is important for adaptive immune responses with a key role during the affinity maturation of antibodies in activated B-lymphocytes. AID targets the trinucleotide motif WRC (W=A/T, R=A/G) on (ssDNA), but neither its enzymatic mechanism(s) nor its modes of binding to ssDNA are known, largely due to a lack of protein structural information. A model of the enzymatic structure of AID suggested two potential ssDNA binding grooves on its surface. To confirm or refute the presence of two substrate binding grooves (Groove 1 and 2); amino acid residues were mutated singly or in combination in Grooves 1 and 2 to evaluate changes in DNA binding affinities and AID enzyme activity velocity compared to wild-type (wt) protein. Some AID mutants were inactive and had little or no binding to DNA, while others retained these properties relative to the wt enzyme. Contrary to our hypothesis, electrostatic interaction is not integral to AID and substrate interaction as previously believed. Along with activity and binding, conformational changes were confirmed in AID mutants, which could explain the differences in enzymatic functionality. Our research suggests AID conformation is key for substrate targeting and binding, and that putative Groove 1 likely contains contact points with ssDNA substrate.

## **Acknowledgements**

I would like to thank the Canadian Institutes of Health Research, my Masters supervisor Dr. Mani Larijani, the staff and students of the Valerie Booth Lab, and Dr. Valerie Booth for their support in completing my research, the use of equipment, supplies, and financing. I would also like to extend a thank-you to my peers and research assistants in the Larijani Lab for their support and assistance throughout my time in the program.

Extended gratitude is deserved of my committee members, Ann Dorward and Rodney Russell, for their time and advice during the editing process. Thank-you.

To my family and friends, thank-you for all that you do to support me in all that I do. You've made all the difference in the world.

## List of Figures

Figure 1. Y shaped structure of antibody molecule	4
Figure 2. Affinity Maturation of B-cells.	7
Figure 3. The mechanism of Somatic Hypermutation and Class Switch Recombination pathways.	13
Figure 4. Binding Affinity versus Enzyme Activity.	22
Figure 5. Human wild-type AID ribbon diagram.	25
Figure 6. AID Model and Putative Binding Grooves 1 & 2.	26
Figure 7. Linear AID sequence.	28
Figure 8. APOBEC-3G Surface Model and combined Surface and ribbon model.	29
Figure 9. Mutant Analysis.	31
Figure 10. Representative Gel of Alkaline Cleavage assay.	46
Figure 11. Representative gel image of Alkaline Cleavage Kinetic assay for K52A AID mutant.	47
Figure 12. Representative gel image of EMSA Assay for three AID mutants (K52D, R50D and R63D).	48
Figure 13. Representative gel image of kinetic EMSA Assay.	49
Figure 14. Alkaline Cleavage assay bar graphs	50
a. Groove 1: Mutants compared to wild-type using percentage of product formed	50
b. Groove 2: Mutants compared to wild-type using percentage product formed	50
Figure 15. EMSA Binding Bar Graph	51
a. Groove 1: Mutants compared to wild-type using percentage of substrate bound	51
b. Groove 2: Mutants compared to wild-type using percentage substrate bound	51
Figure 16. Groove 1 - K22 mutations and assays	53
Figure 17. Groove 1 & 2 - R24 mutations and assays	55
Figure 18. Groove 1 - R25 mutations and assays	57
Figure 19. Groove 1 - R63 mutations and assays	59
Figure 20. Groove 1 - D89 mutations and assays	61
Figure 21. Groove 1 - D96 mutations and assays	63
Figure 22. Groove 1 - D118 mutations and assays	65
Figure 23. Groove 2 - E26 mutations and assays	67
Figure 24. Groove 2 - R50 mutations and assays	69
Figure 25. Groove 1 - K52 mutations and assays	71
Figure 26. Groove 1 - Multiple Mutant K22D/K52D/R63D assays	73
Figure 27. Groove 1 and 2 - Multiple Mutant R24D/R63D	75
Figure 28. Groove 1 - Multiple Mutant D89R/D96R assays	77
Figure 29. Binding Affinity versus Enzyme Velocity.	84
a. Binding Affinity versus Enzyme Velocity for Groove 1 Mutants	84
b. Binding Affinity versus Enzyme Velocity for Groove 2 Mutants	84
c. Binding Affinity versus Enzyme Velocity for Multiple Mutants	84
Figure 30. Models of R24 in wild-type AID, R24D mutation and R24A mutation.	93

## **List of Tables**

Table 1. Site directed mutagenesis primers	<b>34</b>
Table 2. Residues mutated according to mutation type and groove location.	<b>45</b>
Table 3. Kinetic Constant Values	<b>80</b>

## List of Abbreviations

µg	microgram
µl	microlitre
°C	Degrees Celsius
A	Adenine
Å	Angstrom(s)
A3F	Apoplipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like 3F
A3G	Apoplipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like 3G
Ab	Antibody
AID	Activation-induced cytidine deaminase
ADAR	Adenosine deaminase acting on dsRNA
Amp	Ampicillin
APE1	Apurinic/aprimidinic endonuclease
ApoB	Apolipoprotein B
APOBEC	Apolipoprotein B mRNA-editing, enzyme-catalytic polypeptide-like
BCL-6	B-cell lymphoma 6
BCR	B-cell receptor
BER	Base excision repair
BL	Burkitt's Lymphoma
Bp	Base pair
BSA	Bovine serum albumin
C	Cytidine
C region	Constant region of Ig
CLL	Chronic lymphocytic leukemia
D	Aspartic acid
dC	Deoxycytidine
DCDT	dCMP deaminases
DE3	<i>Escherichia coli</i> BL21
DLCL	Diffuse large B cell lymphoma
DNA	Deoxyribonucleic acid
DSB	Double-stranded break
DTT	Dithiothreitol
dU	Deoxyuridine
dsDNA	Double-stranded DNA
E	Glutamic acid
EDTA	Ethylenediamine tetraacetic acid
FAS	Fas cell surface death receptor (NFRSF6, APO-1, APT-1, CD95)
FL	Follicular lymphoma
g	Gram
G	Guanine
GST	Glutathione S-transferase

h	Hour(s)
Ig	Immunoglobulin
IPTG	Isopropyl- $\beta$ thiogalactopyranoside
K	Lysine
kDa	Kilodalton
LB	Lysogeny broth (Luria-Bertani medium)
M	Molar
min	Minute(s)
mL	Millilitre
MMR	Mismatch repair
mRNA	Messenger ribonucleic acid
Myc	v-Myc myelocytomatosis viral oncogene homolog (avian)
NES	Nuclear export signal
NLS	Nuclear localization signal
nt	Nucleotide
OD	Optical density
PAX5	Paired box 5
PIM-1	pim-1 oncogene (kinase associated protein)
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
R	Arginine
RHOH	RAS homolog family member H
RNA	Ribonucleic acid
s	Second(s)
S	Switch region of Ig
SDS	Sodium dodecyl sulphate
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM	Standard error of mean
SOC	Super Optimal Broth with Catabolite repression
ssDNA	Single-stranded DNA
T	Thymidine
tadA	tRNA adenosine deaminase
TBE	Tris borate/EDTA electrophoresis buffer
TEMED	Tetramethylethylenediamine
U	Uracil
UDG	Uracil-N-DNA glycosylase
V	Variable region of Ig
wt	wild-type activation-induced cytidine deaminase
Zn	Zinc

## **Table of Contents**

<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 The Immune system and development of the antibody response	1
1.2 Antibody structure, primary and secondary antibody diversification	2
1.3 DNA targets of AID during secondary antibody diversification	8
1.4 Pathways mediating the full spectrum of AID mutagenesis	11
1.5 The role of AID in tumorigenesis	14
1.6 Clinical mutations and translocations of oncogenes and AID activity	15
1.7 The biochemical features and putative structure of AID	17
1.8 Modeling the putative 3-dimensional structure of AID	20
1.9 Identification of two putative ssDNA binding grooves	23
1.10 Hypothesis	30
<b>Chapter 2 Materials and Methods</b>	<b>31</b>
2.1 Overview of AID mutagenesis and analyses:	31
2.2 Site Directed Mutagenesis	31
2.3 AID Protein Purification	36
2.4 AID Substrate Preparation	38
2.5 Deamination Activity Assay	39
2.6 Affinity-Binding Assay	40
2.7 Quantification of Assay Gel Images	41
2.8 Structural modeling of wt and mutant AID	42
<b>Chapter 3 Results</b>	<b>43</b>
3.1 AID Sequence and Mutagenesis	43
3.2 Overview of AID mutagenesis	43
3.2.1 Groove 1: K22 results	52
3.2.2 Groove 1 and 2: R24 results	54
3.2.3 Groove 1 and 2: R25 Results	56
3.2.4 Groove 1: R63 results	58
3.2.5 Groove 1: D89 results	60
3.2.6 Groove 1: D96 results	62
3.2.7 Groove 1: D118 results	64
3.3.1 Groove 2: E26 results	66
3.3.2 Groove 2: R50 results	68
3.3.3 Groove 2: K52 results	70
3.4.1 Multiple Mutants: K22D/K52D, K22D/R63D AND K22D/K52D/R63D	72
3.4.2 Multiple mutants: R24D/R63D and R25D/R63D results	74
3.4.3 Multiple Mutants: D89R/D96R	76
3.5 Mutation Summary	78
<b>Chapter 4 Discussion</b>	<b>81</b>
4.1 Alkaline Cleavage and EMSA	81
4.1.1 Groove 1 Mutations	82
4.1.2 Groove 2 Mutations	87
4.1.3 Multiple Mutants	89

<b>4.2 The relationship between catalytic activity and ssDNA binding affinity of AID</b>	<b>91</b>
<b>4.3 The role of AID surface structure in enzyme activity and binding</b>	<b>94</b>
<b>4.4 Future Studies</b>	<b>98</b>
<b>Chapter 5 Bibliography</b>	<b>100</b>

## **Chapter 1 Introduction**

### **1.1 The Immune system and development of the antibody response**

The human immune system is comprised of two main types of defence against pathogens: innate responses and adaptive responses. The innate immune response is a first line of defence that is constitutively present and consists of physical barriers, chemicals, and cells (e.g. macrophages) that non-specifically block, take up, and degrade invading agents. Meanwhile, the adaptive immune response takes days to weeks to develop after a pathogenic invasion, but is highly antigen specific, ensuring the generation and maintenance of immunological memory for subsequent challenges. The adaptive immune response is largely carried out by cells known as lymphocytes, of which there are two types: B cells and T cells (Fleisher et al., 1997; Lanasa et al., 2011; Warrington et al., 2011).

B cells are integral to adaptive immunity, in that they specifically bind a wide variety of pathogens through their surface-expressed antibodies. The initial antibody repertoire of B cells covers a wide range of antigens with low affinity, since it has not yet encountered any specific antigen. Such B-cells are considered naïve B-cells and originate in bone marrow, a primary lymphoid organ (Gellert, 2002). Immunoglobulins or antibodies are first expressed on the membrane surface of the B-cell as a B-cell receptor (BCR) and then interact with antigens. Pathogens are capable of activating B cells by crosslinking Ig (immunoglobulin) M-type antibodies on the B cell surface. The B cell then releases soluble antibodies that specifically bind

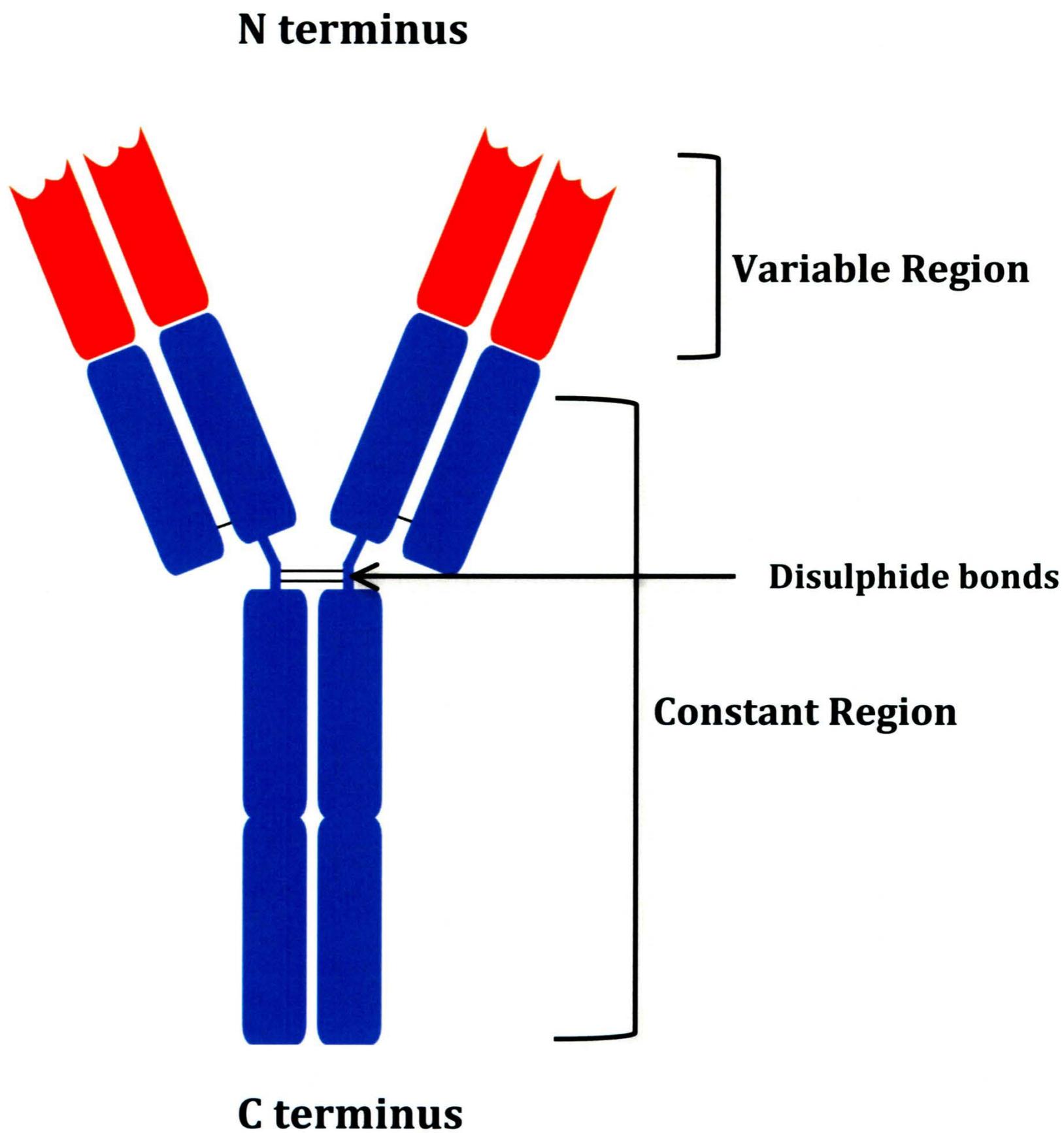
antigens, before migrating to secondary lymphoid organs, lymph nodes, to form germinal centres. Within the germinal centres, B-cells mature in a process known as affinity maturation, undergoing proliferation and activation-induced cytidine deaminase (AID)-mediated mutagenesis of antibody genes (Fleisher et al., 1997; Lanasa et al., 2011; Warrington et al., 2011). During this process, B cells develop antibodies with higher affinity for specific antigens thereby eliciting a stronger and more specific immune response against the current infection and against any future infection by the same pathogen (Martin et al., 2002; Muramatsu, et al., 2000 Okazaki et al., 2002; Revy, et al., 2000). In addition to higher affinity, gene rearrangement can also occur causing B cells to switch different antibody isotypes (IgA, IgE, and IgG), which ultimately determines functionality. Some of these B cells will become antibody secreting plasma cells, while others differentiate into memory cells, retaining a strong affinity for the specific pathogen encountered. This ensures a more efficient immune response during subsequent infection by the same pathogen (Longerich et al., 2006).

## **1.2 Antibody structure, primary and secondary antibody diversification**

The structure of an antibody molecule is loosely “Y” shaped (Fig. 1) and is comprised of 2 heavy chains and 2 light chains held together by disulphide bonds. Heavy chains are composed of constant and variable regions, with the constant regions of the two heavy chains forming the body of the Y structure. This portion of

the antibody performs several tasks: it acts as an anchor to the B cell membrane and is responsible for the isotype and effector functions of the antibody. This is encoded by the constant region of the antibody genes. The isotype determines the biological processes mediated by the antibody, such as opsonisation of pathogens, complement activation and activation of mast cells to release chemical mediators. The "arms" of the Y-shaped antibody molecule are made up of the light chains and the variable regions of the heavy chains. This portion of the antibody molecule contains highly variable loops and mediates specific binding to pathogens. This is the region that is generated by V(D)J gene recombination in and VJ gene recombination in the light chains in the primary (naïve) antibody repertoire and further diversified by the action of activation-induced cytidine deaminase (AID) in the mature (post-challenge) antibody repertoire (Barreto et al., 2005; Fleisher et al., 1997; Longerich et al., 2006; Neuberger et al., 2005).

Primary diversification of the pre-immune naive antibody repertoire is achieved through a process called V(D)J recombination, in which individual gene segments of the Ig locus are recombined to create a functional antibody gene. In the case of B cells, variable (V), diversity (D), and joining (J)- type gene segments are recombined to form a V(D)J gene that encodes for the variable regions of the antibody heavy chain. The variable portion of the light chain is similarly encoded by a gene that is the result of recombination between V and J gene segments. Within the heavy chain gene locus there are 40 V gene segments, 25 D gene segments, and 6



**Figure 1. Y shaped structure of antibody molecule**

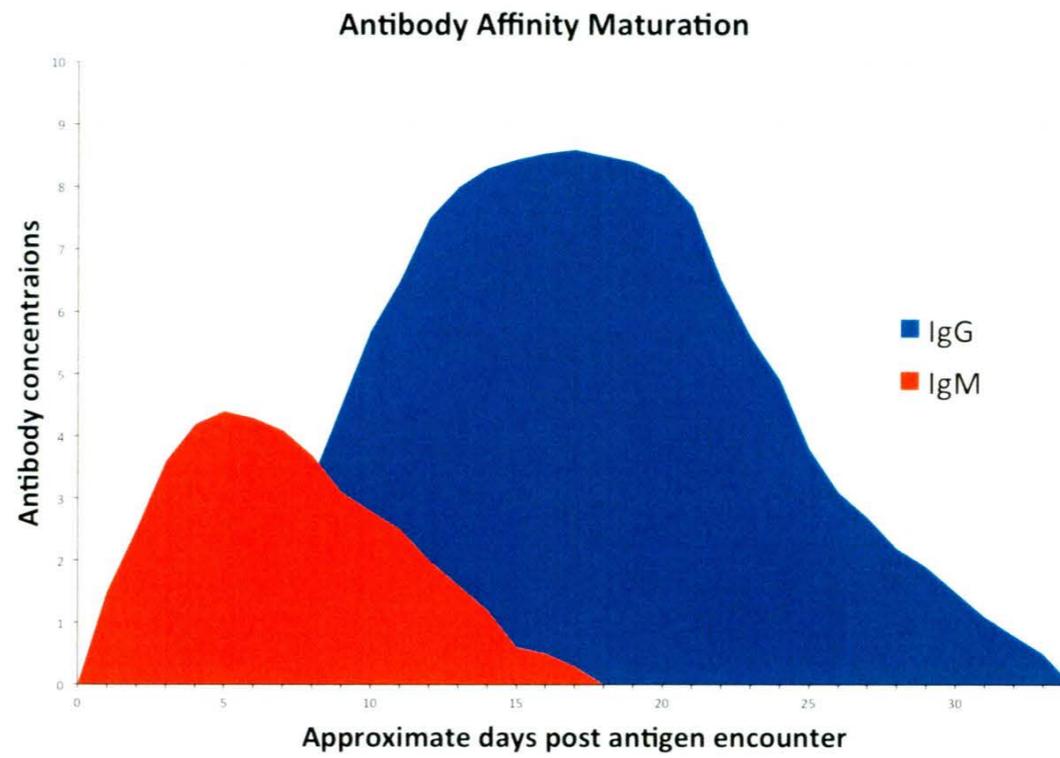
The constant region of an antibody molecule is shown in blue while the variable region located at the N terminus is shown in red. The variable region of antibodies is genetically modified by AID to improve antigen recognition and binding in an adaptive immune response.

J gene segments. The possible rearrangements of these 71 gene segments allow for 6000 possible products. There are two light chains, kappa and lambda, which also recombine in a similar manner. Compounded with imprecise error-prone joining or random insertion and deletion of nucleotides at the borders of the V, D, and J segments, this combinatorial strategy results in millions of different antibody specificities from less than one hundred gene segments (Gellert, 2002).

Naïve B cells initially have a low affinity for any given antigen. To increase the specificity and affinity of the antibody, affinity maturation must take place, which requires secondary antibody diversification (Rajewsky, 1996; Fig. 2). Secondary antibody diversification requires the two AID-driven processes of Somatic Hypermutation (SHM) and Class-Switch Recombination (CSR; Muramatsu et al., 2000; Martin et al., 2002; Okazaki et al., 2002; Revy et al., 2000; Fig. 3). The mutations induced by AID on ssDNA can undergo several repair mechanisms resulting in a full spectrum of mutations in a process called Somatic Hypermutation (SHM). Processing of the mutated base sites can also lead to DNA strand breaks potentially resulting in excision of DNA sequences. When this occurs in the target Ig locus, this is known as Class Switch Recombination (CSR). AID deaminates dC (deoxycytidine, cytidine within the DNA structure) converting it to dU (deoxyuridine, uracil within the DNA structure) at a relatively high rate of  $10^{-3}$ /bp/cell generation, approximately  $10^6$ -fold higher than the spontaneous mutation rate of the human genome (Rajewsky, 1996). Following AID action, the resulting dU in DNA may be

subject to a variety of error-free as well as error-prone DNA repair processes allowing the final mutation spectrum to encompass mutations of any of the four bases, adenine (A), cytosine (C), guanine (G) or thymidine (T) to any of the other three (Di Noia et al., 2007).

While AID randomly mutates throughout the antibody gene, those B cells with higher affinity antibodies will outcompete B cells with lower affinity antibodies for binding to a limited pool of antigen in the germinal centers, thus receiving survival and/or proliferation signals. In this manner, AID-directed mutagenesis of antibody genes followed by cellular selection constitutes affinity maturation of B-cells (Martin et al., 2002; Muramatsu et al., 1999; Muramatsu et al., 2000; Revy et al., 2000; Okazaki et al., 2002).



**Figure 2. Affinity Maturation of B-cells.**

A pictorial representation of antibody levels and “maturation” in response to antigens. IgM isotype antibodies are initially produced and are low affinity. This is followed by an exponential increase in concentration of the more specific antigen binding antibody isotype, IgG. (Adapted from Janeway, 2008).

### 1.3 DNA targets of AID during secondary antibody diversification

AID is a member of the apolipoprotein B mRNA editing enzymes, a catalytic polypeptide-like (APOBECs) family of cytidine deaminase enzymes, of which there are 11 known members in humans (Muramatsu, et al., 1999). AID, a relatively small enzyme, at 24 kDa in size and a length of 198 amino acids in humans, is expressed by activated B-cells undergoing affinity maturation. It was initially presumed to be an mRNA editing enzyme due to homology with other APOBEC family member enzymes that act on RNA (Muramatsu et al., 2000); however, further work showed that it acts solely on single-stranded deoxyribonucleic acid (ssDNA) hotspots (trinucleotide sequence WRC where W = A/T, R = A/C), and not on RNA targets (Klemm et al., 2009; Larijani et al., 2007b; Pham et al., 2003; Sohail et al., 2003). Several lines of evidence substantiate that ssDNA is the sole substrate for AID: a), hypermutation can be induced on transcribed episomal substrates and endogenous genes of hybridomas, fibroblasts, and bacteria with the expression of AID alone, and it is unlikely that the same mRNA target is present in all these species (Martin et al., 2002; Petersen-Mahrt et al., 2002; Yoshikawa et al., 2002); b), AID has been reported to be physically localized at the switch region (S-region) during CSR in B cells, (Chaudhuri et al., 2007; Nowak et al., 2011; Ranjit et al., 2011); c), AID was shown to act on ssDNA and not on RNA or double-stranded DNA (dsDNA) following *in vitro* deamination reactions using purified AID (Bransteitter et al., 2003; Dickerson et al., 2003; Larijani et al., 2007a; Pham et al., 2003); d), the levels of SHM and CSR are diminished by interfering with DNA-specific uracil repair processes (Di Noia et al.,

2002; Rada et al., 2004), indicating that these processes result from the generation of uracil in DNA; e), the absence of Uracil-N-DNA glycosylase (UDG) and MutS protein Homologue 2 (MSH2), two DNA-specific repair factors that process dU only in DNA, causes an altered mutational spectrum at the Ig loci consistent with the lack of uracil processing in DNA; and f), the mutational spectrum and sequence preference of purified AID *in vitro* is nearly identical to that found at the Ig loci in mice and humans (Beale et al., 2004; Klemm et al., 2009; Larijani et al., 2005; Liu et al., 2008; Pasqualucci et al., 2007; Pham et al., 2003; Ramiro et al., 2004; Robbiani et al., 2008; Wang et al., 2004; Xue et al., 2006; Sohail et al., 2003; Yu et al., 2004).

Although it has been well established by several lines of investigation that ssDNA is the target of AID activity, the location(s) and manner(s) in which AID targets ssDNA *in vivo* have not yet been fully elucidated.

The rate of AID-induced mutagenesis at a genetic locus correlates positively with transcriptional status (Peters et al., 1996; Storb et al., 1998; Fukita et al., 1998). SHM and CSR action is eradicated by deletion of Ig promoters, but can be recovered with non-Ig promoters (Fukita et al., 1998). The mechanism proposed for AID access to ssDNA is related to the transcriptional process; in the wake of helix unwinding by the RNA polymerase machinery, ssDNA in the transcription bubble, stemloops and R-loops or supercoiling induced-melted DNA provides AID access to ssDNA (Dayn et al., 1992). Stem loops structures are formed when partial self-complementarity DNA strands are present, allowing "bubbles" of unpaired ssDNA to form. R-loops structures are RNA-DNA hybrids (Thomas, White, & Davis, 1976) and are transiently

formed via preferential pairing of G-rich RNA with the cognate DNA permitting the sister DNA, to be left unpaired. Meanwhile, self-complementary strands are abundant in variable regions resulting in stem loop structures, while GC sequences are found at high frequency in the switch regions (S regions). Thus, various regions of the Ig locus exhibit an unusually high potential for the formation of stem loop and R loop structures during transcription (Yu et al., 2003; Huang et al., 2007). Both can provide ssDNA substrates for AID.

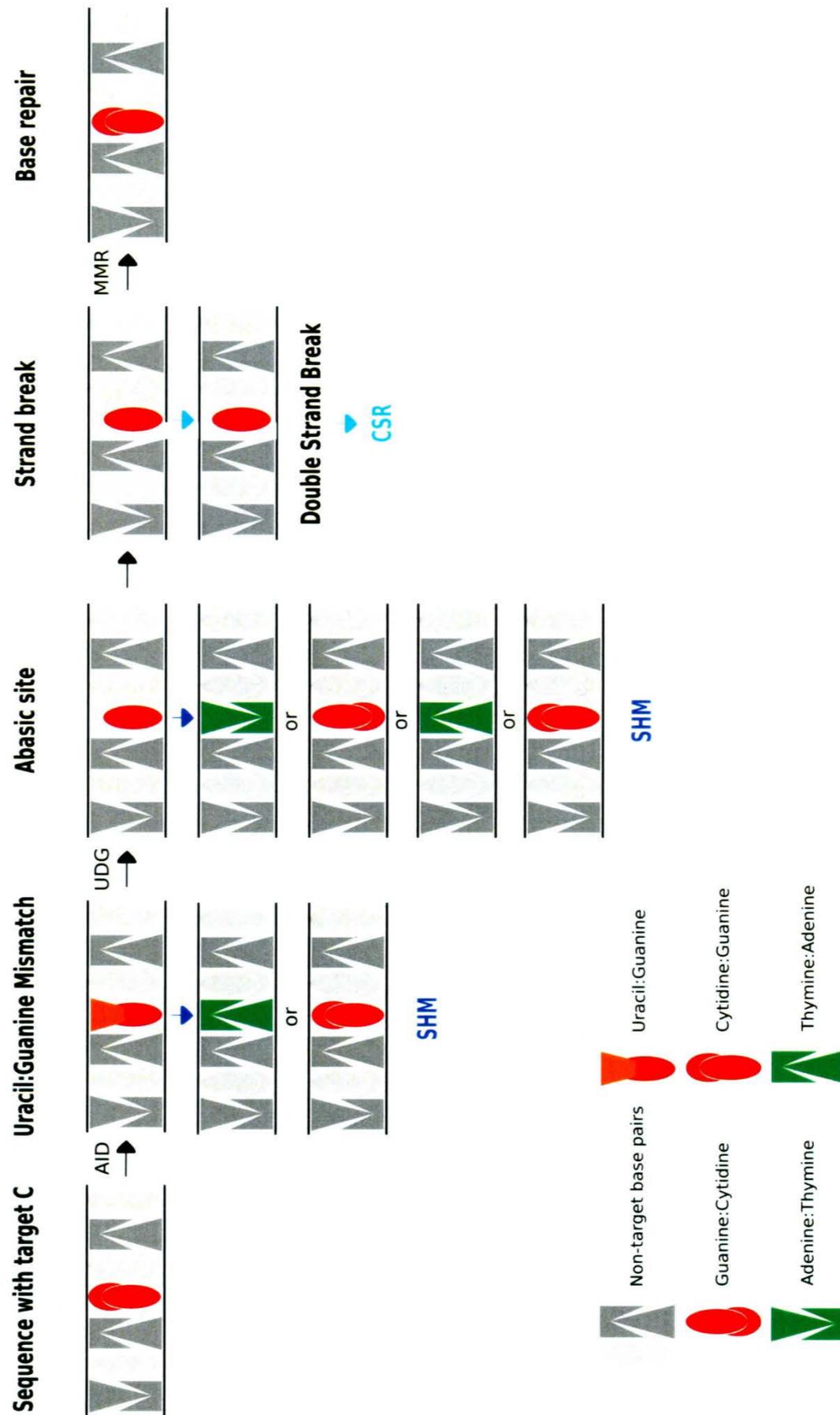
*In vitro* studies have shown AID to have selective activity. AID deaminates dC efficiently in 5-7 nucleotide bubble-type structures (Larijani et al., 2007b) and in R-loop structures (Yu et al., 2005), yet it works inefficiently in stem-loop structures (Larijani et al., 2007b; Ronai et al., 2007; Parsa et al., 2012). Further research showed that AID can mutate both forward and reverse strands of supercoiled plasmid DNA *in vitro* (Shen et al., 2004), similar to the case *in vivo*, where AID induces mutations in the transcribed and untranscribed strands of Ig genes. DNA sequences adjacent to RNA polymerases paused at pre-determined locations undergoing transcription are also mutated by AID (Canugovi et al., 2009). These findings support the discovery that AID mutates DNA containing halted RNA polymerases (Pavri et al., 2010), and taken together, support a model of AID activity wherein AID acts on ssDNA structures of bubble stemloops and R-loops created during the transcription process.

#### **1.4 Pathways mediating the full spectrum of AID mutagenesis**

As AID acts on ssDNA, it deaminates dC to dU, but this sole action can result in the full spectrum of transition and transversion mutations from any of the four bases (A, G, C or T) found in DNA to another base. A transition mutation indicates a purine-to-purine or pyrimidine-to-pyrimidine base change, which is a change from A to G or G to A and from T to C or C to T, respectively. Transversion mutations are purine to pyrimidine or pyrimidine to purine base changes. The different mutation outcomes are due to the action of the various repair pathways that occur following the initial AID-mediated dC to dU conversion (Chadhuri et al., 2004; Neuberger et al., 2005; Fig. 3). Firstly, the dU can go unrepaired and subsequently the DNA replicated, leading to a dC to dT transition mutation. Secondly, the dU:dG mismatch can be recognized by the DNA mismatch repair (MMR) machinery, followed by base excision repair (BER) which can involve error-prone polymerases leading to the introduction of further transition and transversion mutations surrounding the initial dU lesion. Thirdly, dU can be recognized and removed by Uracil-N-DNA glycosylase (UDG) generating an abasic site. As replication continues, transition and transversion mutations can occur across the abasic site (Di Noia et al., 2007). Through this wide variety of downstream mechanisms, the action of AID across the V(D)J-recombined exon and variable portions of the heavy and light chains leads to SHM, CSR and affinity maturation of antibodies.

Repair mechanisms downstream of AID action can lead to abasic sites within ssDNA sequences. Abasic sites are labile and subject to natural or enzymatic

cleavage through apurinic/apyrmidinic endonucleases that cleave the phosphate backbone. Following AID activity, multiple adjacent abasic sites on the two strands can lead to double strand breaks (DSB) in the S-regions located downstream of the constant gene segments of the Ig loci. These DSBs initiate recombination between S-regions of different constant gene segments during CSR, leading to isotype or class switching of antibodies. This occurs through the excision of the first constant gene segment  $C_{\mu}$ , which is then replaced by  $C_{\gamma}$ ,  $C_{\alpha}$ , or  $C_{\epsilon}$  regions. This substitution changes the class or isotype of antibody from the default IgM to IgG, A or E during an immune response (de Yébenes & Almudena, 2006). Individuals with genetically inherited defective mutations of AID suffer from Hyper-IgM type immunodeficiency, and they are unable to mount a proper antibody response against pathogens due to the lack of SHM and CSR pathway initiation. Even common infections can become severe or even lethal for these individuals, which underlines the importance of AID activity for an efficient immune response (Durandy et al., 2005; Durandy et al., 1996).



**Figure 3. The mechanism of Somatic Hypermutation (SHM) and Class Switch Recombination (CSR) pathways.**

AID converts a target C to U within a nucleotide sequence, which is recognized by Mismatch Repair machinery (MMR). U is either replaced by T, repaired to C or removed. If the U is removed by uracil-N-DNA glycosylase (UDG), the abasic site can be filled with either A, G, T all resulting in SHM. If the phosphate backbone of the DNA at the abasic site is cleaved, it can either be repaired or in the event of multiple cleaved sites called double-stranded breaks (DSBs) CSR can occur. (Adapted from de Yébenes, 2006)

### 1.5 The role of AID in tumorigenesis

AID is a DNA mutator that greatly enhances the adaptive immune response; however, it has been shown to mutate off-target loci quite promiscuously. When AID acts on non-Ig genes, it can result in mutations or DSBs in a similar fashion as it does in the Ig genes, leading to dysregulation of gene expression through mutations or chromosomal translocations, resulting in cancers such as B cell lymphoma and leukemia (Kuppers et al., 2001; Kuppers, 2005).

Studies with murine models have elucidated many links between AID activity and oncogene translocations. The IL-6 transgenic BALB/c strain is a well-studied mouse model that develops spontaneous plasmacytomas due to a *c-Myc/IgH* [v-Myc myelocytomatosis viral oncogene homlog (avian)] translocation, mirroring the translocations in Burkitt's lymphoma in humans. When these mice are AID deficient, *c-Myc/IgH* translocations are abolished, as are DSBs at the *IgH* and *Myc* loci (Ramiro et al., 2004). This provides direct evidence that AID is necessary for DSB formation at *IgH* and *c-Myc* translocation sites (Robbiani et al., 2008). AID deficiency was also shown to decrease the frequency of lymphoma development in the *Bcl6* (B-cell CLL/lymphoma 6) transgenic mouse model, in which clonal splenic lymphomas result from *Bcl6* deregulation (Pasqualucci et al., 2007).

Sequencing a large number of genes from AID-deficient and sufficient mice showed that AID mutates up to 50 % of all transcribed genes expressed in B cells (Lieu et al., 2008). Studies measuring the rate of AID-induced mutations in

transgenes throughout the genome, as compared to the rate of AID-induced mutations at the Ig locus, have shown that although AID can mutate genome-wide, it preferentially targets the Ig locus 90 % of the time. Therefore, the highest rate of mutation found in non-Ig transgenes may approach 10 %; however, this is adequate to produce mutations and translocations in non-Ig genes (Robbiani et al., 2008), representing a significant risk to genome integrity.

#### **1.6 Clinical mutations and translocations of oncogenes and AID activity**

In the Western Hemisphere, the vast majority of lymphomas originate in mature B cells (Kuppers, 2005). Many of these cancers have been found to have mutations and translocations of several proto-oncogenes - *cMYC/IgH* in Burkitt's Lymphoma (BL), *BCL2/IgH* in follicular lymphoma (FL), *BCL-6/IgH* in diffuse large B cell lymphoma (DLBCL) - strongly suspected to be due to aberrant AID activity. Promiscuous AID activity has also been suggested to result in other cancers including multiple-myeloma and chronic lymphocytic leukemia (CLL; Shen et al., 1998; Pasqualucci et al., 1998; Kuppers et al., 2001). In human studies, mutations and translocations in signature AID hotspot motifs (WRC) have been reported in oncogenes in lymphoma cells. For instance, *BCL6* gene disruption is associated with various malignancies. In DLBCL, *BCL6* is translocated in 35 % of cases and mutated in 73 % (Migliazza, et al., 1995). AID generated *c-Myc-IgH* translocations have been shown to cause BL in mice. Similarly, point mutations in tumour suppressor genes

such as retinoblastoma (*Rb1*), *p19* and *p53* are displayed in various B-cell tumours and can affect the progression of the disease, as well as its development (Kuppers et al., 2001).

Many oncogenes *c-Myc*, *BCL-6*, *FAS* (Fas cell surface death receptor), *PIM-1* (Pim-1 oncogene), *PAX-5* (Paired box 5), and *RHOH* (Ras homolog family member H) in lymphomas harbour mutations and/or translocations in AID hotspots (Gordon et al., 2003; Muschen et al., 2000; Pasqualucci et al., 1998; Pasqualucci et al., 2000; Shen et al., 1998). The role of AID in the commencement of these diseases is further evidenced by the fact that the majority of tumours originate from the centroblast or post-centroblast stage of B cell development, the stage in B cell development where AID is expressed. Lastly, other translocations of proto-oncogenes almost always involve a translocating partner within the Ig heavy chain locus such as the JH or S-region, the very region that AID targets to mediate SHM and CSR (Kuppers et al., 2001; Willis et al., 2000).

AID expression levels can also be indicative of human prognosis of CLL and DLBCL, where AID levels are associated with poor prognoses for patients (Heintel et al., 2004; Lossos et al., 2004; McCarthy et al., 2003; Reiniger et al., 2006). AID activity is believed to further mutate and destabilize the genomes of lymphomas. For instance, levels of AID were correlated to Imatinib resistance in CML. The drug resistance was shown to arise from hypermutations in tumour suppressor and/or DNA repair genes after the transduction of AID into tumour cells (Klemm et al.,

2009). While some progress has been made in the down regulation of oncogenesis by reducing CSR (Gazumyan et al., 2011) a better understanding of AID targeting could enhance treatment of cancers of B-cell origins. With future insight of the mechanism of AID DNA editing, an inhibitor could be developed as a therapeutic option for patients with cancer resulting from aberrant AID activity, reducing progression of the disease while maintaining immune function.

### **1.7 The biochemical features and putative structure of AID**

Largely due to difficulties with expression and purification of highly genotoxic enzymes such as AID, many aspects of its structure, enzymatic mechanism(s), and modes of AID binding to DNA remain unknown. Although the enzymatic mechanism has not been fully established, evidence suggests that AID acts on dCs that are flipped outside of a DNA strand and no longer aligned in stacking formation with neighbouring bases. Support for this claim includes the precedence of base flipping mechanisms in other deaminase enzymes, including tRNA adenosine deaminase (TadA) and adenosine deaminase acting on dsRNA, (ADAR; Losey et al., 2006; Stephens et al., 2000) and in DNA modifying enzymes like methyltransferases and uracil glycosylases (Holz et al., 1998; Slupphaug et al., 1996; Jiang et al., 2002). Based on our model of AID (Fig. 4), dC would have to be flipped out of the DNA sequence away from adjacent bases to fit into AID's catalytic pocket, a process which occurs for a related enzyme in the APOBEC family, APOBEC3G

(Deng, et al., 2010). Moreover, the size of the enzyme is too large to fit within the small 5-7 nucleotide ssDNA bubble structures that AID preferentially deaminates (Larijani et al., 2007a). Within these small nucleotide bubbles, space is limited and dC is more likely to be unstacked and flipped out of the sequence due to torsional constraints with adjoining bases, further supporting this model.

Biochemical analysis of AID enzymatic catalysis shows an unusually slow rate of reaction compared to most other enzymes. AID has a rate of approximately one deamination every few minutes (Larijani et al., 2007b) as opposed to hundreds to thousands of reactions per second for most enzymes. Remarkably, the aforementioned ADAR deaminase also has a slow enzymatic velocity, and utilizes a base-flipping mechanism on its dsDNA substrate, suggesting that a slow enzymatic velocity may be an inherent feature of a base-flipping mechanism (Stephens et al., 2000; Yi-Brunozzi et al., 2001). While this sluggish rate may be due to the *in vitro* AID production via bacterial expression purification processes (Coker et al., 2006), it has been confirmed by another research group using an independent AID expression system, (Pham et al., 2003) validating slow activity as an innate characteristic of AID enzymatic activity. A slow rate could be a protective mechanism guarding the genome against aberrant AID mutations and directing AID's preference for the Ig loci over off-target genes. This idea is reinforced by the higher rate of activity evidenced by high levels of mutations and chromosomal breaks induced in bacterial lines by mutant human AID, produced through random mutagenesis and selected by bacterial papillation assay (Wang et al., 2009). The intrinsic properties of AID may

very well have evolved as an essential protective mechanism against its promiscuous nature.

AID has an unusually high ssDNA binding affinity in the nM range, several orders of magnitude above typical DNA-binding enzymes. This characteristic, coupled with a long complex half-life of approximately 8 minutes, provides an explanation for the slow catalysis and *in vitro* enzymatic processivity of AID (Larijani et al., 2007b). AID processivity refers to the fact that in a pool of DNA substrates, AID mutates very few targets, but these targets are deaminated quite heavily. This has been suggested to be due to the high DNA binding affinity, such that there is a much higher likelihood of AID dissociating and re-associating with the same substrate once it has acted on ssDNA. This could potentially facilitate further recruitment and interaction with other proteins involved in SHM and CSR processes (Ranjit, et al., 2011).

Variation in the activity of AID coincides with the different binding affinities of AID for each different ssDNA substrate structure studied (Larijani et al., 2007); thus, the binding process is a significant factor when determining AID's enzymatic velocity. This effect was found to occur due to residue mutations affecting both activity and binding affinities. These mutations were predicted to be within the core of the structure of AID (Wang et al., 2009). Such evidence would imply that there are likely two sets of regions of AID that are responsible for regulating its activity: a select few residues found near the catalytic zinc (Zn) coordinating site that engage in

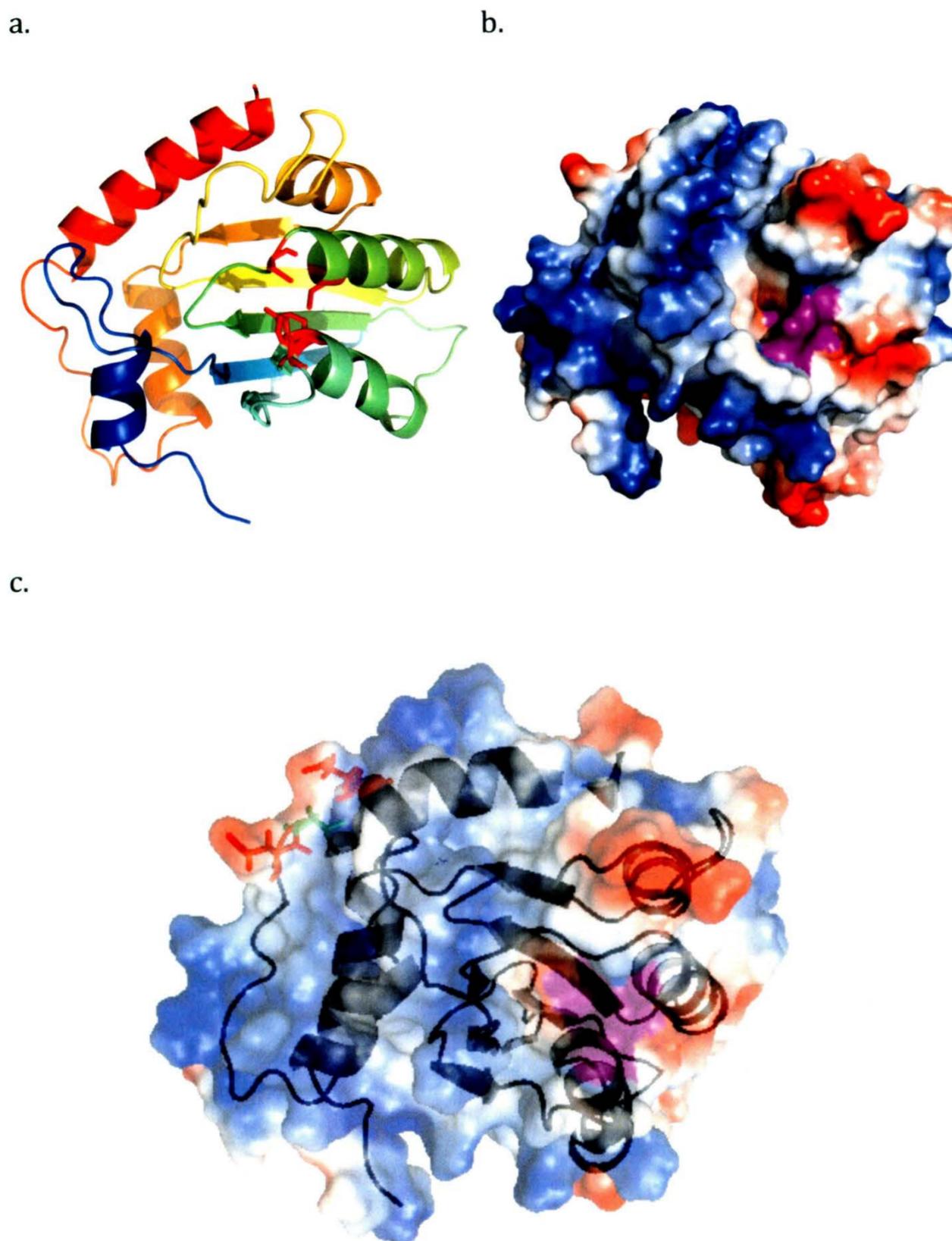
enzymatic deamination, and several residues found on the surface involved in ssDNA binding and potential multimerization.

### **1.8 Modeling the putative 3-dimensional structure of AID**

To date, the structure of the AID protein has not been resolved. In order to study its mechanism and substrate binding functionalities, the structure is of utmost importance. Based on the experimentally determined X-ray structure of the protein family member APOBEC3G, the core and surface topography structures of AID were modeled; from which it was hypothesized there are two potential ssDNA binding grooves. As with APOBEC3G, the model shows AID as having a core of five beta strands surrounded by six alpha helices (Fig. 4). There are several reasons to believe that this model is accurate: Firstly, the model recapitulates the canonical position of the Zn-coordinating residues [two cysteines (C) and one histidine (H)], which are contributed by the ends of opposing alpha helices in all other deaminases. In the AID model, the residues H56, C87, and C90 are located proximal to a proton-donor glutamic acid (E58). Thus, the model generates a viable catalytic site akin to that of other crystallized deaminases (Ireton et al., 2003; Carter Jr., 1995; Conticello et al., 2007; Chen et al., 2008; Holden et al., 2008). In support of this notion, mutagenesis of H56, C87, C90, and E58 leads to an inactive enzyme (Dickerson et al., 2003; M. Larijani, unpublished observations). Secondly, the model recapitulates secondary structure predictions based on the solved structure of other bacterial, yeast, and

human cytidine deaminases (Ko et al., 2003; Losey et al., 2006; Chen et al., 2008; Holden et al., 2008). Thirdly, analysis of the AID surface topography in this model reveals a highly charged surface with a large proportion of positive residues, such as Arginine (R) and Lysine (K), consistent with our previous finding of exceptionally high affinity binding of AID to negatively charged ssDNA. This feature may also explain difficulties in the purification (and consequently crystallization) of AID, necessitating relatively bulky tags such as GST (Sohail et al., 2003; Larijani et al., 2007a) and MBP (Kohli et al., 2009). Fourthly, two putative DNA-binding grooves were observed passing over the catalytic pocket (Fig. 4). These grooves are ~10 Angstroms (Å) in width at their widest points and are lined with a high proportion of positively charged residues. Given that double-helical dsDNA is ~ 20 Å wide, the size of these grooves is consistent with the restriction of AID activity to ssDNA (Pham et al., 2003; Sohail et al., 2003; Larijani et al., 2007a) and its makeup is consistent with high affinity binding to ssDNA (Larijani et al., 2007a). Lastly, a recent random mutagenesis study identified dozens of AID single, double, triple, and multiple mutants with altered enzymatic velocities and the position of many, but not all, of these residues are predicted to be proximal to the catalytic site (Wang et al., 2009).

The prediction model shown in Figure 5 may be a relatively accurate prediction of an AID monomer structure, but it does not account for the quaternary structure of AID. Though the stoichiometry of enzymatically active AID has been a contentious issue in the literature, recent evidence suggests that the native



**Figure 4. Human wild-type AID (a.) ribbon diagram, (b.) surface diagram, and (c.) combined diagram.**

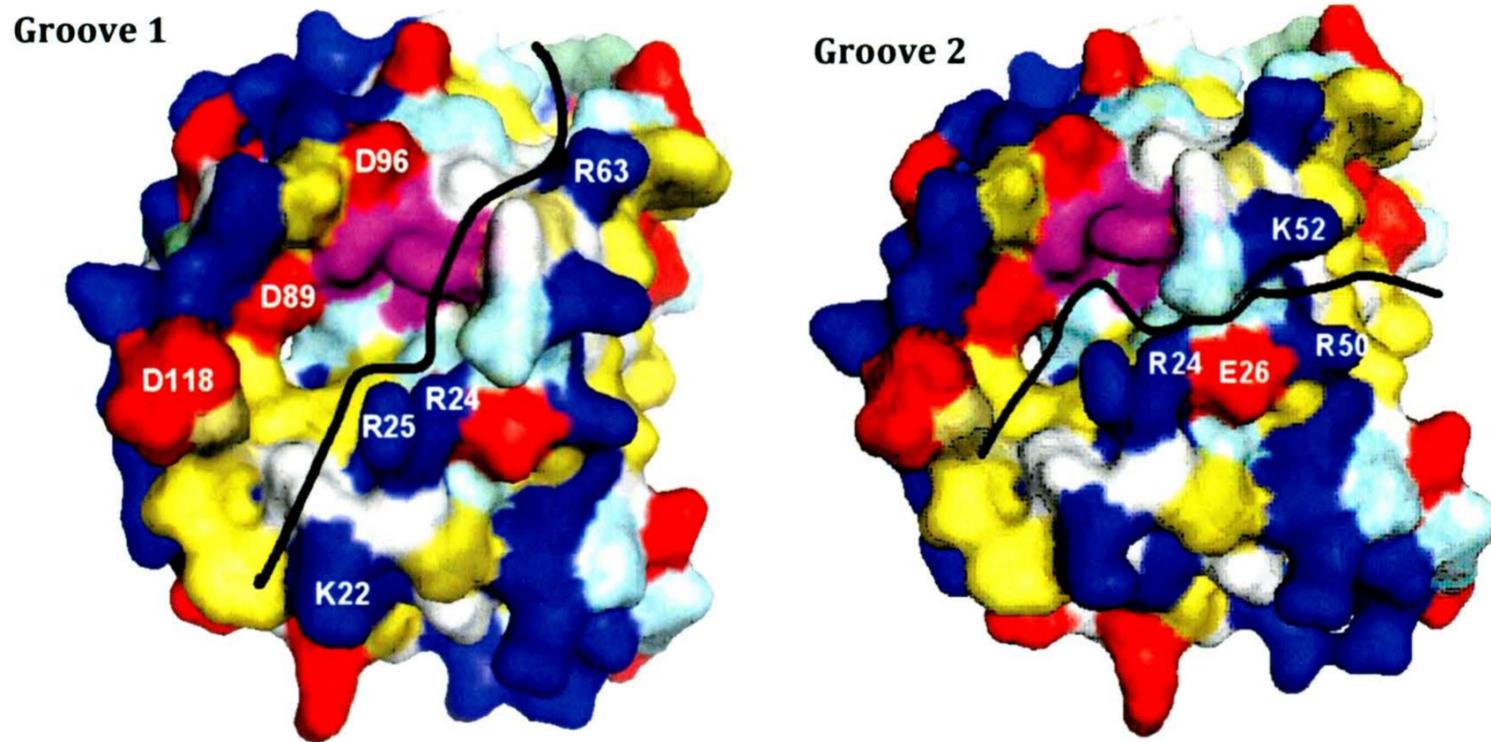
(a.) Ribbon model, (b.) surface model and (c.) combined model was predicted and generated using Pymol v1.30 based on the experimentally determined X-ray structure of the protein family member. As with APOBEC3G, the model shows AID as having a core of five beta strands surrounded by six alpha helices, coloured for visual distinction. The surface model is coloured according to residue; positively charged residues are shown in blue, negatively charged residues are shown in red, aromatic residues are yellow and purple is the Zn binding pocket.

stoichiometry of AID is likely a dimer or tetramer. Purified AID appears to migrate as a tetramer (Dickerson et al., 2003) and when two differently tagged versions of AID were co-expressed in the same cell, co-immunoprecipitations indicated AID forms a dimer or multimer (Ta et al., 2003). By separating purified GST- or His-tagged AID on a size-exclusion column and measuring the activity of various fractions, the highest activity levels in fractions consistent with the size of a tetramer was found (Larijani et al., 2007a). APOBEC2 is 75 % identical to AID, and exists as a tetramer formed by head-to-head interaction of two dimers (Prochnow et al., 2007). Mutational studies disrupting putative dimerization domains of AID based on homology of primary sequence with other family members, as well as the residues involved in multimerization of APOBEC2, have yielded an enzyme with decreased activity (Prochnow et al., 2007). Whilst the majority of the literature supports a model where AID is a tetramer, there are studies that challenge this notion. For instance, a study using atomic force microscopy measurement of free AID in solution found that enzymatically active AID acts as a monomer (Brar et al., 2008). For the final word on the tertiary and quaternary structure of AID, we await the solution of AID structure by NMR (Nuclear magnetic resonance) or X-ray crystallography.

### **1.9 Identification of two putative ssDNA binding grooves**

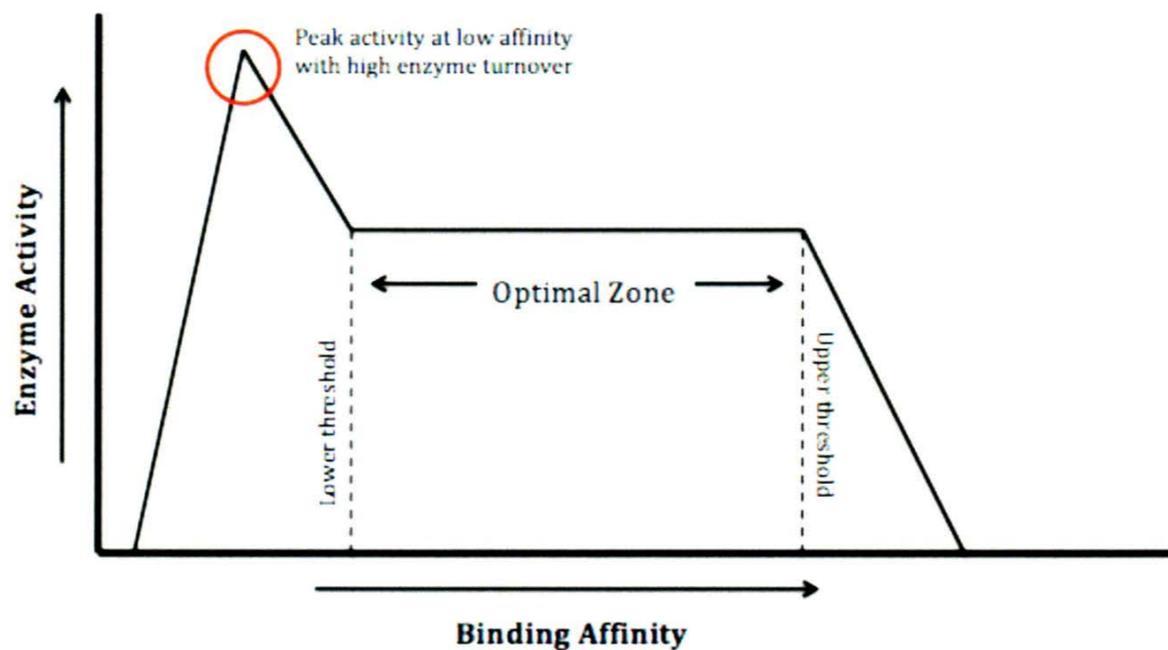
The AID structural model revealed two putative ssDNA binding grooves along the surface – Groove 1 and Groove 2 (Fig. 5). We chose these two grooves due to

their close proximity to the zinc-coordinating site, their correct diameters to accommodate ssDNA, but not dsDNA ( $\sim 6 \text{ \AA}$ ), and the presence of multiple positively charged residues, which are likely candidates for interaction with the negatively charged DNA. Both putative binding grooves pass over the catalytic pocket of AID. Based on these features, it was hypothesized that positively charged residues within this groove interact with DNA and are involved in binding and activity. It was further reasoned that if positively charged residues were mutated to amino acids with negative or neutral charges, ssDNA binding and enzymatic activity would be altered such that binding and therefore activity would be reduced. The reverse should also be true, whereby negatively charged residues repel negatively charged DNA and mutating these to positively charged or neutral residues would more strongly attract ssDNA, which should result in higher affinity binding. Since work in our lab has previously shown that higher ssDNA binding affinity by AID correlates with higher deamination activity, we would also expect to produce mutated enzymes with higher deamination activity; however, it is possible that there is an optimal ssDNA binding affinity, beyond which, higher affinities would actually diminish the enzymatic activity, as the enzyme would remain bound to each substrate for a longer period and reduce the rate of the catalytic cycle (Fig. 6).



**Figure 5. AID Model and Putative Binding Grooves 1 & 2.**

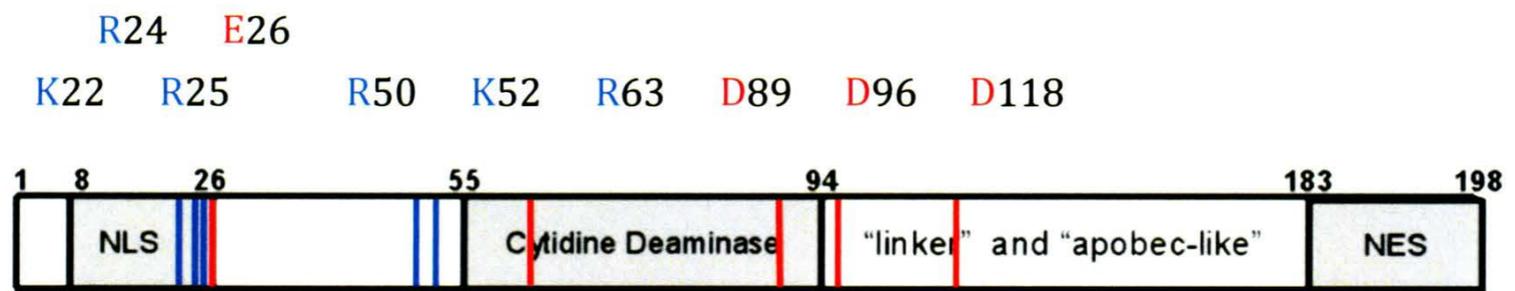
Proposed AID structure was modelled on AID sequence and sequentially homologous APOBEC-1 structure. Positively charged residues are shown in blue, negatively charged residues are shown in red, aromatic residues are yellow and purple is the Zn binding pocket. The track of Groove 1 and Groove 2 are highlighted by black lines. The AID model was rotated slightly to highlight Groove 2.



**Figure 6. Binding Affinity versus Enzyme Activity.**

A proposed relationship between the binding affinity of enzyme and the resulting activity. At lower binding affinity an enzyme can operate on a substrate and easily progress to the next substrate. The activity increases with binding affinity and maximizes at a peak. These high levels can cause errant over activity. As the binding affinity increases further the enzyme is less able to release substrate and act on another. As a result, enzyme activity decreases. The enzyme activity level will reach a level that is neither too promiscuous nor too low when the binding affinity is within this "optimal zone". Binding affinity above this threshold will cause enzymes to dock substrate too tightly and result in a low substrate turnover rate and consequently, a very low activity level.

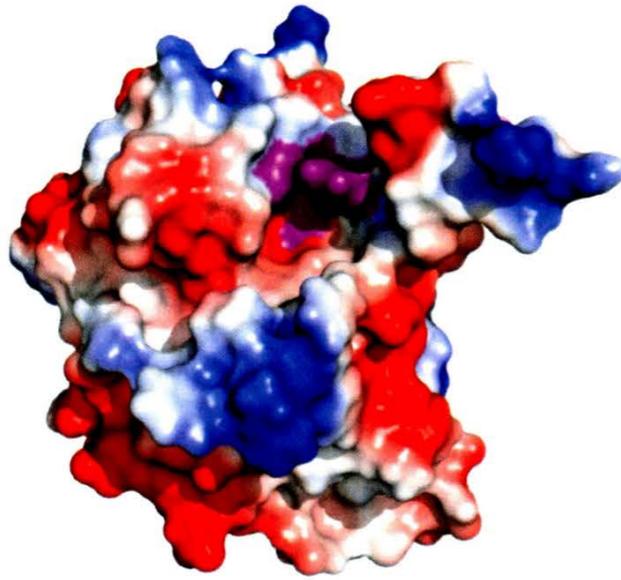
To study this enzyme, we have modeled the structure based on the linear structure of AID (Fig. 7) homologous family member enzymes APOBEC3G and APOBEC2 whose structures have been experimentally determined by NMR or X-ray crystallography (Fig. 8; Prochnow et al., 2007; Chen et al., 2008; Holden et al., 2008). We then identified two putative DNA binding grooves and made various point and multi-residue mutations in and around each of these two grooves, illustrated in a linear representation of the sequence of AID (Fig. 7). We then studied the enzymatic velocities and binding affinities of the mutant enzymes in order to ascertain the involvement of each putative groove in ssDNA binding. Using enzyme mutagenesis, the involvement of specific residues in AID activity and binding can be effectively explored, as shown in APOBEC and other enzyme studies (Kao et al., 1996; MacGinnite et al., 1995; Mu et al., 2012; Hu et al., 2013).



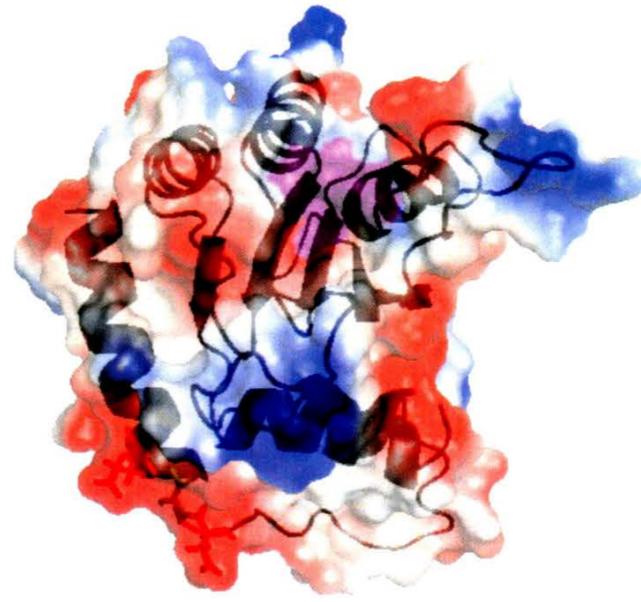
**Figure 7. Linear AID structure.**

A linear representation of AID with known domains indicated. NLS = Nuclear Localization Sequence. NES = Nuclear Export Signal. The amino acid residues selected for mutation are indicated by blue (positively charged) and red (negatively charged) bands.

a.



b.



**Figure 8. (a.) APOBEC-3G Surface Model and (b.) combined Surface and ribbon model.** Proposed charge surface model and ribbon and charge surface model was predicted using the AID sequence and sequentially homologous APOBEC-1 structure. Positively charged residues are shown in blue, negatively charged residues are shown in red and purple is the Zn binding pocket.

### **1.10 Hypothesis**

Using the 3D model of AID (Fig. 5), it is hypothesized that positively charged residues lining Groove 1 and Groove 2 adjacent to the catalytic pocket are relevant for negatively charged ssDNA substrate binding and control of AID activity. If these positively charged residues are altered, enzyme activity and binding should likewise be altered. Changing the positive residues to those with neutral or negative charge is expected to reduce binding and activity of the resulting mutant enzymes.

Conversely, converting negatively charged residues within the putative binding grooves to positively charged residues is expected to increase binding to negatively charged substrate and therefore increase enzyme activity.

## Chapter 2 Materials and Methods

### 2.1 Overview of AID mutagenesis and analyses:

The wt AID sequence was altered using site-directed mutagenesis as a strategy to create individual or compound AID mutants. The modified strands were subsequently verified with sequencing by *Macrogen* and base analysis with DNA Star software. Mutant AID sequences were translated into protein and the purified enzyme was assayed for overall activity, substrate binding and subsequent kinetic assays, if sufficient AID activity was retained. Figure 9 outlines the overall experimental process.



#### Figure 9. Mutant Analysis.

Mutant AID enzymes were designed, produced and tested using various assays to study enzyme activity and binding.

### 2.2 Site Directed Mutagenesis

The positively charged amino acids, lysine (K) and arginine (R), within putative binding grooves 1 and 2 were chosen for mutation to a negatively charged amino acid, aspartic acid (D), or to the relatively inert residue alanine (A). Negatively charged amino acids, aspartic acid, and glutamic acid (E) were likewise converted to arginine. If associated with DNA binding, these mutations should disrupt AID activity between the mutant AID protein and the ssDNA substrate.

Primer pairs of approximately 30 nucleotides were designed for each amino acid residue under investigation (Table 1). GC clamps were included at either end of the primer to ensure specific binding to the mutation site. Lyophilized primers (IDT) were resuspended in MilliQ H<sub>2</sub>O to 100 nM; 12.5 ng of primer and of its complementary strand were used in a 50 µL site-directed mutagenesis reaction containing 32.5 µL MilliQ water, 10 µL High Fidelity 5 X PCR Buffer, 1 µL High Fidelity DNA polymerase (NEB), 1.5 µL 10 mM dNTP mixture, and 25 ng of pGEX-5x-3 GST H long linker AID plasmid template. The thermal cycler steps were as follows: 98 °C for 1 min, 98 °C for 30 s, 50 °C for 45 s, and 72 °C 6 min repeated 30 times, followed by a final extension step of 72 °C for 10 min followed by a 15 °C hold. The reactions were stored overnight at 4 °C and digested the following morning with 0.7 µL DpnI methylation dependent restriction endonuclease for 2 h (hours) at 37 °C. Samples were then dialysed to remove impurities and excess salts using MilliPore 0.025 µm membrane filters on MilliQ water for 1 h at room temperature. The purified samples were transferred to 1.5 mL microcentrifuge tubes and concentrated under vacuum using a SpeedVac Concentrator (*Savant*). Fifty µL of Top10 DH5α cells were transformed with 5 µL of the concentrated PCR product, stirred gently, and left to sit on ice for 20 min. They were then heat shocked at 42 °C for 90 s followed by an addition of 200 µL of Super Optimal Broth with Catabolite repression (SOC) media (Super Optimal Broth with added glucose used for bacterial transformation). The samples were incubated at 37 °C in a shaker for 1 h, after which 250 µL was plated on LB (Lysogeny broth or Luria-Bertani medium)-

ampicillin (37 g LB + Agar/1 L MilliQ dH<sub>2</sub>O, 100 µg/mL ampicillin) and incubated overnight at 37 °C. Five to twenty of the resulting colonies were selected for each mutant AID and grown in 5 mL LB + ampicillin [100 µg/mL] media overnight at 37 °C and gently stirred at 225 rpm. Each culture was transferred to a 1.5 mL micro centrifuge tube and pelleted by centrifugation. Plasmid DNA was extracted using a QIAprep Spin Miniprep kit (*Qiagen*) and microcentrifuge. The resulting plasmids were verified for cloned inserts via EcoR1 restriction digest (0.3 µL, 2 h at 37 °C). Mutations were confirmed by DNA sequencing (*Macrogen Inc.*, Korea) and confirming amino acid signals using DNA Star software. Multiple mutants (Table 2) were created through site directed mutagenesis using plasmid of a confirmed mutant as the template with the designed primers of a secondary mutant sequence.

---

**Table 1. Designed Primers**

Primer sequences and complementary strands were designed using approximately 30-nt bases of wt AID sequence. Point mutations were introduced by conservatively changing the coding bases for each point mutation to preserve maximum amount of wt AID sequence.

---

<b>K22D</b>	<b>ccgctgggctgacggtcggcgtgagacctacc</b>
<b>K22D</b> comp	ggtaggtctcacgccaccgtcagcccagcgg
<b>K22A</b>	<b>ccgctgggctgcgggtcggcgtgagacctacc</b>
<b>K22A</b> comp	ggtaggtctcacgccaccgtcagcccagcgg
<b>R24D</b>	<b>ccgctgggctaagggtgaccgtgagacctacc</b>
<b>R24D</b> comp	ggtaggtctcacggtcacccttagcccagcgg
<b>R24A</b>	<b>ccgctgggctaagggtgcgcgtagacctacc</b>
<b>R24A</b> comp	ggtaggtctcacgcgacccttagcccagcgg
<b>R25D</b>	<b>ggctaagggtcgggatgagacctacctgtgc</b>
<b>R25D</b> comp	gcacaggtaggtctcatcccacccttagcc
<b>R25A</b>	<b>ggctaagggtcgggctgagacctacctgtgc</b>
<b>R25A</b> comp	gcacaggtaggtctcagcccacccttagcc
<b>E26R</b>	<b>ggctaagggtcggcgtaggacctacctgtgctacg</b>
<b>E26R</b> comp	cgtagcacaggtaggtcctacgccacccttagcc
<b>E26A</b>	<b>gctaagggtcggcgtgacacctacctgtgc</b>
<b>E26A</b> comp	gcacaggtaggtcgcacgccacccttagc
<b>R50D</b>	<b>ggactttggttatcttgacaataagaacggctgcc</b>
<b>R50D</b> comp	ggcagccgttcttattgtcaagataaccaaagtcc
<b>R50A</b>	<b>ggactttggttatcttgccaataagaacggctgcc</b>
<b>R50A</b> comp	ggcagccgttcttattggcaagataaccaaagtcc
<b>K52D</b>	<b>ggttatcttcgcaatgacaacggctgccacgtgg</b>
<b>K52D</b> comp	ccacgtggcagccggtgtcattgcaagataacc

---

---

K52A	ggttatcttcgcaatgcgaacggctgccacgtgg
K52Acomp	ccacgtggcagccgttcgcattgcgaagataacc
R63D	ggaattgctcttcctcgactacatctcggactggg
R63Dcomp	cccagtccgagatgtagtcgaggaagagcaattcc
R63A	ggaattgctcttcctcgcctacatctcggactggg
R63Acomp	cccagtccgagatgtaggcgaggaagagcaattcc
D89R	ggagcccctgctaccgctgtgcccgacatgtgg
D89Rcomp	ccacatgtcgggcacagcggtagcaggggctcc
D89A	ccgctacatctcggactgggccctagaccctggcc
D89Acomp	ggccacatgtcgggcacagcgtagcaggggctcc
D96R	gcccgacatgtggcccgtttctgaggggaacc
D96Rcomp	gggttcctcgcagaaagcgggccacatgtcgggc
D118R	cctctacttctgtgagcgcgcaaggctgagcc
D118Rcomp	ggctcagccttgcggcgctcacagaagtagagg

---

### 2.3 AID Protein Purification

Protein production began with transformation of 50  $\mu$ L *Escherichia coli* BL21 (DE3) cells with 5  $\mu$ L of sequenced AID plasmid, letting the cells sit on ice for 20 min. They were then heat shocked for 90 s at 42 °C and supplemented with 200  $\mu$ L of SOC media before incubation at 37 °C and 225 rpm for 1 h. Two hundred and fifty  $\mu$ L of culture was plated on LB-ampicillin and incubated at 37 °C overnight. The next day, a 250 mL LB-ampicillin (100  $\mu$ g/mL) culture was inoculated with one colony of glycerol stock of transformed DE3 cells, and then incubated at 37 °C in an orbital shaker (*Forma Scientific*) until the culture reached the log phase of growth. Protein production was induced in the cells with 250  $\mu$ L 1M isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) to activate the *lac* gene promoter. The culture was then incubated overnight at 16 °C and approximately 180 rpm in a PsychoTherm controlled environment incubator shaker. The next morning, the culture was transferred to a 250 mL Nalgene polypropylene bottle and pelleted using a Sorvall Evolution RC Ultra Centrifuge at 4°C, 5000 rpm for 12 min. After decanting the supernatant, the pellet was re-suspended in 20 mL of cold 1 X Phosphate buffered saline (PBS) and transferred to a 50 mL Falcon tube on ice.

A French pressure cell press (*Thermospectronic*) was used to lyse the re-suspended cells twice, rinsing the French pressure cell with cold 1 X PBS before and after use. The lysate was ultra centrifuged in a Nalgene polypropylene Falcon tube at 5000 rpm and 4 °C for 11 min (minutes) to remove cell waste. The AID protein was

isolated from the lysate supernatant following application to Glutathione sepharose beads (*Amersham*) and subsequently eluted as per the manufacturer's recommendations. Fractions of the elution were collected at 4 °C in 500 µL volumes of 15-30 aliquots for each protein and kept on ice. The fractions were then tested for optical density (OD), protein concentration and purity by nanodrop spectrophotometry (*Thermo Scientific*) at 260 and 280 nm wavelengths. Fractions containing more than 1 mg/mL of protein were combined and dialyzed using Snakeskin<sup>®</sup> Pleated Dialysis Tubing (*Thermo Scientific*) in Dialysis Buffer (1mM DL-Dithiothreitol, 20mM Tris-HCl and 100 mM NaCl) overnight at 4 °C. The buffer was replaced in the morning to dialyze the fractions 3-5 h further, after which the purified protein was aliquoted 50 µL per 1.5 mL micro centrifuge tube and flash frozen in liquid nitrogen for 30 s. The aliquots were then stored at -80 °C.

Each preparation of purified GST-AID protein was run against Bovine Serum Albumin (BSA) standards on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie staining to test for protein concentration, purity, and expression levels. AID protein concentrations were thus determined in comparison to BSA titrations and used to equalize AID volumes for assay use. Deamination and Electrophoretic Mobility Shift Assays (EMSA) were performed with AID to test for activity and complex formation with substrate.

## 2.4 AID Substrate Preparation

The 56 base pair long ssDNA substrate, H2S2bub7, was prepared with a 7-nucleotide single-stranded target site known as a bubble, containing one target cytidine. The bottom strand of the substrate contains the target site trinucleotide sequence thymidine-adenosine-cytidine (TAC) and was purified by High Performance Liquid Chromatography. The top strand used was BHagtop, purified by desalting. Their sequences are as follows, with \* indicating the target cytidine:

BHagtop:

5'-AGA TCC TGC CCC GGC ACT TCG CCC GGG TTT TTC CAG TCC CTT CCC GCT TCA  
GTG AC- 3'

HS2bub7 (TAC):

5'-GTC ACT GAA GCG GGA AGG GAC TGT GTA C\*TT CCG GGC GAA GTG CCG GGG CAG  
GAT CT- 3'

The substrate was radioactively labelled using the following reaction: 1  $\mu\text{L}$  of 2.5 pmol/ $\mu\text{L}$  TAC, 1  $\mu\text{L}$  10X polynucleotide kinase (PNK) buffer, 1  $\mu\text{L}$  PNK enzyme, 3  $\mu\text{L}$  15  $\mu\text{Ci}/\mu\text{L}$  ATP- $[\gamma\text{-}^{32}\text{P}]$ , and 4  $\mu\text{L}$  of autoclaved MilliQ water. Upon mixing, the reaction was incubated at 37 °C for 53 min then at 65 °C for 10 min to deactivate the enzyme. It was then brought to 4 °C, briefly centrifuged, mixed with 10  $\mu\text{L}$  of 1X Tris-EDTA (TE) buffer and separated from unlabelled oligonucleotides using a mini Quick Spin Column (*Roche*). The reaction was applied to the column and spun at 3000 rpm for 3 min. To the resulting flow through, 5  $\mu\text{L}$  1 M KCl, 3  $\mu\text{L}$  2.5 pmol/ $\mu\text{L}$  BHagtop

and adequate autoclaved MilliQ water was added to a final volume of 50  $\mu\text{L}$ . This mixture was incubated at 96  $^{\circ}\text{C}$  for 2 min, then at 1  $^{\circ}\text{C}$  lower temperature each minute for 14 min, 1  $^{\circ}\text{C}$  decrease every 30 s for 47 min, followed by a decrease of 2  $^{\circ}\text{C}$  every 30 s until 6  $^{\circ}\text{C}$  was reached. The labelled substrate was then stored at -20  $^{\circ}\text{C}$ .

## **2.5 Deamination Activity Assay**

The alkaline cleavage assay was used to measure and quantify the enzymatic deamination activity of AID. For this assay, 1.6  $\mu\text{g}$  of AID protein was incubated with 0.15-50 fmol of labelled TAC substrate in 100 mM phosphate activity buffer at pH 7.28 to a final reaction volume of 10-15  $\mu\text{L}$ , at 37 $^{\circ}\text{C}$  for 2 h. The cytidines converted to uracils were removed using 1 unit of Uracil-N-DNA glycosylase (UDG) by adding 0.2  $\mu\text{L}$  of UDG enzyme (*NEB*), 2  $\mu\text{L}$  UDG buffer, and 7.8  $\mu\text{L}$  autoclaved MilliQ water to the reaction, followed by incubation at 37  $^{\circ}\text{C}$  for 30 min. The resulting alkali-labile abasic site in the DNA phosphate backbone was cleaved by adding 2 M NaOH to a final concentration of 200 mM, followed by incubation at 96  $^{\circ}\text{C}$  for 5 min. The reaction mixture was incubated under the same conditions upon the addition of 10  $\mu\text{L}$  Formamide loading dye (0.25% Bromophenol Blue dye and 95% Formamide). The reactions were then loaded onto a 14 % denaturing acrylamide gel (14 % acrylamide:bisacrylamide 19:1, 25 % Formamide, 1X Tris Borate EDTA (TBE) and electrophoresed in 1X TBE running buffer for 3 h at 300 V and room temperature.

The resolved gel was then exposed to a blanked Phosphor Screen GP (Kodak Storage, *BioRad*) for 16-24 h and imaged using a Phosphorimager scanner (*Molecular Dynamics*).

## **2.6 Affinity-Binding Assay**

Substrate-AID binding affinity was measured using EMSA; 1.6 µg of AID was incubated with 0.15-50 fmol of labelled TAC substrate in EMSA binding buffer (2 µM MgCl, 1 mM DTT, 50 mM NaCl, and 50 mM Tris pH 7.5) for a final reaction volume of 10-15 µL at 25 °C for 1 h. Reactions were UV-crosslinked on ice using the optimum setting of a UV-crosslinker (*Stratagene*), rotated 180 degrees and repeated. EMSA loading dye (0.25 % Bromophenol Blue dye, 0.25 % Xylene Cyanol dye, 49.75 % Glycerol, and 49.75 % autoclaved MilliQ water) was added to each sample, before they were electrophoresed on an 8 % native gel (6 % Glycerol, 8 % acrylamide:bisacrylamide 19:1), in 0.5X TBE buffer for 3 h, at 300 V and 4 °C. Following electrophoresis, gels were dried onto chromatography paper (*Whatman*) using a slab dryer, at 80 °C for 2 h, then exposed to a blanked Phosphor Screen GP (*Kodak Storage, BioRad*) for 16-24 h and imaged using a Phosphorimager scanner (*Molecular Dynamics*).

## 2.7 Quantification of Assay Gel Images

Gel images were then quantified 3 times independently to ensure accuracy of data analysis using Quantity One 1-D Analysis Software (*BioRad*). Activity data was plotted as substrate concentration versus percentage product formed. The velocity ( $f\text{mole}/\text{min}/\mu\text{g}$ ) of the enzyme/substrate reaction was calculated using the following formula:

$$\frac{((\text{Fraction of substrate product}) \times [\text{substrate concentration (fM)}]) / (\text{time of incubation, in mins})}{\text{mass of AID (mg)}}$$

Velocity data was then graphed using GraphPad Prism software as substrate concentration versus velocity, with repeated experiments averaged with standard error of mean (SEM) calculated and shown on plots using the GraphPad Prism software. The Michaelis-Menten constant,  $K_m$ , was determined from the plotted activity data, using in-software non-linear regression, where  $K_m$  is the substrate concentration at half of the maximum reaction velocity, taken at saturation.

Binding data was plotted as free versus bound concentration of substrate. The dissociation constant was determined from the graphed data using the GraphPad Prism software's non-linear regression. The law of mass action formula can be rearranged to find  $K_d$ :

$$\text{Law of Mass Action: } [\text{Bound}] = \{[\text{Bound}_{\text{max}}] \times [\text{free}]\} / \{K_d + [\text{free}]\}$$

$$K_d = \{[\text{Bound}_{\text{max}}] \times [\text{free}]\} / [\text{Bound}] - [\text{free}]$$

The maximum concentration of bound substrate is taken at saturation and is assumed to signify total active AID in molar quantities.

## **2.8 Structural modeling of wt and mutant AID**

The structures of wt AID and mutant AID were modeled using the Swiss protein data bank (<http://swissmodel.expasy.org/repository/>). The resolved crystal structure of family member APOBEC-3G (Fig. 8) catalytic domain (PDB ID: 3EIU) was chosen as the best-fit template by the database and lowest energy state predictions were generated. Analysis of predicted models and generation of ribbon and surface charge models were performed using Pymol v1.30 (Fig. 4; Chen, Harjes, Gross, Fahmy, Lu, & Shindo, 2008; Holden, Prochnow, Chang, Bransteitter, Chelico, & Sen, 2008).

## Chapter 3 Results

### 3.1 AID Sequence and Mutagenesis

Although all residues selected for individual or compound mutagenesis line the suggested binding Groove 1 or 2 on the surface of the AID model, they represent different regions within the linear primary sequence of AID (Fig. 7). Residues within these grooves were mutated to other residues bearing an opposite or neutral charge, followed by comparison with wt AID for activity and ssDNA binding. To evaluate if combined mutations would have an additive effect on AID action, selected multiple mutants were designed and tested. Mutant AID proteins that were enzymatically active or bound substrate underwent further testing to determine the precise enzyme velocity kinetics,  $K_m$ , and ssDNA binding affinities,  $K_d$  (Table 3).

### 3.2 Overview of AID mutagenesis

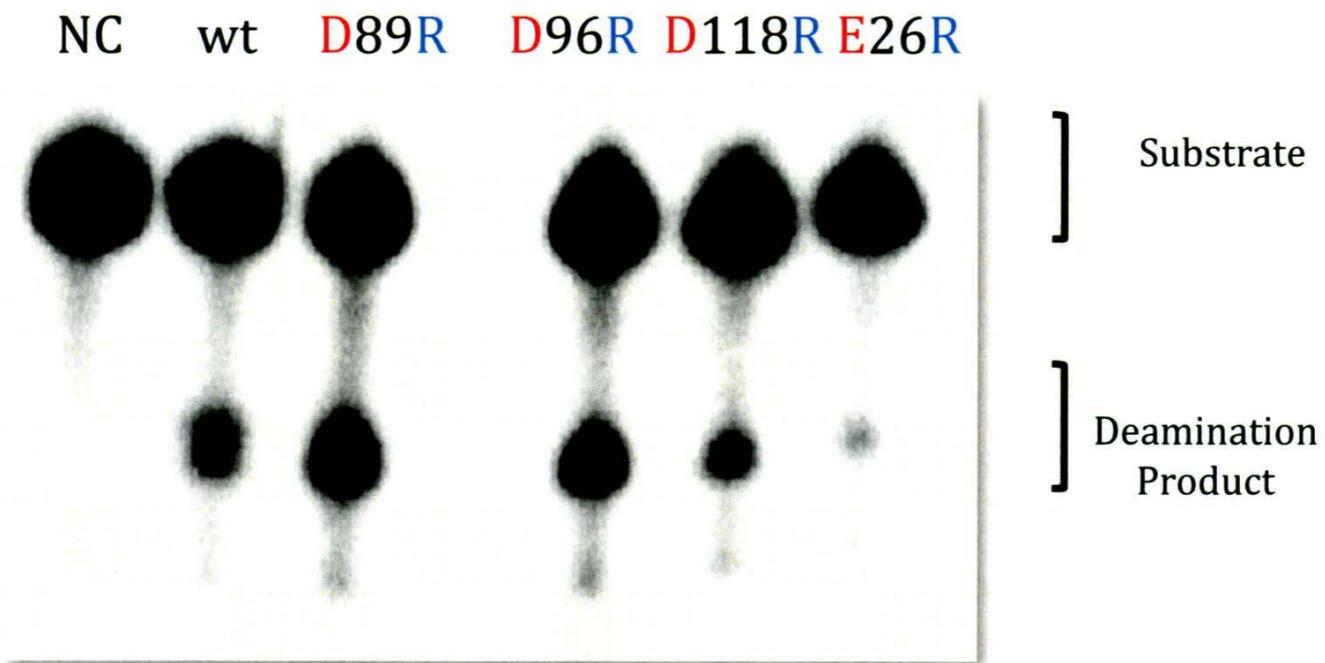
All together, 24 AID mutants were constructed: 18 single and 6 multiple mutants (Table 2). Of the single mutants, 6 were mutations of positive to negative, 6 positive to neutral, 4 were mutations of negative to positive and 2 were negative to neutral residues. Twelve of these mutations were in putative Groove 1 and 8 were in putative Groove 2. Of note, the R24D and R24A mutations are located in both Grooves and are included in the number of mutants for each groove.

Each mutant AID enzyme was tested for activity compared to wt AID using the alkaline cleavage assay. The imaged gels show resolved substrate where deaminated substrate indicates enzymatic activity (Fig. 10). For mutant enzymes that showed activity, the alkaline cleavage experiments were repeated with varying levels of substrate and a standardized amount of enzyme to determine kinetic activity, or velocity, of each mutant (Fig. 11).

In addition to activity assays, each mutant AID enzyme was also tested for binding ability compared to wt AID with the EMSA assay. The imaged gels show resolved bound and unbound substrate (Fig. 12). All mutant enzymes demonstrated substrate binding ability and were subsequently tested repeatedly with the EMSA assays using varying levels of substrate and a standardized amount of enzyme to determine kinetic binding, or binding affinity, of each mutant (Fig. 13).

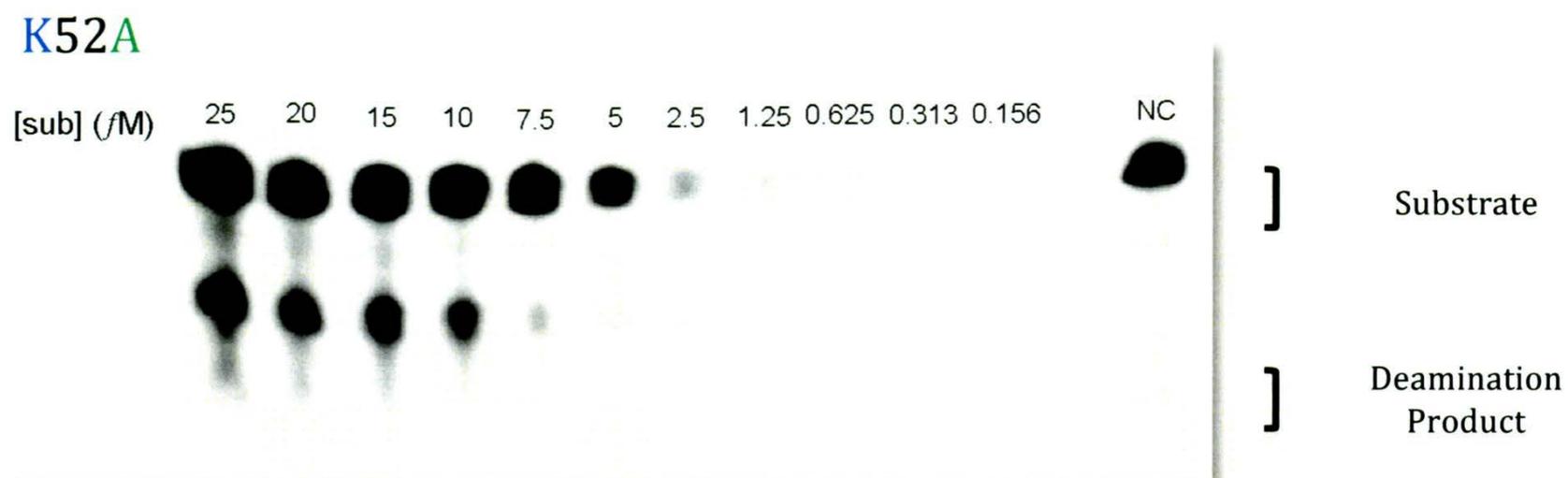
<b>Table 2.</b>	<b>Mutation</b>	
<b>Amino Acid</b>	<b>Opposite Charge</b>	<b>Neutral Charge</b>
Groove 1		
<b>K22</b>	<b>K22D</b>	<b>K22A</b>
<b>R24</b>	<b>R24D</b>	<b>R24A</b>
<b>R25</b>	<b>R25D</b>	<b>R25A</b>
<b>R63</b>	<b>R63D</b>	<b>R63A</b>
<b>D89</b>	<b>D89R</b>	<b>D89A</b>
<b>D96</b>	<b>D96R</b>	<b>D96A*</b>
<b>D118</b>	<b>D118R</b>	<b>D118A*</b>
Groove 2		
<b>R24</b>	<b>R24D</b>	<b>R24A</b>
<b>E26</b>	<b>E26R</b>	<b>E26A</b>
<b>R50</b>	<b>R50D</b>	<b>R50A</b>
<b>K52</b>	<b>K52D</b>	<b>K52A</b>
Multiple Mutants		
	<b>K22D/ K52D</b>	
	<b>K22D/ R63D</b>	
	<b>R24D/R63D</b>	
	<b>R25D/R63D</b>	
	<b>K22D/ K52D/ R63D</b>	
	<b>D89R / D96R</b>	
Positive - Negative - Neutral *Not Done		

**Table 2. Residues mutated according to mutation type and groove location.**



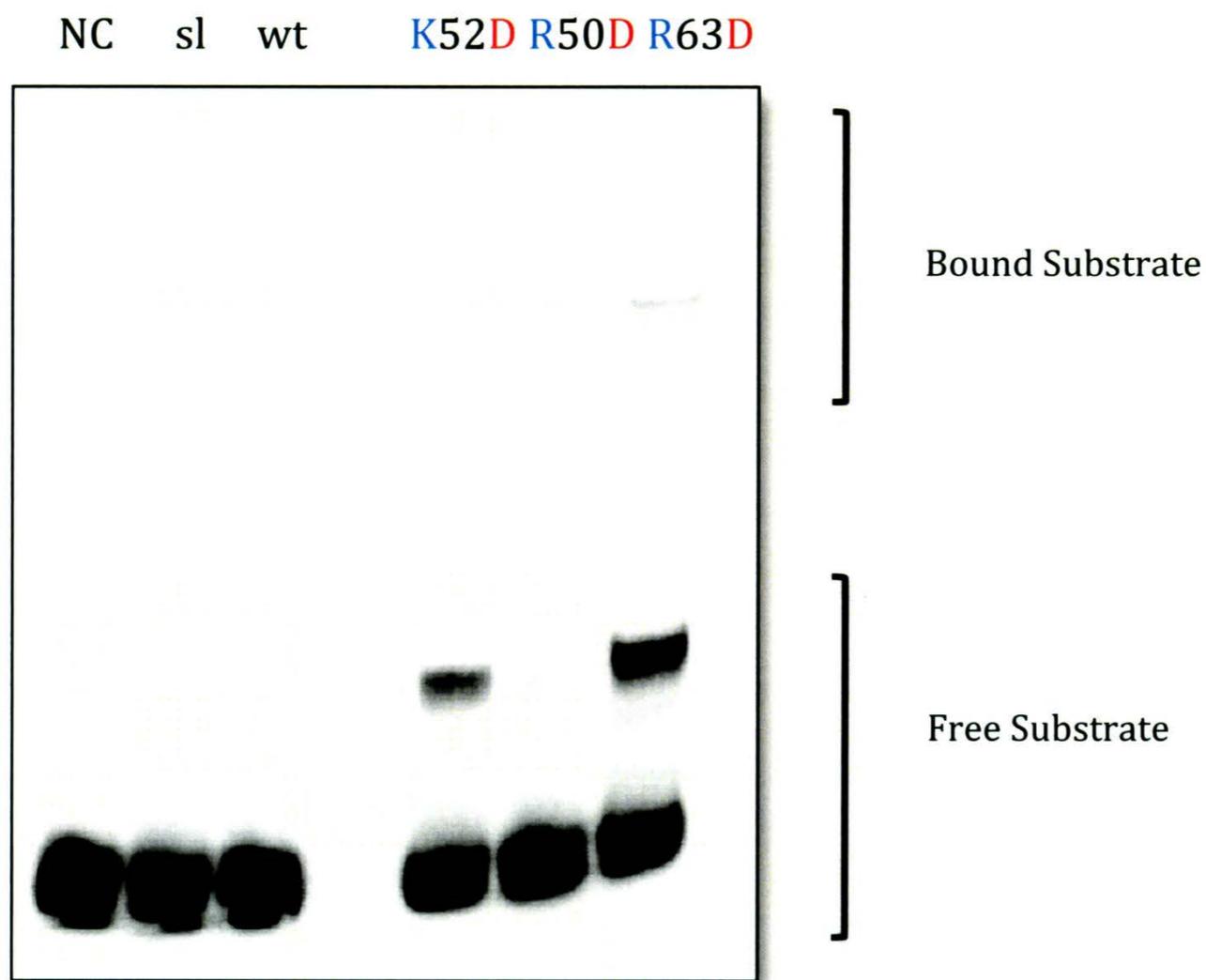
**Figure 10. Representative Gel of Alkaline Cleavage assay.**

AID enzyme was incubated with radioactively labelled substrate at 37°C followed by subsequent 37°C incubations with UDG, NaOH and with denaturing dye. The DNA substrate and deamination product were electrophoretically resolved on denaturing gels and imaged using a phosphoimager. NC = negative control of substrate without AID enzyme, wt = wild type AID as a positive control. Each mutation label indicates the AID enzyme with the corresponding mutation.



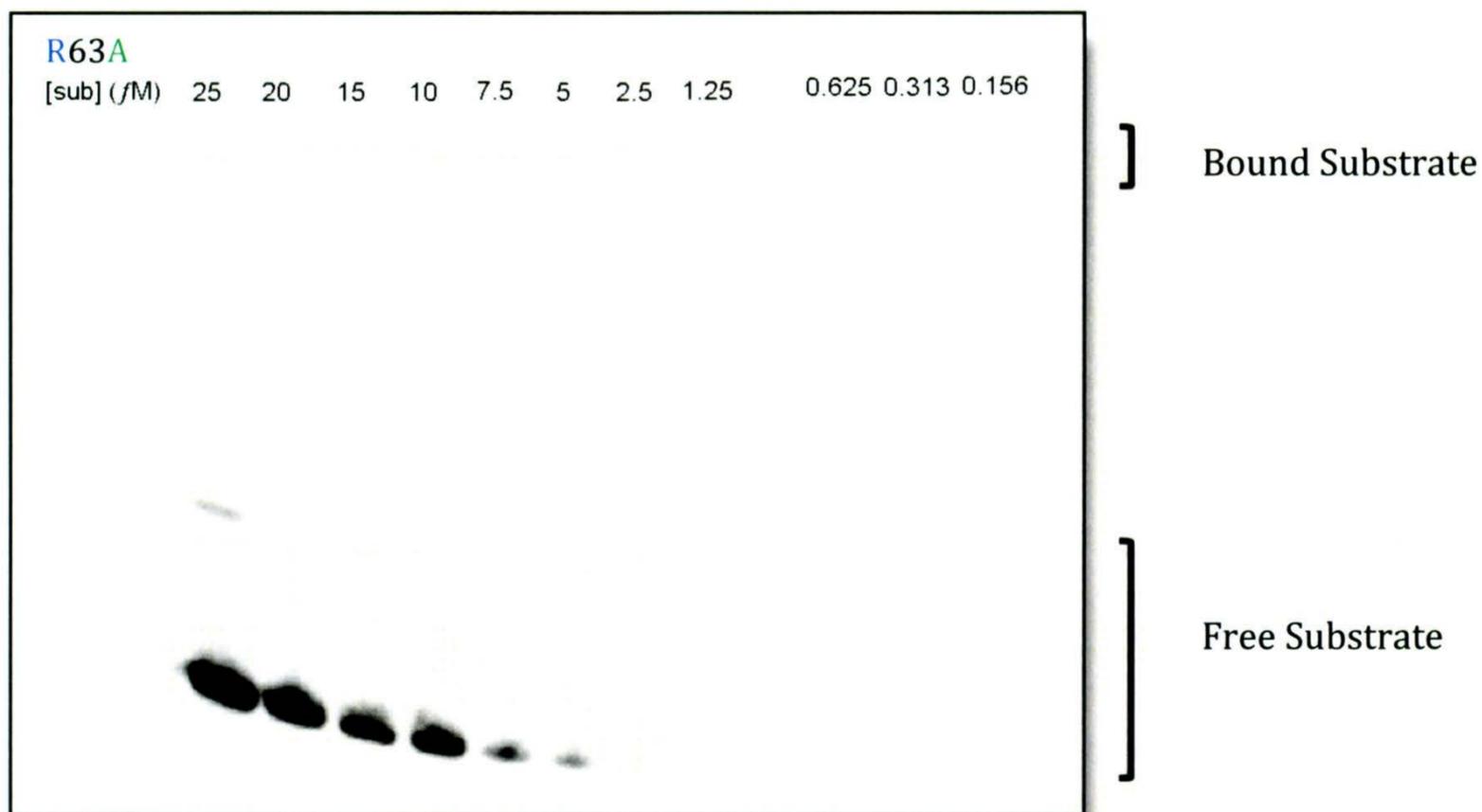
**Figure 11. Representative gel image of Alkaline Cleavage Kinetic assay for K52A AID mutant.**

Varying amounts (0.156 – 25 fM) of purified AID enzyme were incubated with radioactively labelled substrate at 37 °C followed by subsequent 37°C incubations with UDG, NaOH and with denaturing dye. The DNA substrate and deamination products were electrophoretically resolved on denaturing gels and imaged using a phosphoimager. NC = negative control of substrate without AID enzyme, wt = wild type AID as a positive control.



**Figure 12. Representative gel image of EMSA Assay for three AID mutants (K52D, R50D and R63D).**

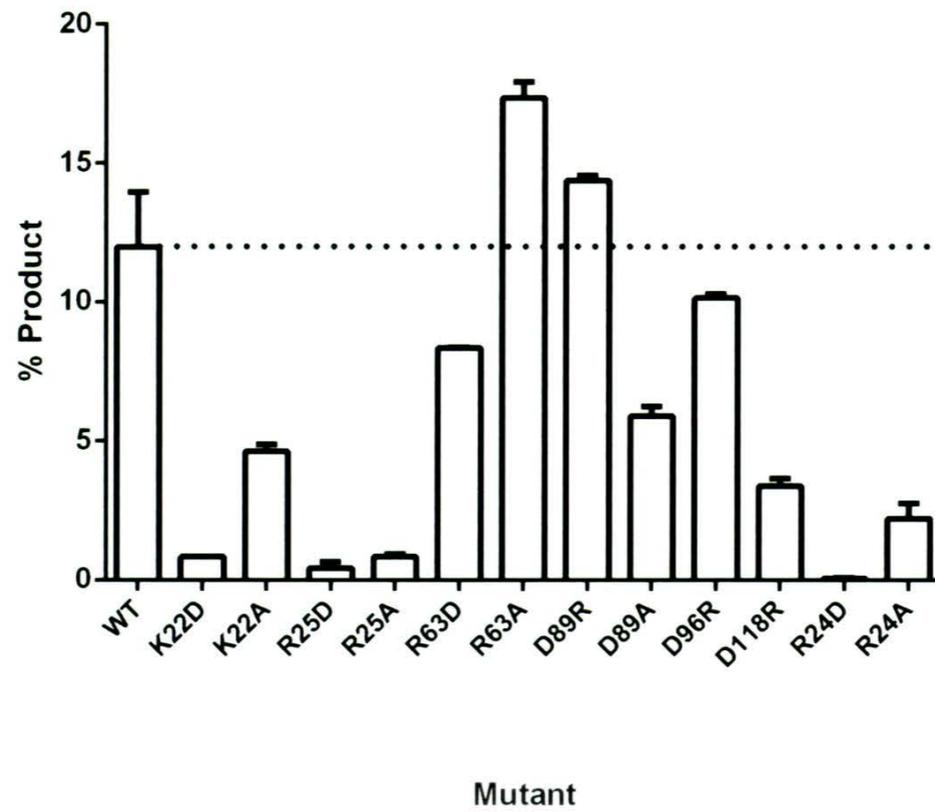
AID enzymes were incubated with radioactively labelled substrate at 37°C then UV-crosslinked to preserve the AID:Substrate complex. Bound and unbound substrates were electrophoretically resolved on native gels and imaged using a phosphoimager. NC = negative control of substrate with no AID enzyme, sl = short linker wtAID and wtAID are positive controls. Each mutation label indicates the AID enzyme with the corresponding mutation.



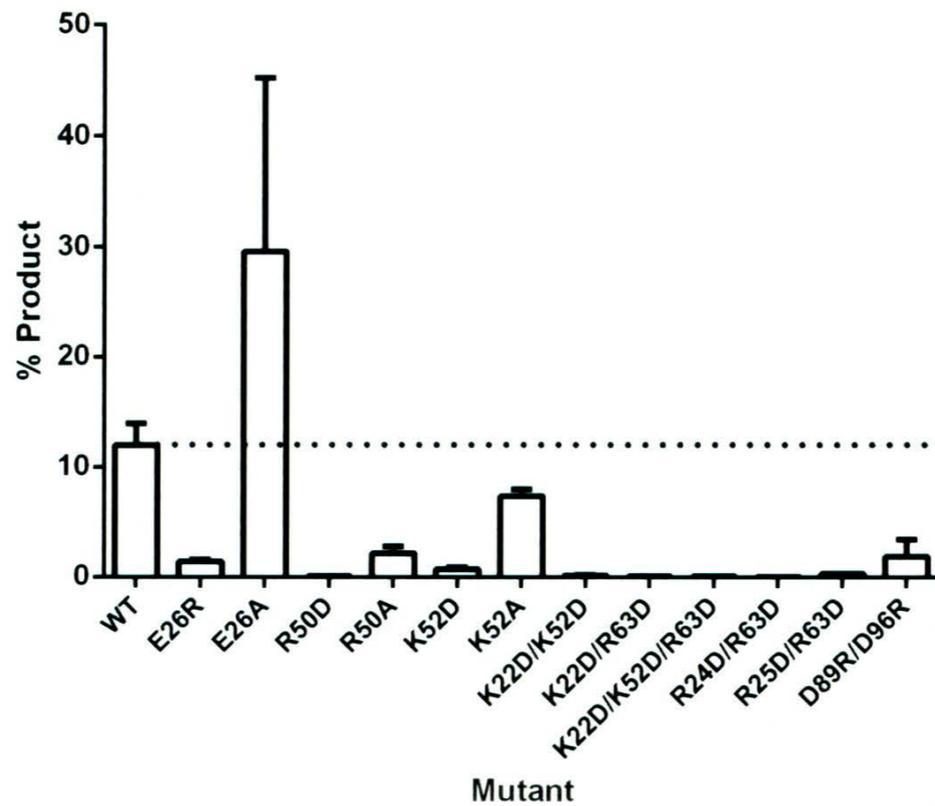
**Figure 13. Representative gel image of kinetic EMSA Assay for R63A.**

Varying amounts (0.156 – 25 fM) of purified AID enzyme were incubated with radioactively labelled substrate at 37°C then UV-crosslinked to preserve the AID:Substrate complex. Bound and unbound substrates were electrophoretically resolved on native gels and imaged using a phosphoimager. The DNA substrate and deamination product were electrophoretically resolved on denaturing gels and imaged using a phosphoimager.

a. Groove 1: Mutants compared to wild-type using percentage of product formed



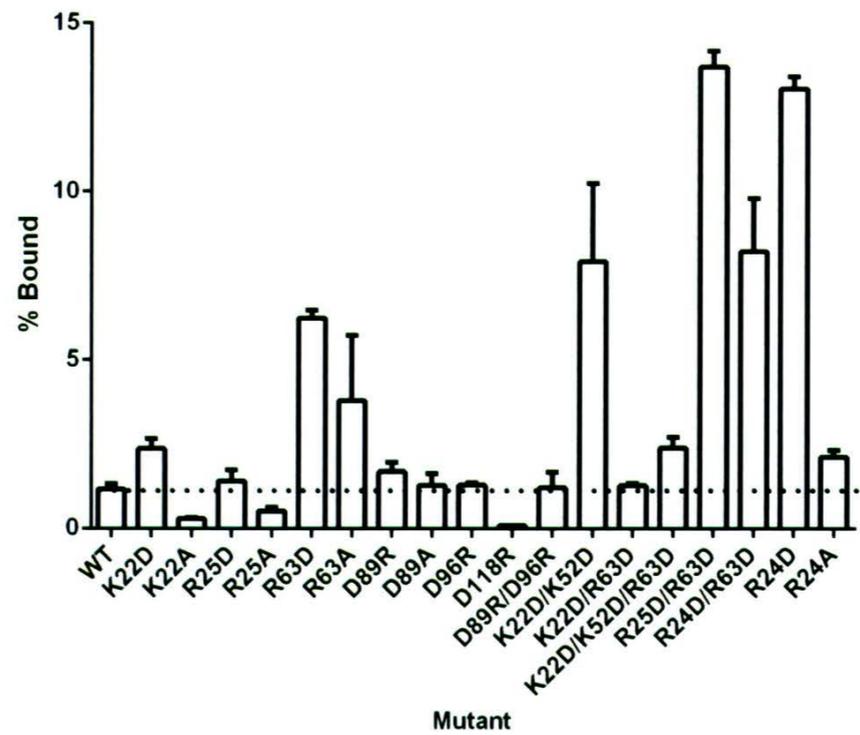
b. Groove 2: Mutants compared to wild-type using percentage product formed



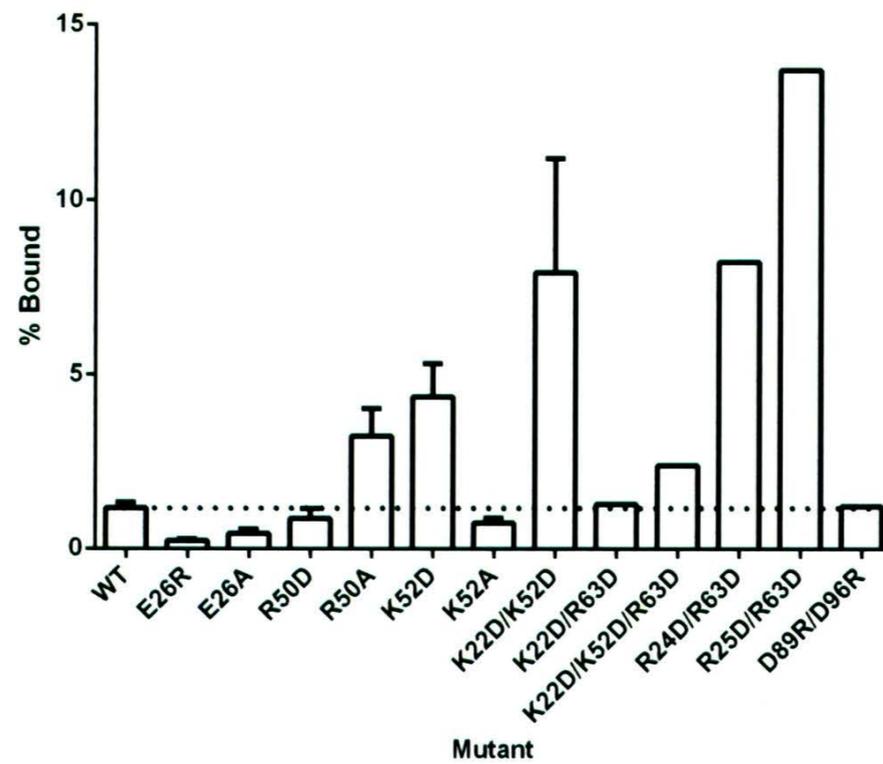
**Figure 14. Alkaline Cleavage assay bar graphs for mutants in (a.) Groove 1 and (b.) Groove 2**

Graph represents average percent product of repeated Alkaline Cleavage experiments for each indicated mutant AID. SEM is indicated by error bars.

a. Groove 1: Mutants compared to wild-type using percentage of substrate bound



b. Groove 2: Mutants compared to wild-type using percentage substrate bound



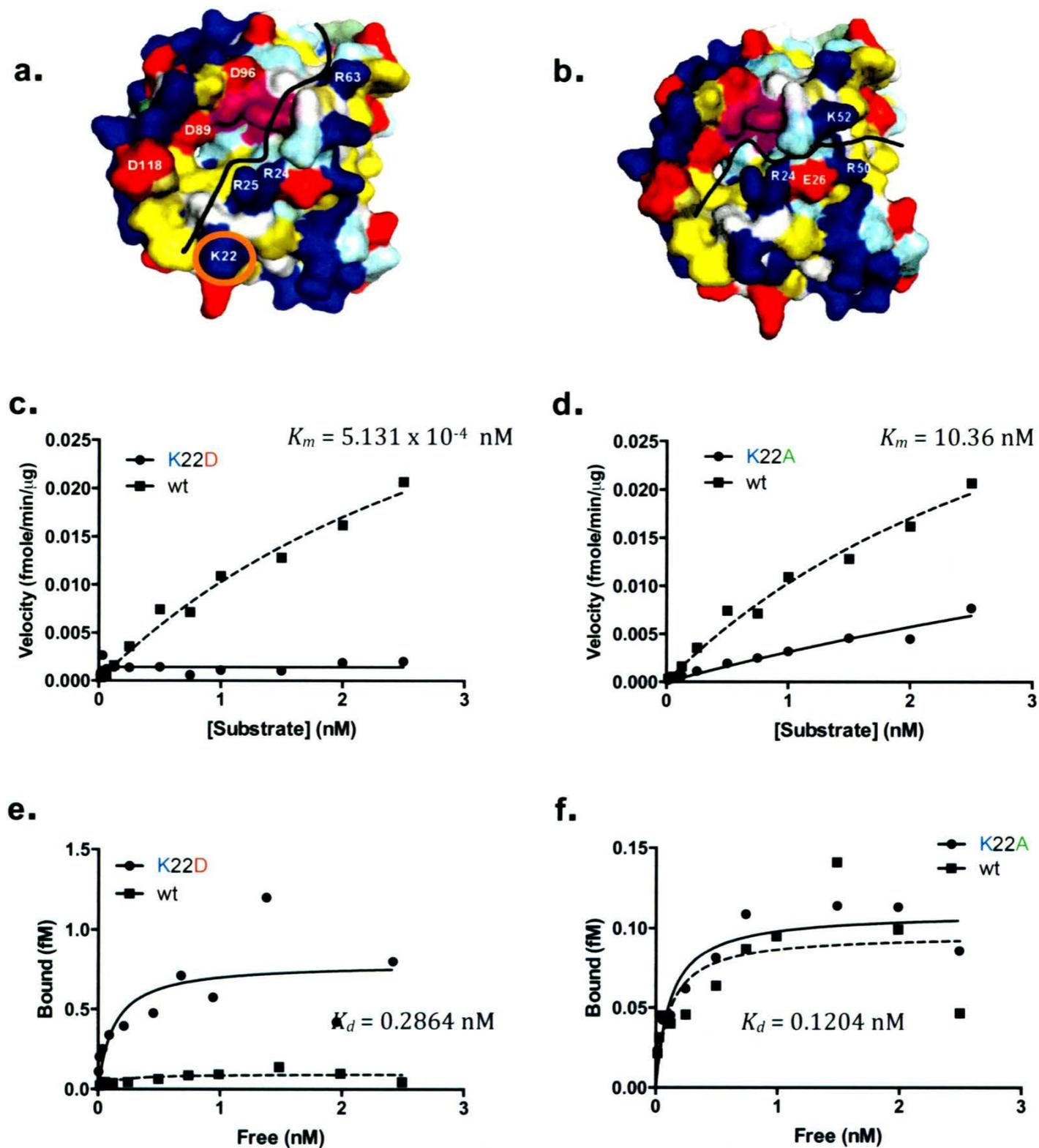
**Figure 15. EMSA Binding Bar Graph**

Graph represents average percent bound substrate of repeated EMSA binding experiments for each indicated mutant AID. SEM indicated by error bars.

### 3.2.1 Groove 1: K22 results

K22 is found within putative Groove 1 and is part of the NLS (Fig. 7, Fig. 16a). A substantial decrease in AID enzymatic activity occurred with mutations at residue K22; overall activity was reduced to 5 % for K22D, when compared to wt enzyme that generated 12 % deamination product under defined assay conditions (Fig. 14a). The K22A mutation was a less drastic change in residue charge but still led to a reduced AID activity, with approximately 3.5 % deamination product formed under similar assay conditions (Fig. 14a). The K22D mutation caused a decrease in activity according to kinetic curves, coincident with a decrease in  $K_m$  to  $5.313 \times 10^{-4}$  nM from 3.824 nM for wt AID (Fig. 16, Table 3). K22A demonstrated a decrease in enzyme efficiency, with a  $K_m$  value of 10.36 nM (Fig. 16, Table 3).

Changes in mutant AID activity may be influenced by the duration of ssDNA substrate binding. An increase in the percentage of bound ssDNA was evident for K22D over wt AID, with  $\approx 3$  % bound product compared to  $\approx 1.5$  %, respectively (Fig. 15a). K22A bound-complex was much less than wt, at  $\approx 0.5$  % (Fig. 15a). In terms of dissociation constants,  $K_d$  for K22D increased to 0.2864 nM, relative to 0.1137 nM for wt AID, indicating an overall decrease in binding affinity (Fig. 16e, Table 3). K22A exhibited very similar ssDNA substrate binding affinity to wt AID, with a similar  $K_d$  value of 0.1204 nM (Fig. 16f, Table 3).



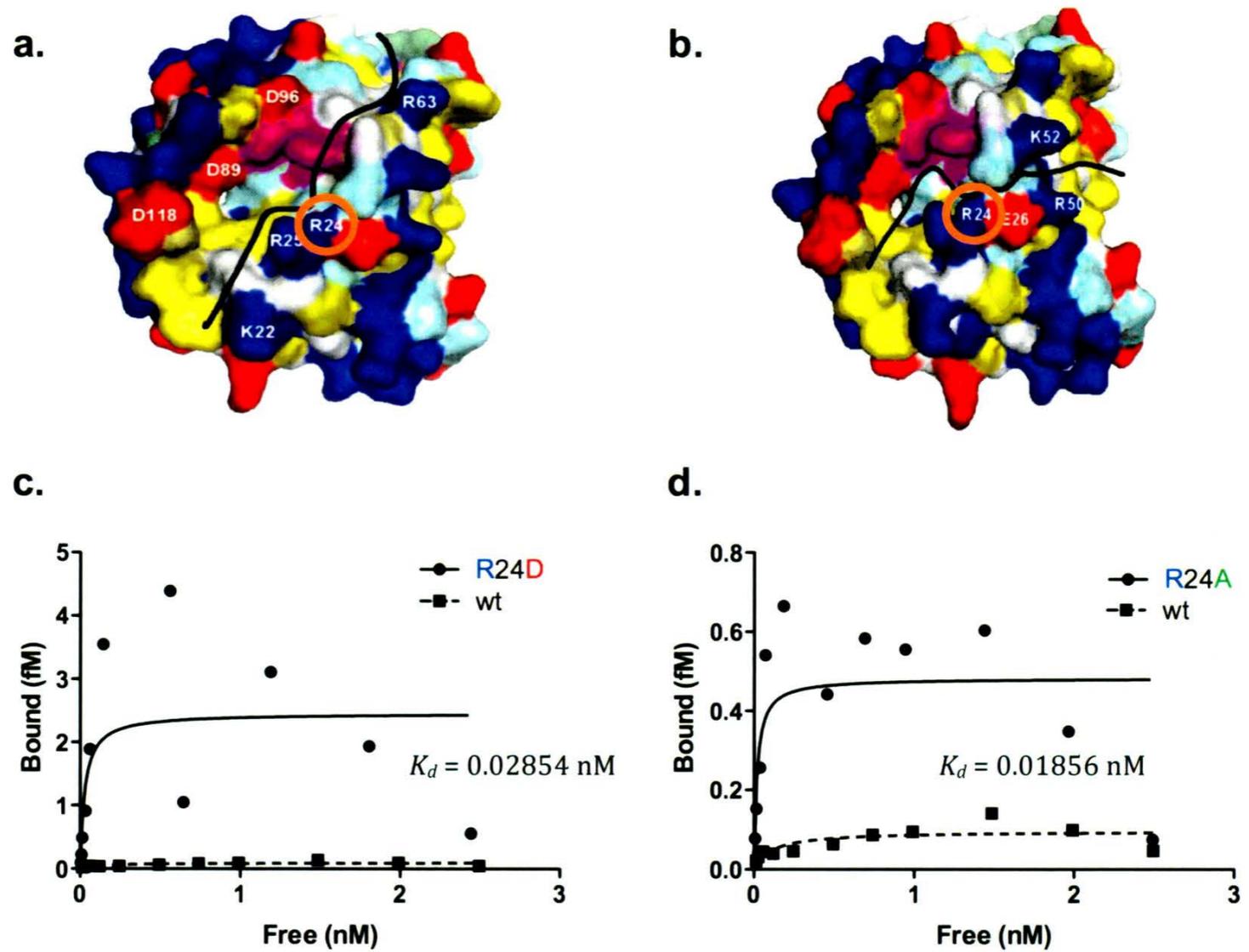
### Figure 16. Groove 1: K22 mutations and assays

Purified AID was assayed for enzymatic reaction velocity and for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b.) of the enzyme. Velocity (c. and d.) and binding affinity (e. and f.) of K22D and K22A, respectively, relative to wt. Data represents three independent assays.  $K_m$  and  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.2.2 Groove 1 and 2: R24 results

The R24 residue is located in an overlap between Groove 1 and 2 and thus must be considered as evidence for both Grooves as a potential docking site of AID substrate (Fig. 17a, & b). R24 also lies within the NLS domain of AID (Fig. 7). The mutations of R24 were nearly enzymatically dead compared to wt enzyme activity, the only mutation of Groove 2 to decrease activity. The activity of R24A and R24D produced less than 5 % deamination product versus 12% of product formed for wt (Fig. 14a). Due to such a great loss in activity, no kinetic velocity experiments were performed on R24 mutant enzymes.

The decrease in activity may be due to an increased in substrate binding time. R24A yielded an increase of bound ssDNA product with ~3 % (Fig. 15a) compared to 1.5 % that of wt AID, an approximate 2-fold increase. R24D exhibited one of the most significant increases in binding, with 14 % product versus ~1.5 % of wt, an approximate 9-fold increase. Binding kinetics for these R24A and R24D resulted in lower dissociation constants of 0.01856 nM and 0.02854 nM respectively, indicating increased binding affinity over wt with a  $K_d$  value of 0.1204 nM (Fig. 17c & d, Table 3).



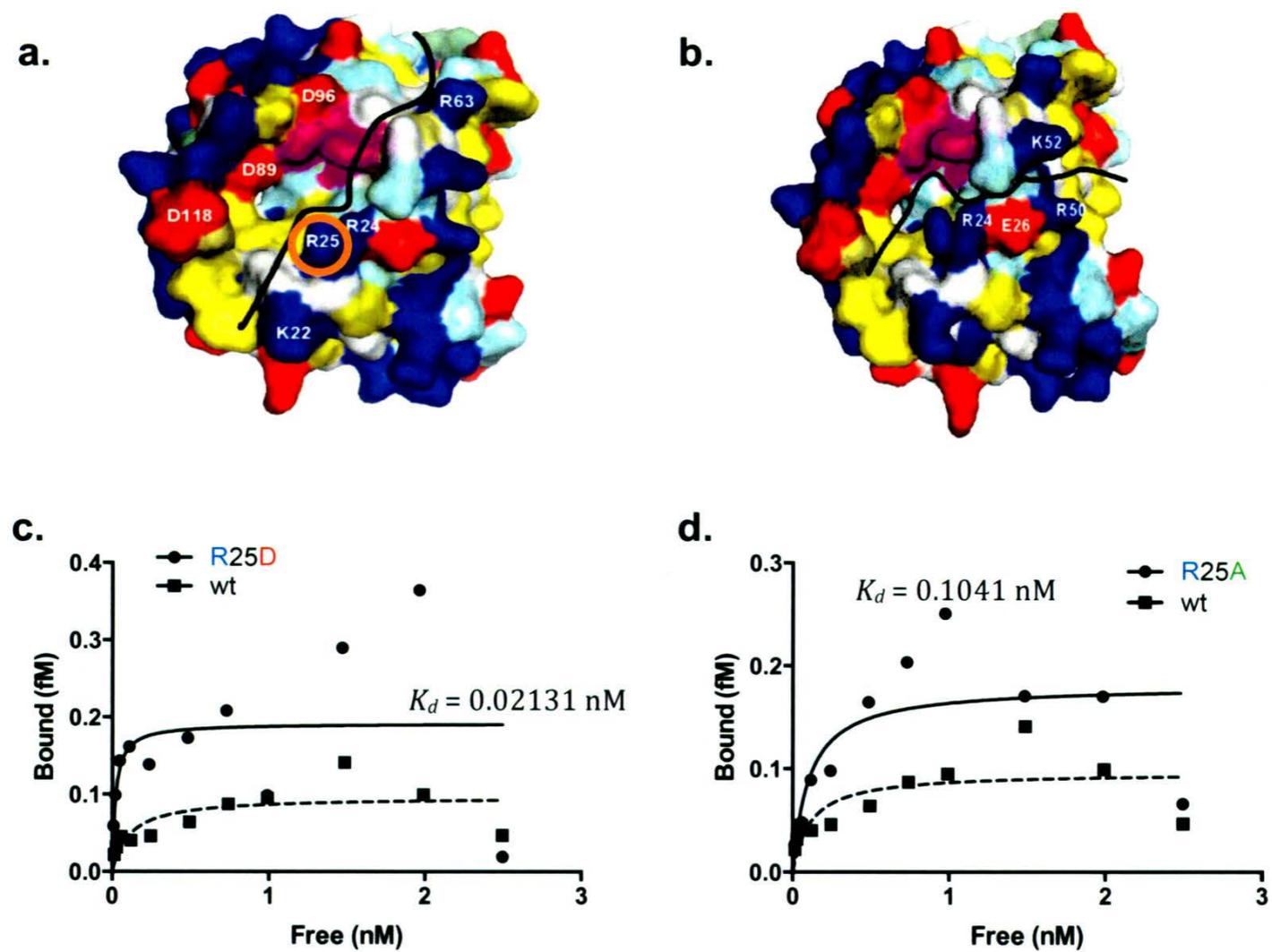
**Figure 17. Groove 1 & 2: R24 mutations and assays**

Purified AID was assayed for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b) of the enzyme. Binding affinity of R24D and R24A (relative to wt). Binding affinity (c. and d.) of R24D and R24A, respectively, relative to wt. Data represents three independent assays.  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.2.3 Groove 1 and 2: R25 Results

R25 is found in both Groove 1 and Groove 2 and must be considered as evidence for both grooves (Fig. 18a). A significant decrease in enzyme activity was found for both R25D and R25A mutants. Each mutant enzyme formed approximately 1.5 % one tenth of the product which wt formed (Fig. 14a). Since little activity was shown, velocity experiments were not performed for either R25 mutant.

Compared to the binding results of wt AID, R25D exhibited similar binding, with ~2 % of bound substrate formed. R25A showed a decrease in bound product, yielding approximately half of bound substrate wt AID binds (Fig. 15a). Binding affinity constant  $K_d$  of 0.02131 nM (Fig. 18, Table 3) for R25D demonstrated a 5-fold increase in affinity for substrate in comparison to the wt values of 0.1137 nM. R25A had a slight increase in binding affinity with  $K_d$  of 0.1041 nM (Fig. 18c & d, Table 3).



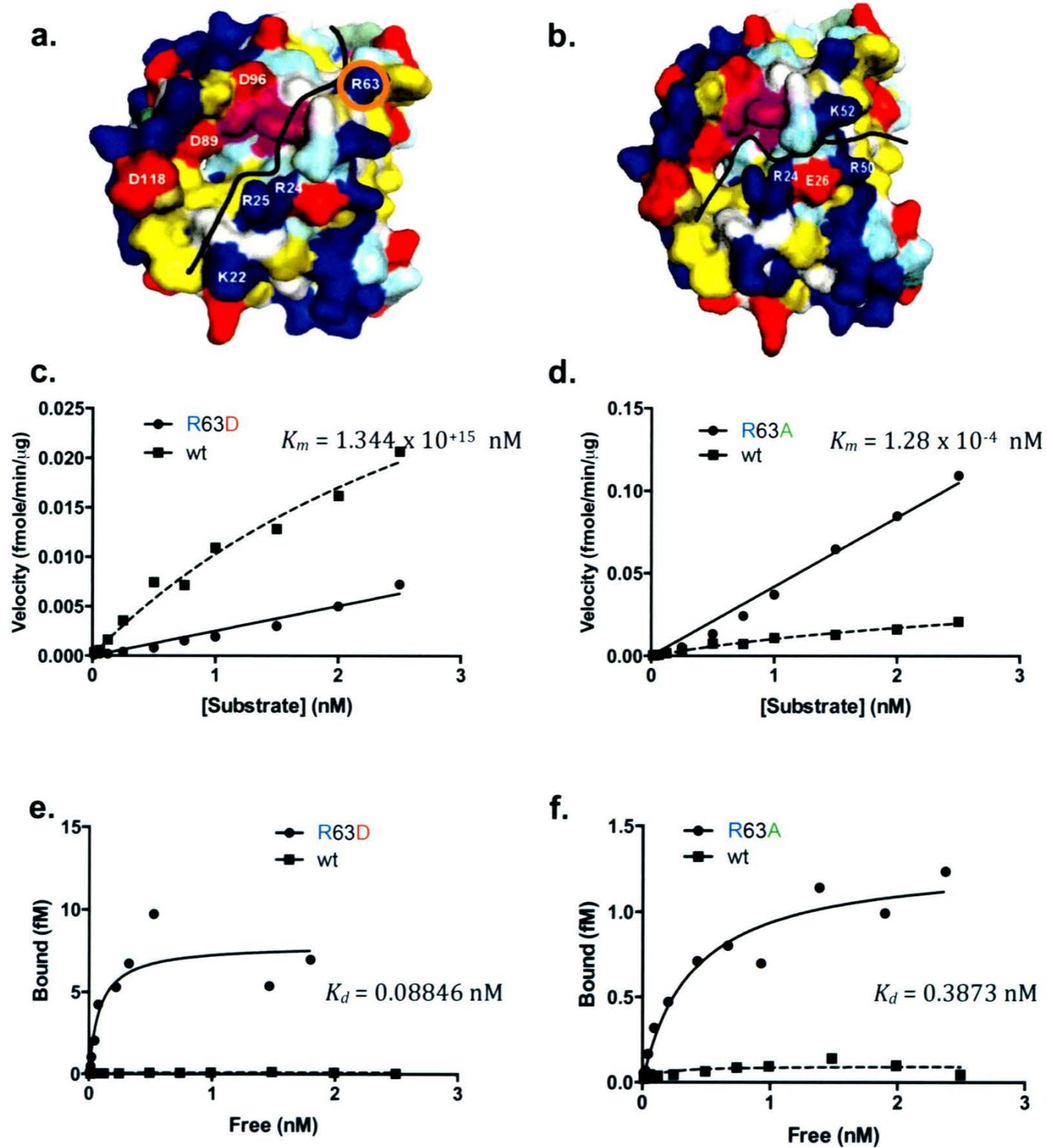
**Figure 18. Groove 1: R25 mutations and assays**

Purified AID was assayed for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b.) of the enzyme. Binding affinity of R25D (c.) and R25A (d.), relative to wt. Data represents three independent assays.  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.2.4 Groove 1: R63 results

R63 is found within Groove 1 of the AID surface and within the cytidine deaminase domain of the sequence (Fig. 19 a). The R63D mutation resulted in approximately one-third reduction of activity, forming 8 % product compared to 12 % for wt AID (Fig. 14a). However, R63A showed an increase in product formed with 17 %, an approximate 1.5-fold increase (Fig. 14a). The enzyme velocity experiments were contrary to the activity assay results. Both R63D and R63A showed  $K_m$  values exponentially higher than that of wt,  $1.344 \times 10^{+15}$ ,  $1.28e \times 10^{+14}$  and 3.824 nM, respectively (Fig. 19c & d, Table 3). This suggests a highly significant decrease in reaction velocity for R63D and R63A.

The binding assay for R63D displayed a 4-fold increase of bound substrate from approximately 1.5 % bound for wt to 6 % for R63D (Fig. 15a). Likewise, the binding affinity calculated for R63D from kinetic experiments,  $K_d$  0.08846 nM was one order of magnitude smaller than that of wt, 0.1137 nM, indicating a greater binding affinity (Fig. 19e, Table 3). R63A displayed a 2-fold increase with 4 % bound substrate over 1.5 % bound by wt (Fig. 15a). Contrary to this result, kinetic experiments elucidated a  $K_d$  value of 0.3873 nM an approximate 3-fold increase of that of wt indicating an increase in binding affinity for R63A (Fig. 19f, Table 3).



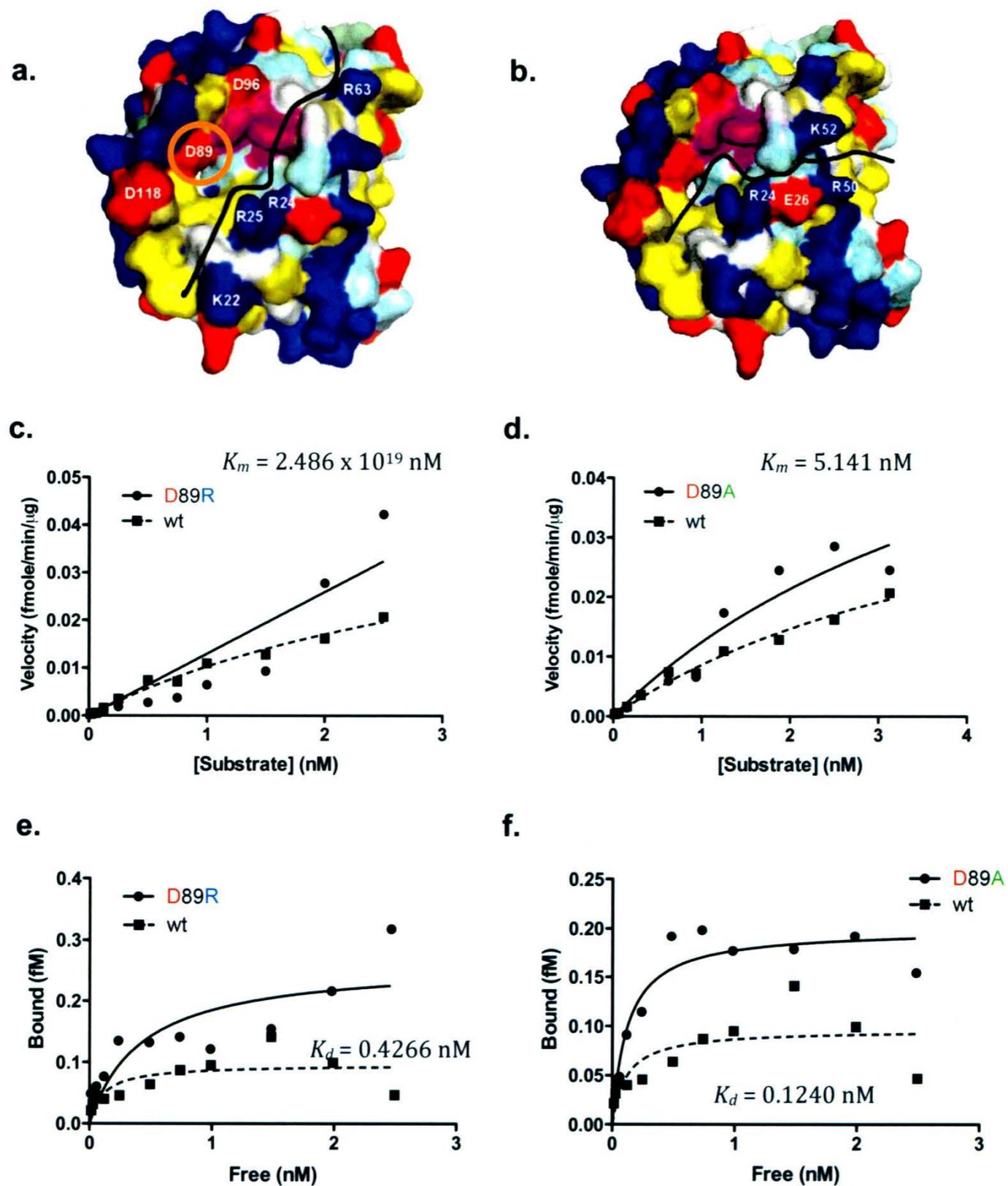
**Figure 19. Groove 1: R63 mutations and assays**

Purified AID was assayed for enzymatic reaction velocity and for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b) of the enzyme. Velocity and binding affinity of R63D (c. and e., respectively) and R63A (d. and f., respectively), relative to wt. Data represents three independent assays.  $K_m$  and  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.2.5 Groove 1: D89 results

D89R showed a slight increase in enzyme activity with 14 % product formed versus 12 % produced by wt (Fig. 14a). The velocity for D89R increased to 0.4266 over 3.824 nM of wt, which demonstrated a reduced enzyme velocity (Fig. 20c, Table 3). D89A, a neutral charge mutation, halved the activity of wt AID, producing only 6 % of product (Fig. 14a). The calculated velocity for D89A opposes this reduction of activity with a decreased  $K_m$  value of 5.141 compared to 3.824 nM of wt (Fig. 20d, Table 3). This suggests an increase in enzyme velocity.

The binding assays show little change in bound substrate for D89 mutations versus the substrate bound by wt AID. D89R produced 2 % of bound substrate and D89A, 1.5 % a slight increase and similar values respectively to wt bound substrate of 1.5 % (Fig. 15a). D89A also maintains a similar binding affinity of 0.1240 nM to wt, 0.1137 nM with an approximate one-tenth increase, which indicates a reduction in binding activity (Fig. 20f, Table 3). On the contrary, the  $K_d$  of D89R is 0.4266 nM, an approximate 4-fold increase over the wt value, which is a significant decrease in binding affinity (Fig. 20e, Table 3).



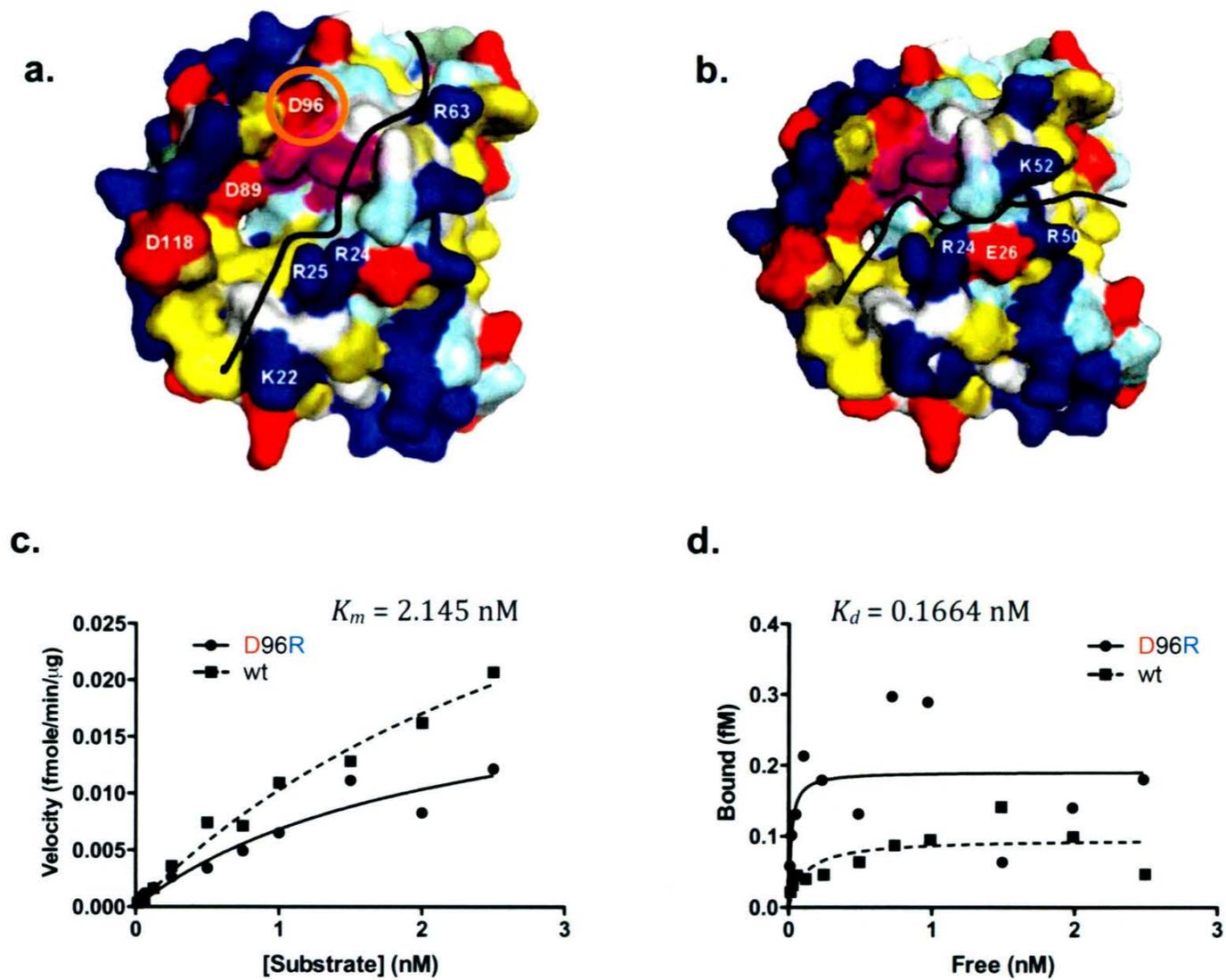
### Figure 20. Groove 1: D89 mutations and assays

Purified AID was assayed for enzymatic reaction velocity and for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b) of the enzyme. Velocity (c. and d.) and binding affinity (e. and f.) of D89R and D89A, relative to wt. Data represents three independent assays.  $K_m$  and  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.2.6 Groove 1: D96 results

The enzymatic activity of D96R is comparable to that of wt, which resulted in 11 % of product and 12 % of product, respectively (Fig. 14a). Despite the similarity in activity, the velocity determined for D96R, 2.145 nM, is just over half of that of wt at 3.824 nM, suggesting an increase in enzyme efficiency (Fig. 21c, Table 3).

Binding assays for D96R and wt AID both resulted in 1.5 % of bound substrate indicating no significant change in binding (Fig. 15a). The  $K_d$  of D96R was calculated as 0.1664 nM, approximately one tenth of wt  $K_d$  0.1137 nM, which suggests a significant increase in binding affinity (Fig. 21d, Table 3).



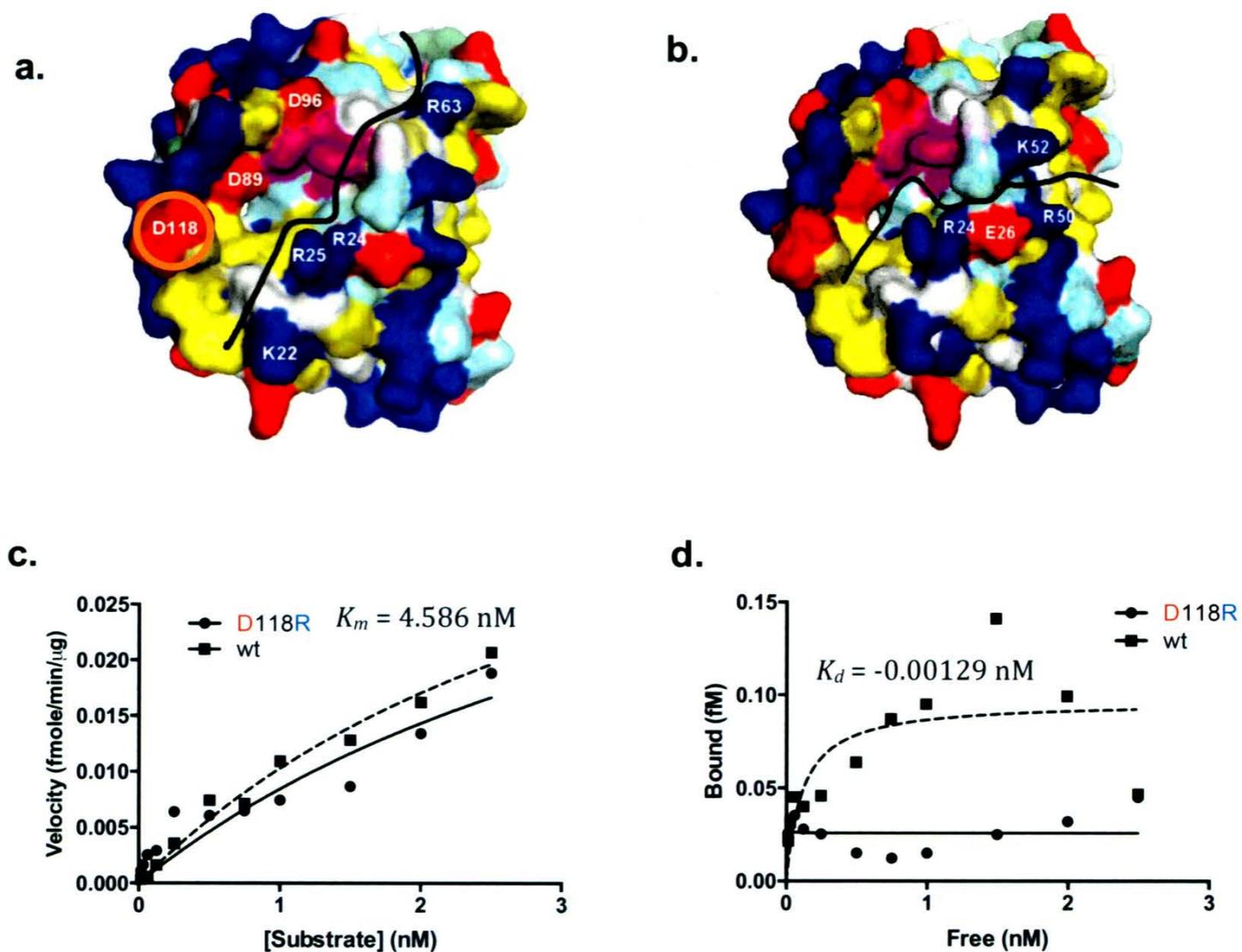
**Figure 21. Groove 1: D96 mutations and assays**

Purified AID was assayed for enzymatic reaction velocity and for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b) of the enzyme. Velocity (c.) and binding affinity (d.) of D96R, relative to wt. Data represents three independent assays.  $K_m$  and  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.2.7 Groove 1: D118 results

The enzymatic activity of D118R showed a significant decrease to 4 % product approximately one third of that of wt at 12 % product (Fig. 14a). Coincident reduction was seen in enzyme velocity with an increased  $K_m$  of 4.586 nM over 3.824 nM of wt (Fig. 22c, Table 3).

Binding of D118R was approximately 0 %, a significant decrease from wt bound product of 1.5 % (Fig. 15a). The calculated  $K_d$  for D118R was -0.001209 nM, 2 orders of magnitude smaller than wt, which indicates a stronger binding affinity for substrate (Fig. 22d, Table 3).



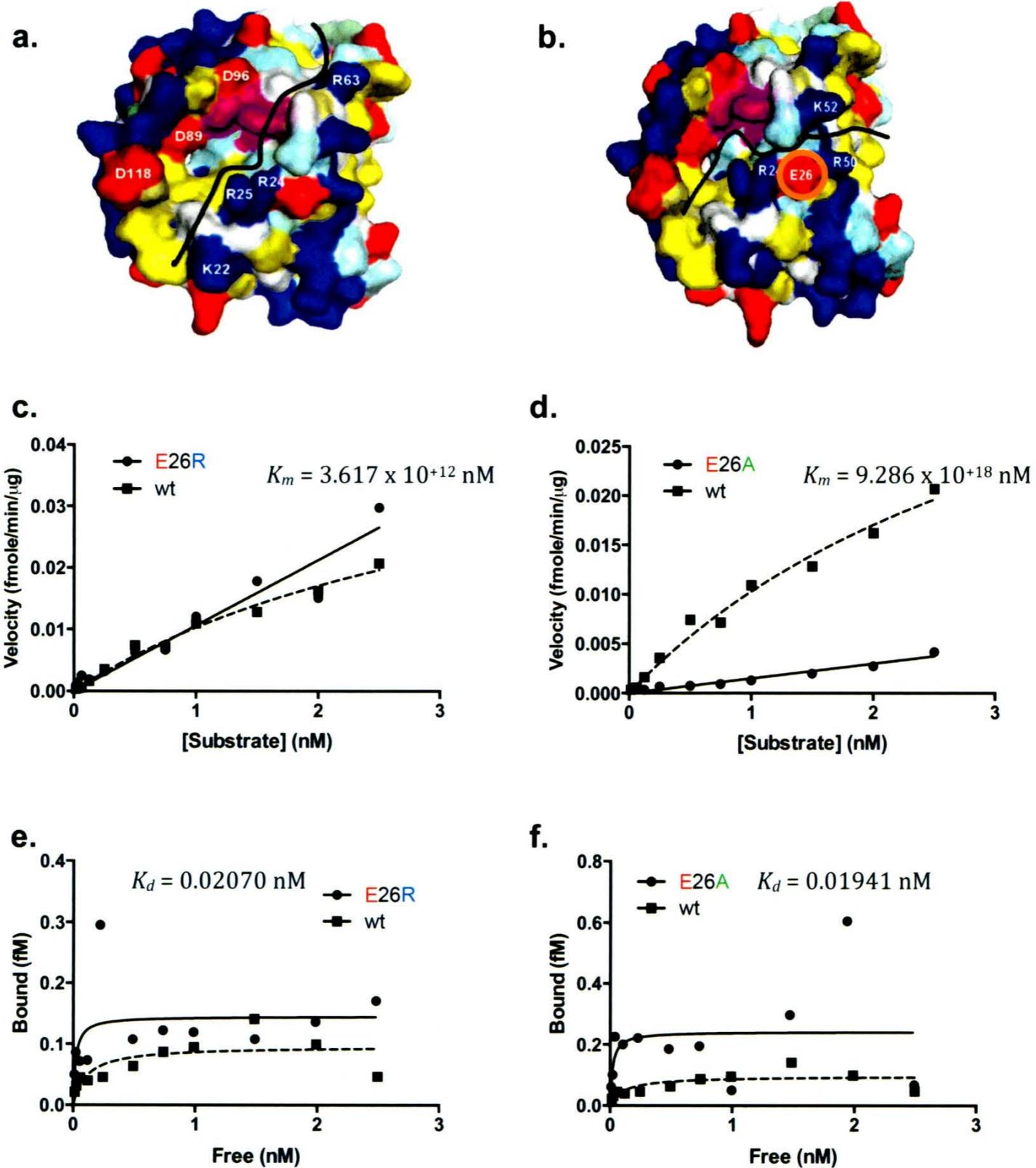
**Figure 22. Groove 1: D118 mutations and assays**

Purified AID was assayed for enzymatic reaction velocity and for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b) of the enzyme. Velocity (c.) and binding affinity (d.) of D118R, relative to wt. Data represents three independent assays.  $K_m$  and  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.3.1 Groove 2: E26 results

The two mutations of residue E26 had opposing effects on enzyme activity. E26R produced approximately one sixth of the product relative to wt AID, a comparison of 2 % versus 12 %, respectively (Fig. 14b). The  $K_m$  for E26R is in agreement with this result with a calculated value of  $3.617 \times 10^{+12}$  nM, a theoretical saturation level, indicating very low catalytic rate compared to the wt value of 3.824 nM (Fig. 23c, Table 3). Conversely, E26A produced almost 2.5 times the amount of product of wt with a result of 30 %. Kinetic experiments yielded a  $K_m$  value of  $9.286 \times 10^{+18}$  nM, which like E26R represents a dramatic decrease in catalytic activity (Fig. 23d, Table 3).

Both E26R and E26A bound to less substrate with approximately 0.25 % and 0.5 %, respectively versus 1.5 % substrate bound by wt (Fig. 15b). Contrary to these binding results, the  $K_d$  values for E26R and E26A are 0.02070 nM and 0.01941 nM, similar values but also an order of magnitude smaller than that of wt at 0.1137 nM (Fig. 23e & f, Table 3). These values indicate an increase in binding affinity of these mutant enzymes for ssDNA.



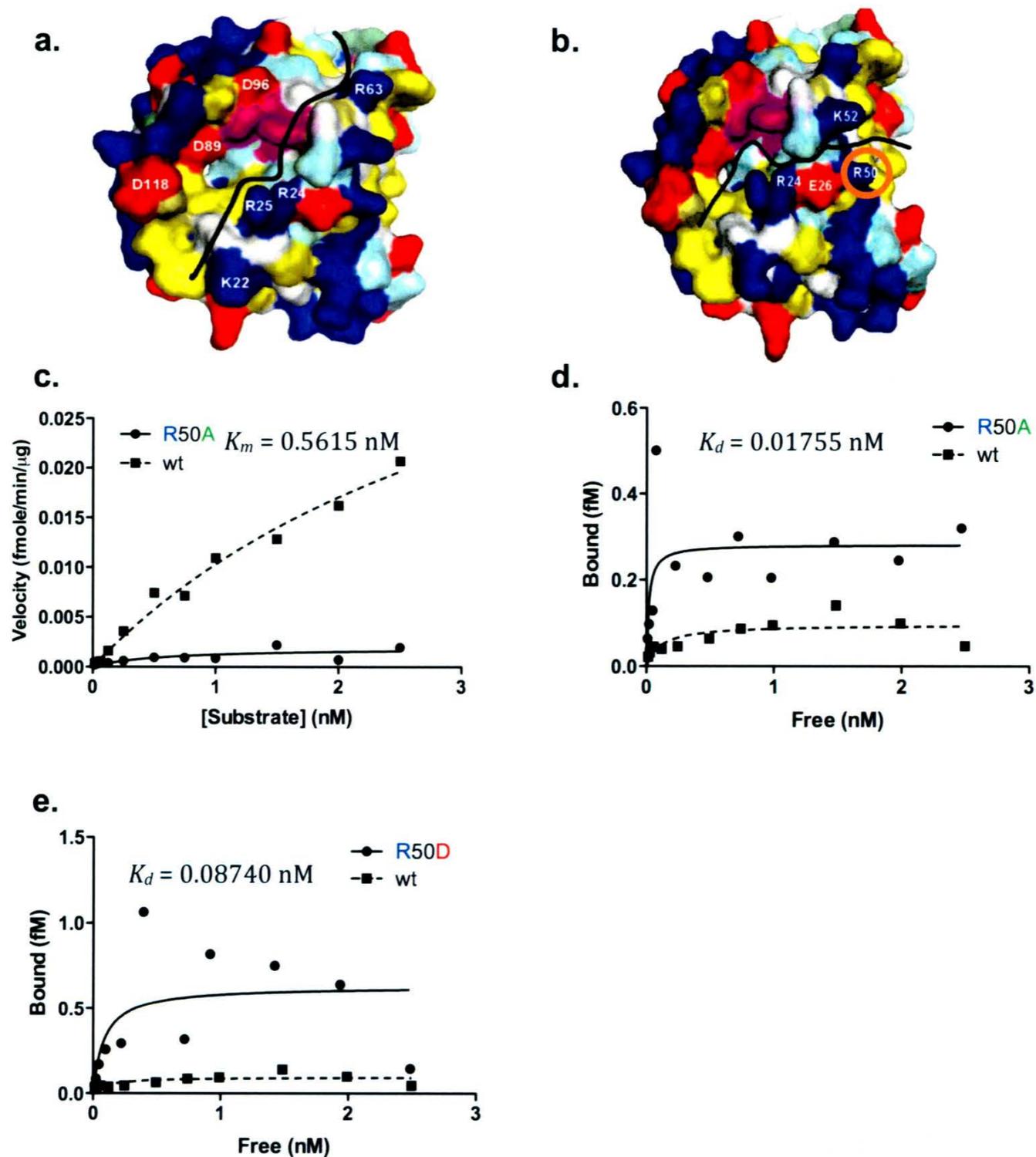
### Figure 23. Groove 2: E26 mutations and assays

Purified AID was assayed for enzymatic reaction velocity and for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b) of the enzyme. Velocity (c. and d.) and binding affinity (e. and f.) of E26R and E26A, relative to wt. Data represents three independent assays.  $K_m$  and  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.3.2 Groove 2: R50 results

The enzymatic activity of R50D was also diminished with less than 1 % product in comparison to 12 % of wt (Fig. 14b). Due to such little activity; no velocity was determined for this mutant enzyme. R50A similarly displayed a greatly reduced product, at 3 %, approximately one quarter that of wt (Fig. 14b). R50A had a greatly reduced calculated enzyme velocity of 0.5615 nM, approximately one-seventh that of wt, 3.824 nM indicating an increase in reaction velocity (Fig. 24c, Table 3).

R50D also showed reduced bound substrate of 1 % versus 1.5 % of wt (Fig. 15b). However, kinetic experiments yielded a 10-fold reduced  $K_d$  value of 0.08740 nM compared to wt of 0.1137, suggesting an increased in binding affinity (Fig. 24e, Table 3). Opposing these results, R50A displayed an increase in bound substrate compared to wt, with 3 % over 1.5 %, respectively. Likewise, R50A displayed an increase in binding affinity with a ten-fold reduced  $K_d$  value from wt of 0.1137 to 0.01755 nM of R50A (Fig. 24d, Table 3).



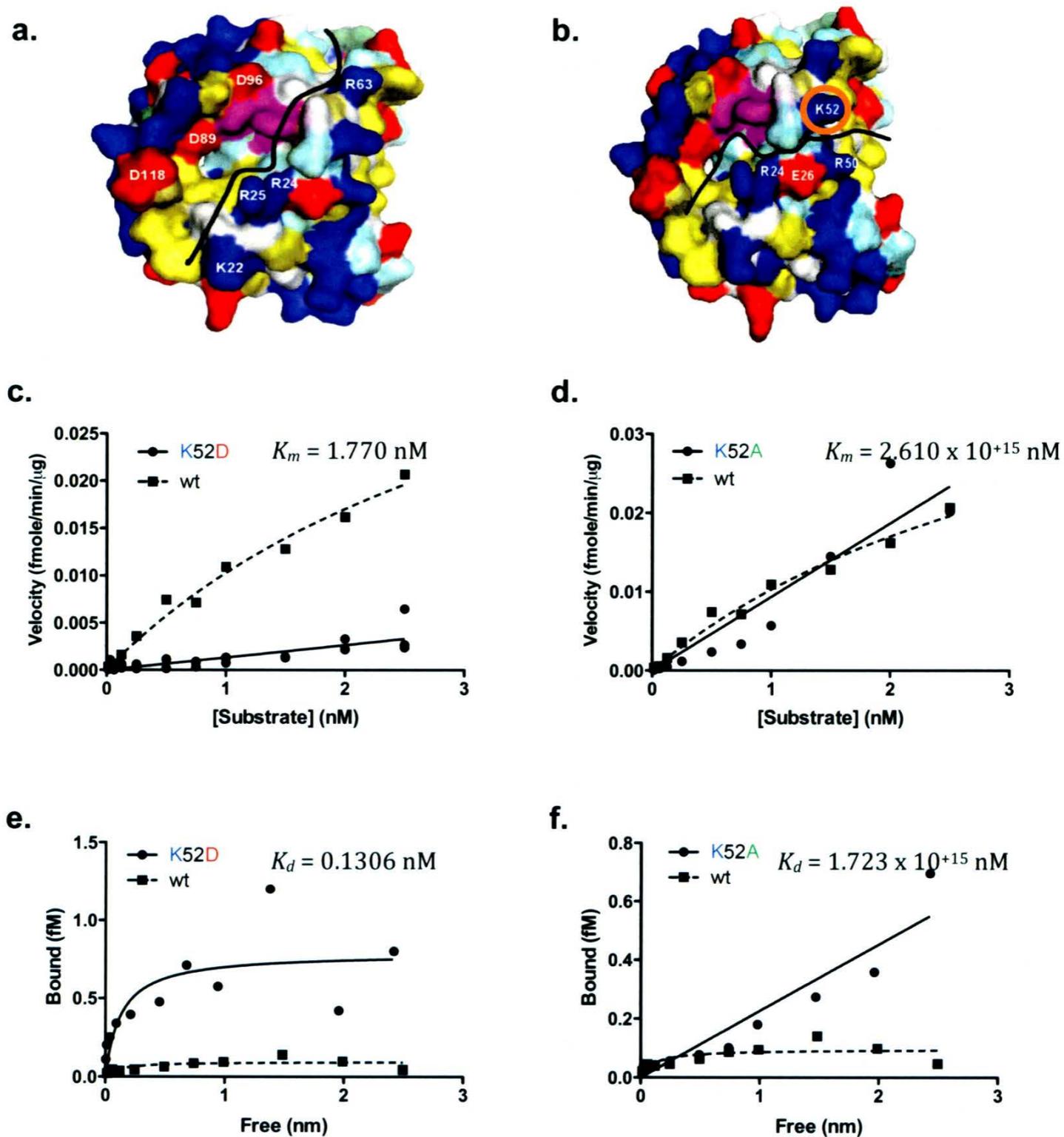
**Figure 24. Groove 2: R50 mutations and assays**

Purified AID was assayed for enzymatic reaction velocity and for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b.) of the enzyme. Velocity of R50A (c.) and binding affinity of R50A (d.) and R50D (e.), relative to wt. Data represents three independent assays.  $K_m$  and  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.3.3 Groove 2: K52 results

The enzymatic activity of K52D was reduced compared to wt with 1 and 12 %, respectively, suggesting reduced enzyme activity (Fig. 14b). Contrary to these results, the determined  $K_m$  for K52D was 1.770 nM approximately half of the wt value at 3.824 nM indicating an increase in reaction velocity (Fig. 25c, Table 3). K52A displayed decreased substrate product, at 7 % just over half of that of wt of 12 % indicating decreased activity (Fig. 14b). Likewise, K52A had an exponentially higher calculated  $K_m$  of  $2.610 \times 10^{15}$  nM compared to wt of 3.824 nM indicating a great decrease in reaction velocity (Fig. 25d, Table 3).

K52D showed increased bound substrate of 4 % versus 1.5 % of wt (Fig. 15b). However, kinetic experiments yielded a similar  $K_d$  value of 0.1306 nM compared to wt of 0.1137 nM, suggesting a slight decrease in binding affinity (Fig. 25c, Table 3). Opposing these results, K52A displayed a decrease in bound substrate compared to wt, with 1 % over 1.5 %, respectively (Fig. 15b). Likewise, K52A displayed a decrease in binding affinity with an exponentially increased  $K_d$  value from wt of 0.1137 nM to  $1.723 \times 10^{15}$  nM of K52A (Fig. 25, Table 3).



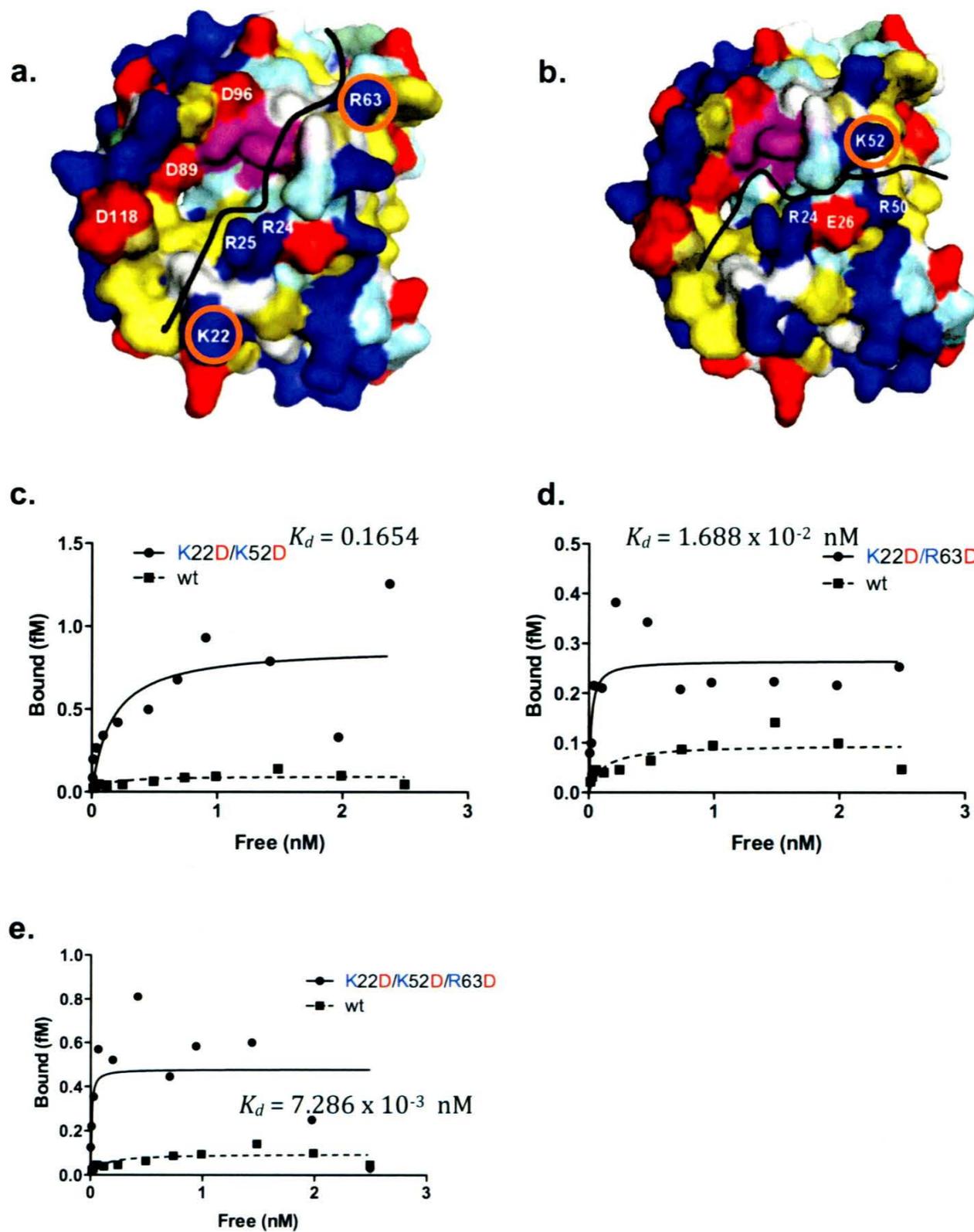
### Figure 25. Groove 2: K52 mutations and assays

Purified AID was assayed for enzymatic reaction velocity and for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b) of the enzyme. Velocity (c. and d.) and binding affinity (e. and f.) of K52D and K52A, relative to wt. Data represents three independent assays.  $K_m$  and  $K_d$  values were calculated from data plots in-software using non-linear regression.

#### 3.4.1 Multiple Mutants: K22D/K52D, K22D/R63D AND K22D/K52D/R63D

K22D/K52D and K22D/K52D/R63D have mutations along Grooves 1 and 2 while K22D/R63D has mutations solely in Groove 1. K22D/K52D, K22D/R63D and K22D/K52D/R63D showed no enzyme activity with approximately 0 % product formed during Alkaline Cleavage Assays compared to 12 % produced by wt (Fig. 14b). Due to lack of activity, no kinetics were run for these mutant enzymes and therefor no  $K_m$  values were determined.

A variety of results were found for binding assays. K22D/K52D showed an approximate 5-fold increase in bound substrate with 8 % versus 1.5 % of wt (Fig. 15b). Contrary to this results, kinetic binding experiments yielded a slight increase in  $K_d$  value compared to wt with 0.1654 nM and 0.1137 nM, respectively, indicating a slight decrease in binding affinity (Fig. 26c, Table 3). K22D/R63D yielded the same amount of bound substrate as wt with 1.5 % (Fig. 15b) for each. The  $K_d$  of K22D/R63D, however, was an order of magnitude smaller than that of wt, with 0.016883 nM compared to 0.1137 nM (Fig. 26d, Table 3). This corresponds with the decrease in  $K_d$  value but the K22D/R63D mutation yielded a significantly smaller value of 0.007286 nM compared to 0.1137 nM of wt, indicating a far high binding affinity for AID substrate (Fig. 26e, Table 3).



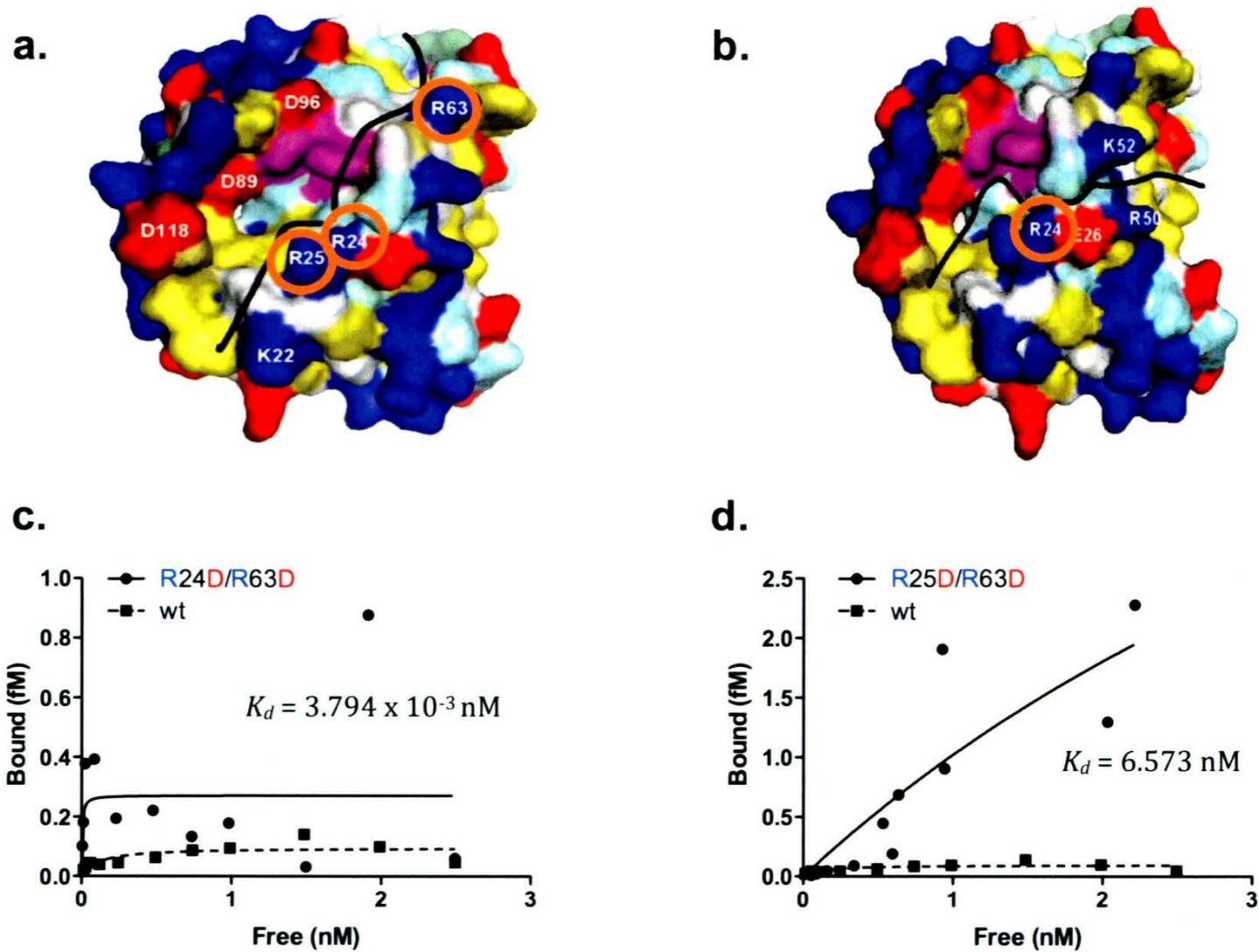
**Figure 26. Groove 1 and 2: Multiple Mutant K22D/K52D/R63D assays**

Purified AID was assayed for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b.) of the enzyme. Binding affinity of K22D/K52D (c.), K22D/R63D (d.), and K22D/K52D/R63D (e.), relative to wt. Data represents three independent assays.  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.4.2 Multiple mutants: R24D/R63D and R25D/R63D results

R25D/R63D and R24D/R63D although containing similar mutation yielded different results. Their activity values were respectively and approximately 0.5 % and 0 % product compared to 12 % of wt (Fig. 14b). Considered enzymatically dead, these mutants were not tested kinetically and have no calculated  $K_m$  values.

Binding experiments showed that R25D/R63D had an approximate 9-fold increase in bound product with 14 % compared to wt with 1.5 % and R24D/R63D showed an approximate 5-fold increase with 8 % (Fig. 15b). The  $K_d$  found for R25D/R63D does not correspond to an increase in binding, as it is far greater than wt of 0.1137 nM at 6.573 nM indicating a far smaller binding affinity for substrate than wt (Fig. 27d, Table 3). R24D/R63D resulted in a significantly smaller  $K_d$  value of 0.003794 nM compared to wt of 0.1137 nM suggesting a far higher binding affinity compared to wt (Fig. 27e, Table 3).



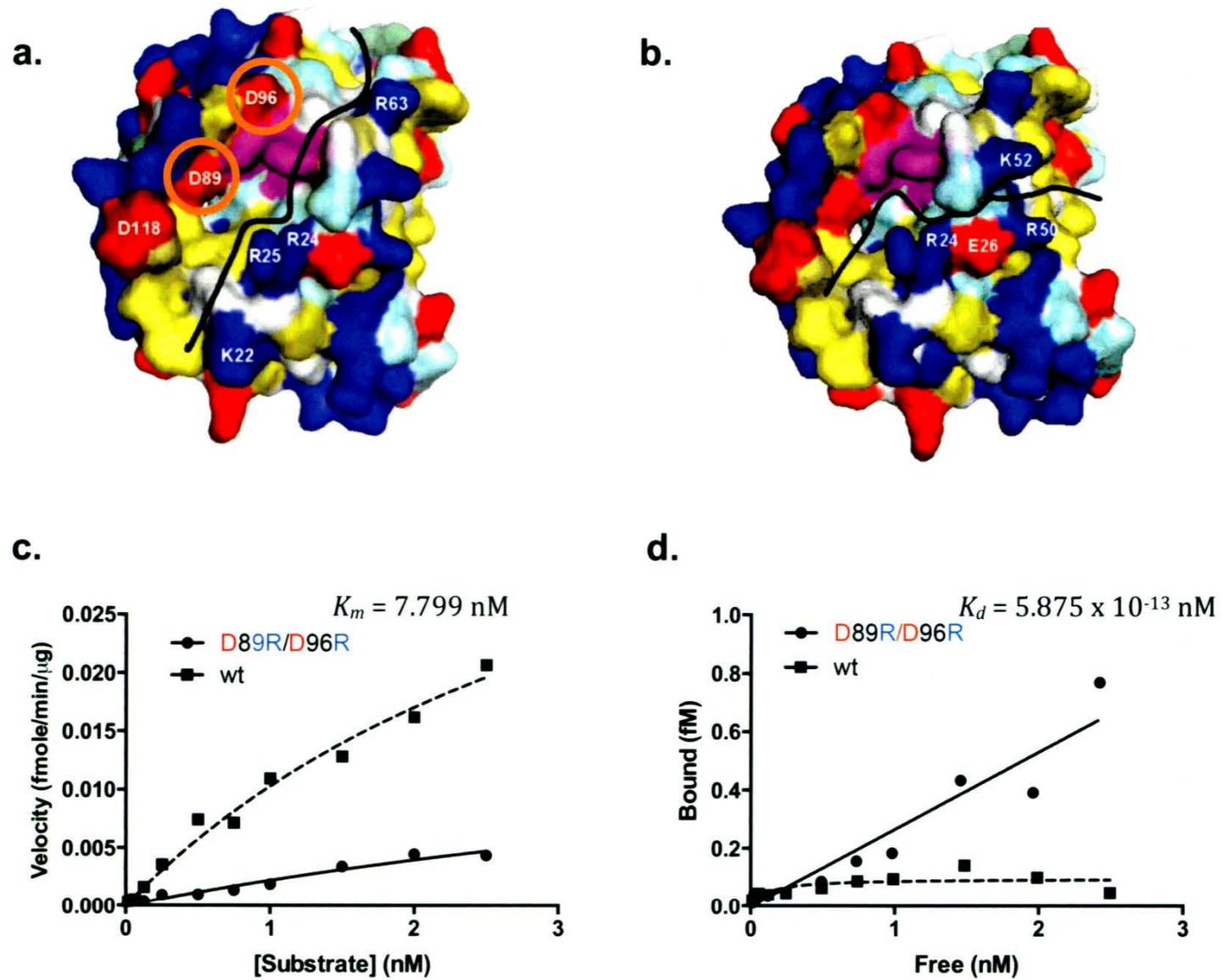
**Figure 27. Groove 1 and 2: Multiple Mutant R24D/R63D**

Purified AID was assayed for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b.) of the enzyme. Binding affinity of R24D/R63D (c.) and R25D/R63D (d.), relative to wt. Data represents three independent assays.  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.4.3 Multiple Mutants: D89R/D96R

The product yielded from D89R/D96R is approximately one sixth of that yielded from wt, 2 and 12 % respectively (Fig. 14b). This is consistent with the larger  $K_m$  value found for D89R/D96R of 7.799 nM versus 3.824 nM found for wt (Fig. 28b, Table 3).

In terms of bound substrate, D89R/D96R bound 1.5 % of substrate, equivalent to wt (Fig. 15b). Despite these values, D89R/D96R had an exponentially higher  $K_d$  compared to wt with  $5.875 \times 10^{+13}$  nM and 0.1137 nM respectively (Fig. 28c, Table 3). This suggests that the double mutation D89R/D96R has an exceptionally low binding affinity for ssDNA.



**Figure 28. Groove 1: Multiple Mutant D89R/D96R assays**

Purified AID was assayed for substrate binding affinity against a single-stranded DNA substrate in its native sequence (wt) or following the introduction of single or combination amino acid mutations in Groove 1 (a.) or Groove 2 (b.) of the enzyme. Velocity (c.) binding affinity (d.) of D89R/D96R, relative to wt. Data represents three independent assays.  $K_d$  values were calculated from data plots in-software using non-linear regression.

### 3.5 Mutation Summary

In terms of activity, Groove 1 mutations yielded two that increased activity and enzymatic velocities, R63A and D89R. Eight mutations decreased activity to such an extent that no kinetics were performed and samples were deemed enzymatically dead: R24D, R25A, R25D, double mutants R24D/R63D, R25D/R63D, K22D/K52D, K22D/R63D, and the triple mutant K22D/K52D/R63D. The following mutations were consistent in their results of decreasing enzymatic activity and kinetics, including calculated Michaelis constants ( $V_m$ ): K22D, K22A, R24A, R50A, K52D, D96R, D118R, and double mutant D89R/D96R. Of special note were some mutations, which produced inconsistent results, including D96R – although this mutation showed decreased deamination activity and kinetics, the  $V_m$  for this enzyme was actually higher than wt AID. D89A showed a decrease in general activity, but an increase in both catalytic activity and  $V_m$ . R63D, however, showed an increase in general activity but a decrease in catalytic activity and a decrease in  $V_m$ .

For Groove 2, R50D was deemed enzymatically dead. R50A and K52D were both consistent with decreased activity, kinetic, and  $V_m$  results. K52A, however, showed a decrease in activity but an increase in enzymatic velocity and  $V_m$  results. E26R and E26A produced very inconsistent results. E26R decreased activity, increased enzymatic velocity, and decreased  $V_m$ . E26A, however, increased activity, decreased enzymatic velocity, and yet increased  $V_m$ .

The majority of mutations showed consistent results in terms of binding and subsequent kinetic experiments. From Groove 1, K22D, R24D, R24A, K2D/K52D, K22D/K52D/R63D, R24D/R63D, R25D/R63D, R63D, R63A, and D89R were all consistent in increasing binding to ssDNA substrate in EMSA and kinetic assays. R25D, D89A, D96R, K22D/R63D, and D89R/D96R also increased AID substrate binding; but marginally – their bound substrate levels were comparable to wt AID. These all showed increased binding kinetics. D118R showed a decreased in bound substrate that was reflected in its kinetic experiments.

From mutations within Groove 1, K22A and R25A showed inconsistent results. They both decreased binding in the standard assay, yet their kinetics revealed a higher binding affinity.

For Groove 2, R50A and K52D showed an increase in binding in the standard and kinetic assays. Inconsistent results for E26R, E26A, R50D, and K52A were apparent. They all showed a decrease in substrate binding, however, they all showed an increase in their kinetic assays. It is of note that E26R and E26A also showed inconsistent results for deamination activity assays.

Mutant	Michaelis Constant, $K_m$ (nM)	Dissociation Constant, $K_d$ (nM)
wild-type	3.824	0.1137
K22D	5.313e-4	0.2864
K22A	10.36	0.1204
R24D	-	2.854e-2
R24A	-	1.856e-2
R25D	-	2.131e-2
R25A	-	0.1041
R63D	1.344e+15	8.846e-2
R63A	1.378e+14	0.3873
D89R	2.486e+19	0.4266
D89A	5.141	0.1240
D96R	2.145	1.664e-2
D118R	4.586	-1.209e-3
E26R	3.617e+12	2.070e-2
E26A	9.286e+18	1.941e-2
R50D	-	8.740e-2
R50A	0.5615	1.755e-2
K52D	1.770	0.1306
K52A	2.610e+15	1.723e+015
K22D/ K52D	-	0.1654
K22D/ R63D	-	1.688e-2
R24D/R63D	-	3.794e-3
R25D/R63D	-	6.573
K22D/ K52D/ R63D	-	7.286e-3
D89R / D96R	7.799	5.875e+13

**Table 3. Enzyme Kinetics constant Values**

Constants were calculated based on data from kinetic experiments using in software functions. Repeated experiments had data combined.

## Chapter 4 Discussion

To compare the effects of mutations generated within the enzyme on its activity as well as substrate binding, all experiments were also performed using human wt GST-AID. The  $K_m$ , and dissociation constant,  $K_d$ , were determined for wt AID. The generated kinetic curves and constants were used as a baseline for comparison. Some mutants were found to have significant deamination activity, and were tested kinetically, whereas all mutants underwent kinetic binding studies. The kinetic constants inversely reflect the behaviour they represent; a larger  $K_m$  value indicates a decrease in enzyme efficiency, while a larger  $K_d$  value demonstrates a decreased binding ability in comparison to the wt values.

### 4.1 Alkaline Cleavage and EMSA

For mutants with measurable kinetic activity and substrate binding, the individual alterations in the overall levels of deaminated and bound products were not always consistent with kinetic data used to determine  $K_m$  and  $K_d$ . One possible explanation could be the nature of the experimental strategy, where the means were based on 2-4 individual reactions at a fixed substrate concentration. The kinetic experiments involved 11 reactions over a wide range of substrate concentrations, with reactions at low substrate concentrations being crucial for constant determinations. Trends based on these experiments are more accurate than single concentration assays, as they include multiple data points; however, it should be

noted that some of the most altered  $K_m$  and  $K_d$  values at the extreme high and low ranges are only estimates, because the enzyme activities were affected by mutations to such an extent that saturation of activity or binding could not be reached at higher concentrations.

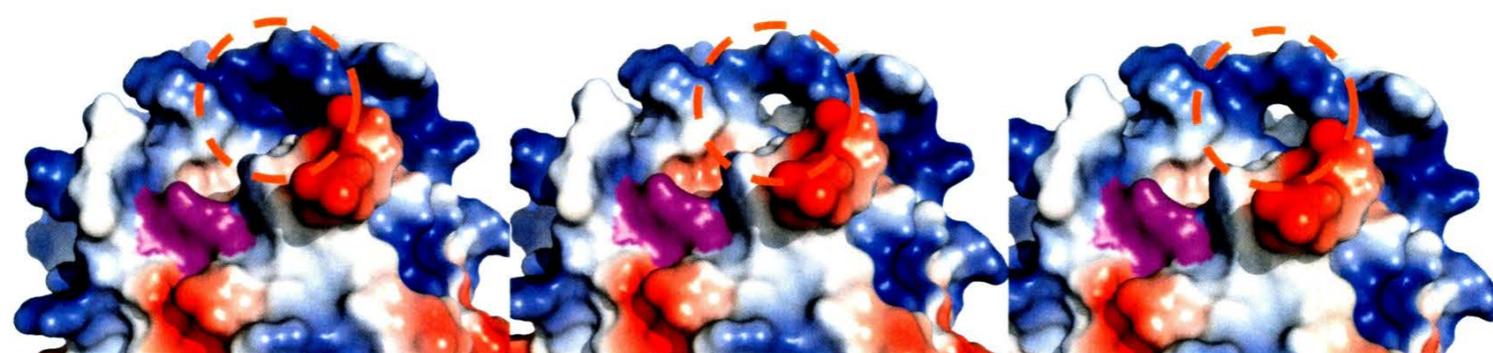
#### 4.1.1 Groove 1 Mutations

Mutation at residue 22 from positive lysine to aspartic acid (K22D) caused a decrease in  $K_m$  as shown in Table 3. Such a large decrease of the Michaelis constant from that of wt AID indicates higher enzymatic efficiency. This disagrees with our original hypothesis that the negatively charged arginine would repel rather than attract negatively charged single-stranded DNA, leading to lower enzyme activity, rather than an increase. This suggests that the charge of residue 22 is less important for activity than other residues, and that the observed changes in behaviour could be due to conformation changes within the protein itself. As well, the mutations lead to a lower binding affinity but this corresponded with an increase in enzyme activity. A mutation to a charge-neutral residue, alanine (K22A), had a lesser affect on activity as seen in Figure 16. The  $K_m$  increased, indicating lower enzyme efficiency. This suggests that the positive charge of lysine does play a role, albeit a minor one, in activity since a neutral mutation reduces enzyme efficiency.

Two other mutated residues located close to K22 were R24D and R24A, both of which appeared to be enzymatically inactive. Following our hypothesis, this could

be due to the loss in overall positive charge along the putative binding grooves from either mutation. Whilst the R24D drastic mutation could alter the overall folding of AID, this is far less likely for the R24A mutation. Therefore, the detrimental effect of these mutations on activity suggests that R24 is important for activity. For binding, both mutations show an overall trend of significantly increased bound substrate. This is further supported by EMSA kinetics where their  $K_d$  values were reduced by more than 10 fold, representing a large increase in binding ability due to R24D and R24A mutations. These effects may be due to conformational changes within the groove (Fig. 29). Although these results are counterintuitive, in that a positive to negative or neutral residue substitution was predicted to decrease DNA binding affinity, they suggest that R24 is somehow involved in DNA binding. Additionally, the vast increase in binding affinity may explain the drastic decrease in enzymatic activity, as the enzyme is unable to release a substrate in order to bind the next substrate molecule.

Both R25D and R25A appeared to be enzymatically inactive. Here, the change to a negative residue caused a more drastic increase in binding than when a neutral amino acid is substituted into the protein. Like the case with R24, this result is the opposite of what we would expect given our hypothesis; therefore, it appears that charge is not the primary component in AID behaviour. This suggests that R25 is also involved in ssDNA binding. As was the case with R24, this result further suggests that activity and binding are not completely interdependent, as some of our mutations exerted opposing affects on these two enzymatic properties.



Wild-type

AID R24D

AID R24A

**Figure 29. Models of R24 in wild-type AID, R24D mutation and R24A mutation.**

Models were generated using  $\pi$ mol software.

Blue= Positive, Red= Negative, White= Neutral, Purple= Catalytic Pocket

The R63A mutation, which resulted in an extremely high  $K_m$  and little enzyme activity, did not produce kinetic values as low as R63D. Binding for R63D showed a decrease in  $K_d$ , indicating an increase in binding efficiency. R63A however, shows more than a 3 times increase in  $K_d$  compared to wt, indicating a decrease in binding efficiency. This indicates that the positive charge of R63 may be important for ssDNA binding, but that the drastic R63D mutation caused a four-fold change. Overall, this data suggests R63 is important for AID activity and for binding, based on the change in kinetic coefficients.

R63A is a positive to neutral mutation that resulted in an increase in  $K_m$ , indicating that the enzyme was rendered less efficient. If the charge of this amino acid is involved in AID activity, the conversion to a neutrally charged residue should decrease activity with negatively charged substrate DNA, which was the observation. Conversion of positive R to a negatively charged residue D rendered AID enzymatically dead, further supporting the involvement of R63 and its positive charge with ssDNA.

Following our hypothesis, a negative to positive or neutral residue mutation would result in an increase in product formation. This was observed for mutants D89R and D89A, both of which resulted in an increase in deamination, D89R more so than D89A. D89R is a negative to positive mutation that increased activity. This is likely due to the increase in overall positive charge within the groove, which would more strongly attract and bind negatively charged single-stranded DNA. According

to alkaline cleavage kinetic constants, both D89R and D89A increased compared to wt AID, suggesting a paradoxical decrease in enzyme efficiency. D89R's  $K_m$  value increased exponentially, indicating an extreme decrease in enzyme efficiency. Both mutants also showed increased  $K_d$  values, indicative of a lower ssDNA binding affinity. Therefore, the negative charge of these two residues does not seem to repel ssDNA, and they are likely not positioned along the ssDNA binding site of AID.

The D96R mutation also results in a decreased  $K_m$  to just over half that of the wt. This decrease suggests higher enzyme efficiency over wt AID. In this case, the change from negative to positive charge can explain the increase in activity, as the negatively charged DNA would be more highly attracted to the positively charged R rather than the negatively charged D originally in its place. This is further supported by the decrease in dissociation constant by approximately 10-fold, interpreted as increased binding. The results suggest that D96 may be positioned close to the ssDNA binding groove.

The D118R mutation led to an increased estimated  $K_m$ , implying a decrease in enzyme efficiency compared to wt, correlating to the decrease in velocity as seen in the kinetic curve. This is unexpected, as the mutation to a positive residue theoretically would have increased activity. One interpretation is that D118 may not be directly involved in AID:substrate association. The estimated  $K_d$  was also vastly reduced, indicating a much higher binding affinity. Overall, D118R appears to decrease deamination and increase binding. This is possible if the enzyme binds to

the substrate for a much longer period of time, reducing the overall amount of deamination.

#### 4.1.2 Groove 2 Mutations

The next residue tested for activity kinetically was E26. Mutation to a positive residue, arginine, shows an increase in enzyme activity by the gross amount of product generated, but an increased  $K_m$  (Fig. 23). Such an increase implies lower enzyme efficiency due to the charge disruption; thus, E26 appears to be important for activity. Mutation of the same residue to neutral-charged alanine produces a similar effect to  $K_m$ , which agrees with the role of E26 in velocity. As shown in Figure 23, the kinetic curve shows an overall decrease in velocity. It can be said that disrupting the negative charge hampers AID activity, suggesting that this residue is involved in AID activity through an unknown mechanism. Since this is a negatively charged residue, contact with the phosphate backbone of ssDNA may be a possibility. In terms of binding, E26R lowered the  $K_d$ , translating to an increase in binding ability, which may be due to the positive charge of arginine attracting negatively charged ssDNA. E26A slightly increased the  $K_d$  value and lowered binding ability, meaning that E26 is involved in AID activity and might be a factor in binding.

R50D also appeared to be enzymatically inactive in the activity assay; however, it caused a 10-fold decrease in  $K_d$  value. This result may imply that this mutant alters the overall folding of the enzyme. R50A resulted in a  $K_m$  value that was

15 % that of wt AID, indicating increased enzyme efficiency. This suggests that the electrostatic attraction of this residue and ssDNA is not important for AID activity. Rather, a change in the size and/or shape of the binding groove could change efficiency of deamination. In terms of binding, the  $K_d$  was reduced even more than R50D, also indicating an increase in binding efficiency of AID due to this mutation. This outcome supports the idea that charge may be a less important factor than size for maintaining the integrity of the putative ssDNA binding groove.

Overall, the unexpected results for many of these mutations may be explained through size and distortion of the suggested binding grooves, rather than through electrostatic forces alone. Alternatively, the fact that many of these mutations from positively charged to negatively charged residues actually enhanced the binding of ssDNA, may indicate that these residues are not involved in direct contact with ssDNA, or that they are involved in ssDNA binding in an indirect and more complex manner, where the change to a negative residue may cause the attraction and change in conformation of a nearby positive residue "into" the putative groove, thereby mediating the overall mutation effects.

K52D is another mutation that increased AID deamination efficiency based on a decrease in  $K_m$ . ssDNA binding kinetics revealed a dissociation constant with an approximate 20-fold increase over that of wt. This suggests that binding ability is decreased, so K52 in its natural state is important for binding, whether due to charge or to conformation. Insight into the role of charge for this residue is evident with the

K52A mutation. Here, activity was eradicated and the mutation reduced binding even further with a higher order magnitude  $K_d$  value. From this, we can gather that this residue may be involved in direct or near direct contact with ssDNA. R63D was the next amino acid for mutation analysis, and it was also found to be enzymatically inactive, and thus initially indicated that both the residue identity and position were important for AID activity.

#### 4.1.3 Multiple Mutants

K22D/K52D showed little or no enzyme activity. Binding experiments showed similar results to that of wt; the overall curve trend of reaction velocity shows an increase in binding, but the  $K_d$  is slightly above that of wt AID, which means it is slightly more effective at substrate binding.

The K22D/R63D multiple mutant showed no deamination activity, with no rescue of enzyme activity by K22D for R63D. This suggests that R63D has a more significant role in AID activity than K22D. When substrate binding was assayed, the resulting  $K_d$  value was about one tenth that of wt, indicating that a much greater binding efficiency was achieved. This value is smaller than either R63D or K22D, the latter of which increased  $K_d$  compared to wt, thus it had a compounded effect.

K22D/K52D/R63D, the only triple mutant, was also enzymatically dead. K22D and K52D mutations cannot alleviate the loss of activity brought on by R63D. This could suggest that R63D is more crucial to deamination activity than either

K22D or K52D. Binding kinetics reveal a dissociation constant about half that of K22D/R63D, and 1/50th of K22D/K52D. Ergo, the triple mutant appears to greatly enhance binding ability over either of its constituent mutations.

R24D/R63D also showed no deamination activity, which was expected given that both mutations showed no AID activity. This can be explained by the loss of positively charged residues or possibly substrate contact site distortion. Binding experiments showed that this is the second smallest  $K_d$  value of all mutations assayed. This double mutation led to increased binding efficiency relative to wt AID, more so than either of its constituent mutations, which verifies their importance in AID binding.

R25D/R63D, similar to other AID enzymes containing multiple mutations, showed no deamination activity. As the charges of the amino acids were converted from positive to negative, repulsion to substrate ssDNA can account for this lack of activity. Likewise, EMSA kinetics experiments resulted in a  $K_d$  value about a 60-fold increase over wt, showing a greatly reduced substrate binding efficiency. This was an unexpected result, as both of the mutations individually increased the binding affinity; thus, it was expected that the combination would compound the increase in substrate affinity. The reduction in activity and binding was most likely due to conformational changes to the putative binding grooves, AID protein folding, or a combination of both.

D89R/D96R mutation showed an increased  $K_m$  suggesting a decrease in enzyme efficiency. The  $K_d$  value increased on the order of  $10^{13}$ , indicating almost complete ablation of ssDNA binding. This is fitting, since little activity resulted from this double mutation. From this evidence, we can gather that D89 is more involved in activity than D96. In terms of binding kinetics, the multiple mutant enzyme also has a higher  $K_d$  value, indicating that the D96R mutation is unable to rescue the drastic effect of the D89R mutation. Therefore, it is likely that D89 is more involved in the overall structural integrity of the protein while D96 plays a more direct role in ssDNA binding.

#### **4.2 The relationship between catalytic activity and ssDNA binding affinity of AID**

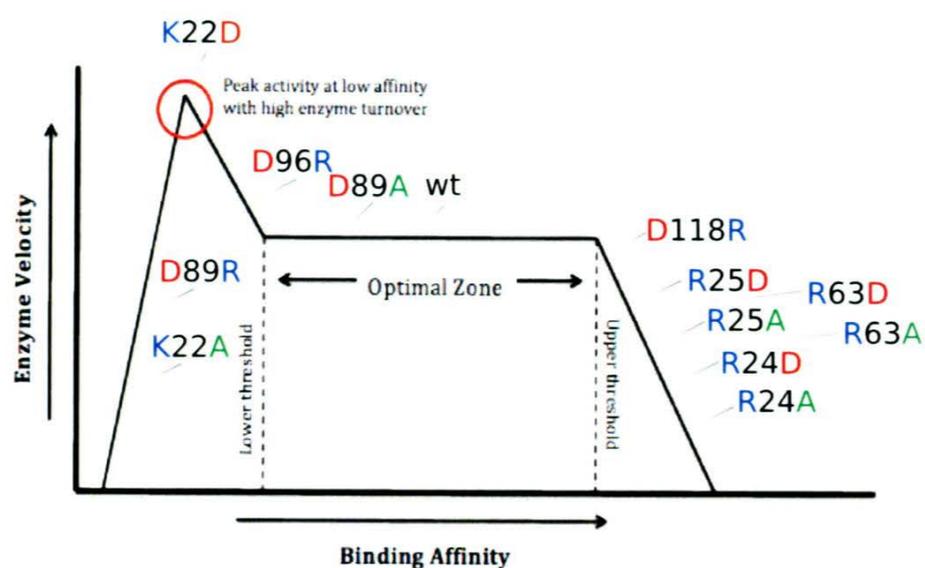
By comparing activity and affinity data, correlation between the two can be evaluated. A model emerges where there exists an optimal enzyme-substrate affinity zone in which the typical levels of AID activity can be achieved as substrate affinity and catalytic activity are balanced. Below this threshold, enzymes would have a very low affinity that results in lower activity. Likewise, there would exist an affinity threshold for optimal activity. If bound in the enzyme:substrate complex too long, the overall AID activity would decrease reducing the potential to deaminate multiple substrate targets (Fig. 30).

The results of our experiments are highly suggestive of such a model, with the optimal zone of moderate enzymatic catalytic activity and binding affinity

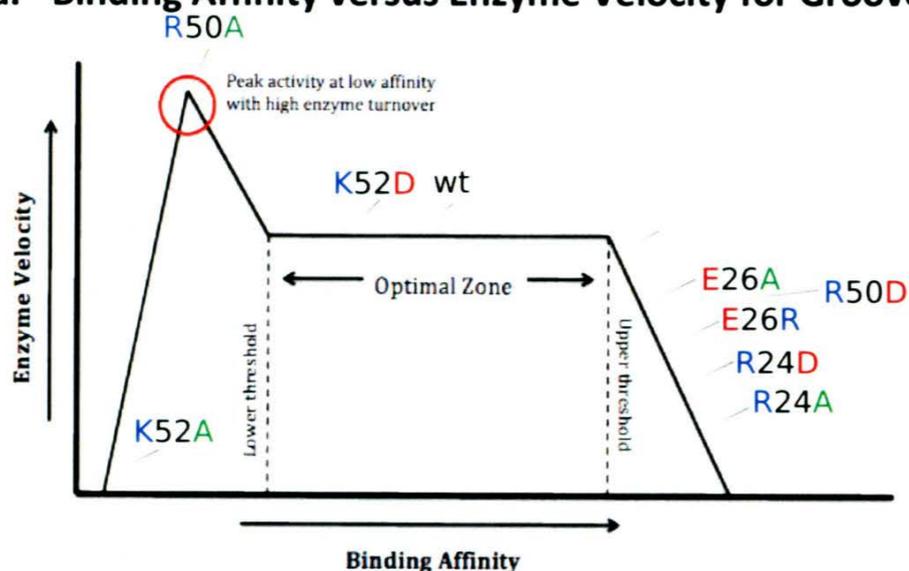
assigned to wt AID levels (Fig. 30). Similar profiles were observed for K52D, D96R, and D89A mutations, which display activity and binding kinetics like wt AID. Mutant AID proteins that fall below the optimal affinity threshold with either low catalytic or high catalytic activity included: D118R, K22A, R63A, R63A, D89R, K22D/K52D, R25D/R63D. Also included in this group was K22D, that exhibited very high enzymatic activity and low binding affinity.

The last category includes mutant enzymes that displayed low catalytic activity, presumably due to overly high substrate binding affinities, such that enzyme turnover rates were too slow to permit high AID activity. This category was the largest grouping, suggesting these amino acid residues are the most sensitive to change and have great importance to AID binding affinity and activity. The mutations in this category included: D89R/D96R, E26R, R63D, K52A, E26A, R24D/R63D, K22D/K52D/R63D, K22D/R63D, R24A, R25D, R24D, R50D, and R25A. Interestingly, this group includes all types of residue switches, ranging from positive to negative, negative to positive, or charged to neutral residue changes.

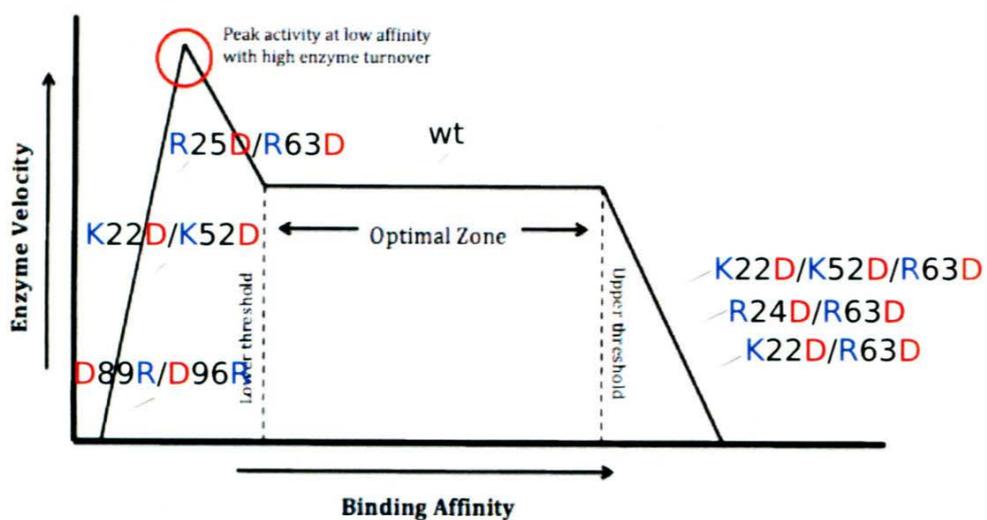
Overall, our data suggests that the binding mechanism of ssDNA by AID is easily disrupted by changes within its amino acid sequence and resulting structure. Subsequently, the enzyme can easily lose its optimal catalytic activity to binding affinity ratio. Based on the disruptions seen in activity and binding, Groove 1 appears to be the most likely site for substrate interaction.



a. Binding Affinity versus Enzyme Velocity for Groove 1 Mutants



b. Binding Affinity versus Enzyme Velocity for Groove 2 Mutants



c. Binding Affinity versus Enzyme Velocity for Multiple Mutants

**Figure 30. Binding Affinity versus Enzyme Velocity.**

Illustration of mutant AID velocity and binding affinity, not to scale for (a.) Groove 1 mutants, (b.) Groove 2 mutants and (c.) multiple mutants.

### **4.3 The role of AID surface structure in enzyme activity and binding**

The mutant enzymes had various levels of activity and binding compared to their wt counterpart. This suggests that both the charge and the size of amino acids can be imperative for AID and substrate interaction based on a delicate balance of residue shape and charge. Attractive forces would be required for substrate and enzyme interaction, while repulsion could act as a means to foster substrate turnover; reducing the time AID is involved in a protein:DNA complex. As seen in side-by-side structural modeling comparisons of wt R24, R24D, and R24A AID models (Fig. 29) the mutations can have dramatic effects, changing not only charge but also shape within the putative binding groove. This example shows a “tunnel-like” effect that develops with the introduction of mutations at this residue, in addition to new charge and charge distribution. It is possible that the other mutations have also introduced a similar disruption in shape, which may explain the unexpected behaviour seen with some of the AID mutations.

Of the twenty-four mutations induced within the AID sequence, only four increased activity efficiency: D96R, K52D, R50A, and K22D. Since three of these reduced the overall positive charge within the putative binding grooves, it can be said that the electrostatic interaction between residues and target ssDNA does not follow the hypothesis suggested. Positively charged residues do not always enhance activity with negatively charged substrate. A revised hypothesis suggests that other factors are responsible for AID activity and interaction.

Based on these results, the mutations that caused the greatest disruption to activity are the most likely residues involved in deamination. According to  $K_m$  values, those residues are on either end of the spectrum:  $K_m$  (largest to smallest) D89R>E26A>R63A>E26R>K22A>D89R/D96R>D89A>D118R>wt>D96R>K52D>R50A>K22D. Except for D96R, D118R, and D89A, the mutations drastically increased enzyme efficiency (2-fold or more) or decreased it (by half). Overall these experiments show that residues R24, R25, E26, R63, and D89, which best align with Groove 1 (Fig. 5), are highly involved in AID activity.

All mutant enzymes were tested kinetically for binding. When listed according to dissociation constant, there is no trend between charged residues and binding ability.  $K_d$  (largest to smallest): K52A>D89R/D96R>R25D/R63A>K22D>E26A>K22D/K52D>D89A>K22A>wt>R25A>R63D>R50D>R24D>R25D>E26R>R24A>R50A>K22D/R63D>D96R>K22D/K52D/R63D>R24D/R63D>D118R. That is, the proposed hypothesis does not hold true for all residues. Those mutations which most greatly affected binding were K52A, D89R/D96R, K22D/K52D/R63D, R24D/R63D, and D118R, which all increased binding efficiency. From this, the residues most actively involved in substrate binding appear to be: K22, R25, K52, R63, D89, D96, and D118. With the exception of K52, these amino acids support Groove 1 as a likely contact site within AID for binding. Introduction of mutations to K52 may disrupt overall protein folding or charge distribution along the protein surface, explaining why changes to this amino acid can be detrimental to AID binding function.

Generated models of mutant AIDs show drastic shape changes within the putative binding grooves as a result of mutations, which can explain some of the unexpected behaviours seen in our assays (Fig. 29). Certain residue changes resulted in loss of protrusions along the grooves. An unexpected change in activity can imply that a residue's role in AID activity is not charge-centred, based on our hypothesis that a positively charged residue in theory would attract negatively charged DNA. Rather, differences in activity can appear due to a conformational change within the binding groove. For instance, by molecular weight, arginine is 174.2 g/mol versus alanine at 89.1 g/mol, the largest amino acid weight change of all the mutations. Decrease in size could allow more room, and thus greater accessibility, for ssDNA to "fit" and contort into the groove site. Additionally, a mutation may cause a lapse in protein folding, thus altering enzyme substrate interaction in this matter. Overall, our work suggests that the change in amino acid size, and thus the shape and size of the suggested binding grooves, may be a much more significant determinant of ssDNA binding and activity, than the change in the charge itself.

Altogether, activity and binding of AID are not solely dependent on the charge distribution of surface residues; however our results do seem to implicate a higher number of residues along putative Groove 1, rather than putative Groove 2. Similar findings support the involvement of surface residues in substrate interactions, including a polynucleotide-binding region alike to APOBEC enzymes (Wang, 2009). Specific implicating residues are R and K surface residues, R24 and residues

surrounding D118 of AID such as A111, R112 and L113 (Mu, et al., 2012; Hu et al., 2013; Doi et al., 2008), which align with our results.

#### 4.4 Future Studies

The results from these experiments shed some light on surface residues of AID that impact ssDNA binding, as well as on the complexity of the relationship between the enzymatic rates of AID and its ssDNA binding affinity. There are several areas that should be explored to better understand AID and these results.

It is noteworthy that the range in size of amino acids mutated varies greatly. This in turn may explain some of the observed differences in activity and binding, in addition to charge/electrostatic interaction. From largest to smallest: lysine (K, 146 g/mol), arginine (R, 174 g/mol), aspartic acid (D, 133.11 g/mol), glutamic acid (E, 147.13 g/mol), and alanine (A, 89.10 g/mol). Because of this, further studies should include mutations that consider a charge to size ratio when selecting amino acids for mutagenesis. The shape of the binding site and the density of charge could offset attraction or repulsion to target DNA – which factor accounts for the changes seen in these experiments is unclear, and could be better resolved. For example, conversion of Glutamic acid or Aspartic acid to Lysine instead of Arginine would allow for a more reliable study of charge difference on AID activity. Glutamine is a likely candidate for conversion to another charge neutral residue closer in size. Since these amino acids are closer in size to the amino acids they are replacing, size and shape would be eliminated as a major factor for any differences in activity and binding exhibited, if the mutations do not cause any major protein folding errors. These would more likely be due to the altered charge rather than due to the altered size.

Finally, in terms of multiple mutations, there are other interesting scenarios to explore. Of the 24 mutations induced, 8 of them decreased deamination activity. Of special note are K22D, R50A, and K52D all of which rendered AID nearly inactive. The significance of each of their roles in activity can be determined by combining these mutations with others that greatly increased binding. If the introduction of an active mutation rescues the activity of either of the "dead" mutants, then the new mutation can be said to have a more detrimental role in activity. If the activity is not rescued, the primary mutation has a more integral role in enzyme activity.

Since charge does not always appear to affect activity or binding, the possibility of other contact sites should be explored more directly using various footprinting techniques. Although far more challenging technically (especially for a highly charged molecule such as AID), these assays would complement our mutagenesis approach and together paint a more accurate picture of residues directly involved in substrate binding.

## Chapter 5 Bibliography

- Barreto, V. M., Pan-Hammarstrom, Q., Zhao, Y., Hammarstrom, L., Misulovin, Z., & Michel, C. N. (2005). AID from bony fish catalyzes class switch recombination. *Journal of Experimental Medicine*, 202, 733-8.
- Beale, R. C., Svend, K. P., Ian, N. W., Reuben, S. H., Rada, C., & Michael, S. N. (2004). Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: Correlation with mutation spectra in vivo. *Journal of Molecular Biology*, 337, 585-96.
- Bhutani, N., Brady, J. J., Damian, M., Sacco, A., Corbel, S. Y., & Blau, H. M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature*, 463, 1042-7.
- Bransteitter, R., Pham, P., Scharff, M. D., & Goodman, M. F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings of the National Academy of Science of the United States of America*, 100, 4102-7.
- Brar, S. S., Sacho, E. J., Tessmer, I., Croteau, D. L., Erie, D. A., & Diaz, M. (2008). Activation-Induced Deaminase, AID, is catalytically active as a monomer on single-stranded DNA. *DNA Repair (Amst)*, 7 (1), 77-87.
- Canugovi, C., Samaranayake, M., & Bhagwat, A. S. (2009). Transcriptional pausing and stalling causes multiple clustered mutations by human activated-induced deaminase. *The FASAB Journal*, 23, 34-44.
- Carpenter, M. A., Rajagurubandara, E., Wijasinghe, P., & Bhagwat, A. S. (2010). Determinants of sequence-specificity within human AID and APOBEC3G. *DNA Repair (Amst)*, 9, 579-87.
- Carter Jr., C. W. (1995). The nucleoside deaminases for cytidine and adenosine: structure, transition state stabilization, mechanism, and evolution. *Biochimie*, 77 (1-2), 92-8.
- Chaudhuri, J., & Frederick, W. A. (2004). Class-switch recombination: Interplay of transcription, DNA deamination and DNA repair. *Nature reviews Immunology*, 4, 541-52.
- Chaudhuri, J., Basu, U., Zarrin, A., Yan, C., Franco, S., & Perlot, T. (2007). Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Advances in Immunology*, 94, 157-214.

- Chen, K. M., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y., & Shindo, K. (2008). Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature*, *452*, 116-9.
- Coker, H. A., & Petersen-Mahrt, S. K. (2007). The nuclear DNA deaminase AID functions distributively whereas cytoplasmic APOBEC3G has a processive mode of action. *DNA Repair (Amst)*, *6*, 235-43.
- Coker, H. A., Morgan, H. D., & Petersen-Mahrt, S. K. (2006). Genetic and in vitro assays of DNA deamination. *Methods in Enzymology*, *408*, 156-70.
- Conticello, S. G., Langlois, M. A., Yang, Z., & Neuberger, M. S. (2007). DNA deamination in immunity: AID in context of its APOBEC relatives. *Advances in Immunology*, *94*, 37-73.
- Dancyger, A. M., King, J. J., Quinlan, M. J., Fifield, H., Tucker, S., Saunders, H. L., et al. (2011). Differences in the enzymatic efficiency of human and bony fish AID are mediated by a single residue in the C terminus modulating single-stranded DNA binding. *The FASEB Journal*, *26*, 1517-1525.
- Dayn, A., Malkhosyan, S., & Mirkin, S. M. (1992). Transcriptionally driven cruciform formation in vivo. *Nucleic Acids Research*, *20*, 5991-7.
- de Yébenes, V. G., & Almudena, R. R. (2006). Activation-induced deaminase: light and dark sides. *TRENDS in Molecular Medicine*, *12* (9), 432-9.
- Deng, L., Velikovskiy, C. A., Xu, G., Iyer, L. M., Tasumi, S., Kerzic, M. C., et al. (2010). A structural basis for antigen recognition by the T cell-like lymphocytes of sea lamprey. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 13408-13.
- Di Noia, J. M., & Neuberger, M. S. (2007). Molecular mechanism of antibody somatic hypermutation. *Annu. Rev. Biochem.*, *76*, 1-22.
- Di Noia, J., & Neuberger, M. S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature*, *419*, 43-8.
- Dickerson, S. K., Market, E., Besmer, E., & Papavasiliou, F. N. (2003). AID Mediates Hypermutation by Deaminating Single Stranded DNA. *The Journal of Experimental Medicine*, *197*, 1291-6.
- Doi, T., Kato, L., Ito, S., Shinkura, R., Wei, M., Nagaoka, H., et al. (2009). The C-terminal region of activation-induced cytidine deaminase is responsible for a recombination function other than DNA cleavage in class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America*, *106* (8), 2758-63.

- Durandy, A., Peron, S., & Fischer, A. (1996). Hyper-IgM syndromes. *Curr Opin Rheumatol*, 18 (4), 751-8.
- Durandy, A., Revy, P., Imai, K., & Fischer, A. (2005). Hyper-IgM syndromes caused by intrinsic B-lymphocytes defects. *Immunol Rev*, 203, 67-79.
- Fleisher, T. A., & Tomar, R. H. (1997). Introduction to diagnostic laboratory Immunology. *JAMA*, 278 (22), 1823-57.
- Fukita, Y., Jacobs, H., & Rajewsky, K. (1998). Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity*, 9, 105-14.
- Gazumyan, A., Timachova, K., Yuen, G., Siden, E., Di Virgilio, M., Woo, E. M., et al. (2011). Amino-Terminal Phosphorylation of Activation-Induced Cytidine Deaminase Suppresses c-Myc/IgH Translocation. *Molecular Cellular Biology*, 31 (3), 442-9.
- Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem*, 71, 101-32.
- Gordon, M. S., Kanegai, C. M., Doerr, J. R., & Wall, R. (2003). Somatic hypermutation of the B cell receptor genes B29 (Ig beta, CD79b) and mb1 (Ig alpha CD79a). *Proceedings of the National Academy of Sciences*, 100, 4126-31.
- Heintel, D., Kroemer, E., Keinle, D., Schwarzinger, I., Gleiss, A., Schwarzmeier, J., et al. (2004). High expression of activation-induced cytidine deaminase (AID) mRNA is associated with unmutated IGHV gene status and unfavourable cytogenic aberrations in patients with chronic lymphocytic leukaemia. *Leukemia*, 18 (4), 756-62.
- Holden, L. G., Prochnow, C., Chang, Y. P., Bransteitter, R., Chelico, L., & Sen, U. (2008). Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature*, 456, 121-4.
- Holz, B., Klimasauskas, S., Serva, S., & Weinhold, E. (1998). 2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases. *Nucleic Acids Research*, 26, 1076-83.
- Hu, Y., Ericsson, I., Torseth, K., Methot, S. P., Sundheim, O., Liabakk, N. B., et al. (2013). A Combined Nuclear and Nucleolar Localization Motif in Activation-Induced Cytidine Deaminase (AID) Controls Immunoglobulin Class Switching. *Journal of Molecular Biology*, 425 (2), 424-43.
- Huang, F. T., Yu, K., Balter, B. B., Selsing, E., Oruc, Z., & Khamlichi, A. A. (2007). Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain S<sub>mu</sub> class switch region. *Molecular and Cellular Biology*, 27, 5921-32.

- Ireton, G. C., Black, M. E., & Stoddard, B. L. (2003). The 1.14 Å crystal structure of yeast cytosine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. *Structure*, *11*, 961-72.
- Jiang, Y. L., Kwon, K., & Stivers, J. T. (2001). Turning On uracil-DNA glycosylase using a pyrene nucleotide switch. *The Journal of Biological Chemistry*, *276*, 42347-54.
- Jiang, Y. L., Stivers, J. T., & Song, F. (2002). Base-flipping mutations of uracil DNA glycosylase: substrate rescue a pyrene nucleotide wedge. *Biochemistry*, *41*, 11248-54.
- Kao, Y.-C., Cam, L. L., Laughton, C. A., & et al. (1996). Binding Characteristics of Seven Inhibitors of Human Aromatase: A Site-directed Mutagenesis Study. *Cancer Research*, *56*, 3451-60.
- Klemm, L., Duy, C., Iaobucci, I., Kuchen, S., Gregor, V. L., & Feldhahn, N. (2009). The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. *Cancer Cell*, *16*, 232-45.
- Ko, T. P., Lin, J. J., Hu, C. Y., Wang, A. H., & Liaw, S. H. (2003). Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution. *Journal of Biological Chemistry*, *278* (21), 19111-7.
- Kohli, R. M., Abrams, S. R., Gajula, K. S., Maul, R. W., Gearhart, P. J., & Stivers, J. T. (2009). A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *The Journal of Biological Chemistry*, *284*, 22898-904.
- Kuppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. *Nature Reviews Cancer*, *5*, 251-62.
- Kuppers, R., & Dalla-Favera, R. (2001). Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*, *20*, 5580-94.
- Lanasa, M. C., & Weinberg, J. B. (2011). Immunoglobulin class switch recombination in chronic lymphocytic leukemia. *Leuk Lymphoma*, *52* (7), 1398-1400.
- Langlois, M., Rupert, C. L., Silvestro, G. C., & Michael, S. N. (2005). Mutational comparison of the single-dominated APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Research*, *33*, 1913-23.
- Larijani, M., & Martin, A. (2007). Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. *Molecular and Cellular Biology*, *27*, 8038-48.

- 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*, 840-5.
- Larijani, M., Frieder, D., Sonbuchner, T. M., Bransteitter, R., Goodman, M. F., & Bouhasira, E. E. (2005). Methylation protects cytidines from AID-mediated deamination. *Molecular Immunology*, *42*, 599-604.
- 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. *Molecular and Cellular Biology*, *27*, 20-30.
- Liu, M., Duke, J. L., Richter, D. J., Vinuesa, C. G., Goodnow, C. C., Kleinstein, S. H., et al. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature*, *451* (7180), 841-5.
- Longerich, S., Basu, U., Alt, F., & Storb, U. (2006). AID in somatic hypermutation and class switch recombination. *Current Opinion in Immunology*, *18*, 164-74.
- Losey, H. C., Ruthenburg, A. J., & Verdine, G. L. (2006). Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nature Structural & Molecular Biology*, *13*, 153-9.
- Lossos, I. S., Levy, R., & Alizadeh, A. A. (2004). AID is expressed in germinal center B-cell-like and activated B-cell-like diffuse large-cell lymphomas and is not correlated with intraclonal heterogeneity. *Leukemia*, *18* (11), 1775-9.
- MacGinnitie, A. J., Anant, S., & Davidson, N. O. (1995). Mutagenesis of apobec-1, the Catalytic Subunit of the Mammalian Apolipoprotein B mRNA Editing Enzyme, Reveals Distinct Domains That Mediate Cytosine Nucleoside Deaminase, RNA Binding, and RNA Editing Activity. *The Journal of Biological Chemistry*, *270*, 14768-75.
- Martin, A., Bardwell, P. D., Woo, C. J., Fan, M., Shulman, M. J., & Scharff, M. D. (2002). Activation induced cytidine deaminase turns on somatic hypermutation in hybridomas. *Nature*, *415*, 802-6.
- Martin, A., Li, Z., Lin, D. P., Bardwell, P. D., Iglesias-Ussel, M. D., Edelman, W., et al. (2003). Msh2 ATPase Activity is Essential for Somatic Hypermutation at A-T Basepairs and for Efficient Class Switch Recombination. *The Journal of Experimental Medicine*, *198*, 1171-8.
- McBride, K. M. (2004). Somatic hypermutation is limited by 346 CRM1-dependent nuclear export of activation-induced deaminase. *Journal of Experimental Medicine*, *199*, 1235-1244.

McCarthy, H., Wierda, W. G., Barron, L. L., Cromwell, C. C., Wang, J., Coombes, K. R., et al. (2003). High expression of activation-induced cytidine deaminase (AID) and splice variants is a distinctive feature of poor prognosis chronic lymphocytic leukemia. *Blood*, *101*, 4903-8.

Migliazza, A., Martinotti, S., Chen, W., Fusco, C., Ye, B. H., Knowles, D. M., et al. (1995). Frequent somatic hypermutation of the 5' noncoding region of the BCL6 gene in B-cell lymphoma. *Proceedings of the National Academy of Sciences U.S.A.*, *92*, 12520-4.

Morgan, H. D., Dean, W., Coker, H. A., Reik, W., & Petersen-Mahrt, S. K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *The Journal of Biological Chemistry*, *279*, 52353-60.

Mu, Y., Prochnow, C., Pham, P., Chen, X., & Goodman, M. F. (2012). A Structural Basis for the Biochemical Behavior of Activation-induced Deoxycytidine Deaminase Class-switch Recombination-defective Hyper-IgM-2 Mutants\*. *The Journal of Biological Chemistry*, *287* (33), 28007-816.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RYNA editing enzyme. *Cell*, *102*, 553-63.

Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., et al. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *Journal of Biological Chemistry*, *274* (26), 18470-6.

Murphy, K., Travers, P., & Walport, M. (2008). *Janeway's Immunology*. New York, NY, USA: Garland Science.

Muschen, M., Re, D., Jungnickel, B., Diehl, V., Rajewsky, K., & Küppers, R. (2000). Somatic mutation of the CD95 gene in human B cells as side effect of the germinal center reaction. *Journal of Experimental Medicine*, *192*, 1833-39.

Neuberger, M. S., Javier, M. D., Rubert, C. L., Gareth, T. W., Yang, Z., & Rada, C. (2005). Somatic Hypermutation at A.T. pairs: Polymerase error versus dUTP incorporation. *Nature Reviews Immunology*, *5*, 171-8.

Nowak, U., Matthews, A. J., Zheng, S., & Chaudhuri, J. (2011). The splicing regulator PTBP2 interacts with cytidine deaminase. *Nature Immunology*, *12*, 160-166.

Okazaki, I., Kinoshita, K., Muramatsu, M., Yoshikawa, K., & Honjo, T. (2002). The AID enzyme induces class switch recombination in fibroblasts. *Nature*, *416*, 340-5.

Parsa, J. Y., Ramachandran, S., Zaheen, A., Nepal, R. M., Kapelnikov, A., & Belcheva, A. (2012). Negative supercoiling creates single-stranded patches of DNA that are

substrates for AID-mediated mutagenesis. *PLoS Genetics, in press*, 8 (2), 10.1371/journal.pgen.1002518.

Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P., & Muramatsu, M. (2007). AID is required for germinal center-derived lymphomagenesis. *Nature Genetics*, 40 (1), 108-12.

Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., et al. (1998). BCL-6 mutations in normal germinal center B cells: Evidence of somatic hypermutation acting outside Ig loci. *Proceeding of National Academy of Sciences U.S.A.*, 95, 11816-21.

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R. S., Küppers, R., et al. (2000). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*, 412, 341-6.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., & Ansarah-Sobrinho, C. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell*, 143, 122-33.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., et al. (2010). Activation-Induced Cytidine Deaminase Targets DNA at Sites of RNA Polymerase II Stalling by Interaction with Spt5. *Cell*, 143 (1), 122-33.

Peters, A., & Storb, U. (1996). Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity*, 4, 57-65.

Petersen-Mahrt, S. K., Harris, R. S., & Neuberger, M. S. (2002). AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*, 418, 99-104.

Pham, P., Bransteitter, R., Petruska, J., & Goodman, M. F. (2003). Processive AID catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*, 424, 103-7.

Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F., & Chen, X. S. (2007). The APOBEC-2 crystal structure and functional implication for the deaminase AID. *Nature*, 445, 447-51.

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. *Molecular Cell*, 16, 163-171.

Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., & Cairns, B. R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell*, 135, 1201-12.

- Rajewsky, K. (1996). Clonal selection and learning in the antibody system. *Nature*, *381*, 751-758.
- Ramiro, A. R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Selina, C.-K., & Muramatsu, M. (2004). AID is required for the c-Myc/IgH chromosome translocations in vivo. *Cell*, *118* (4), 431-8.
- Ranjit, S., Khair, L., Linehan, E. K., Ucher, A. J., Chakrabarti, M., Schrader, C. E., et al. (2011). AID binds cooperatively with UNG and Msh2-Msh6 to Ig switch regions dependent upon the AID C terminus. *Journal of Immunology*, *187*, 2464-75.
- Reiniger, L., Bödör, C., Bognár, Á., Balogh, Z., Csomor, J., Szepesi, Á., et al. (2006). Richter's and prolymphocytic transformation of chronic lymphocytic leukemia are associated with high mRNA expression of activation-induced cytidine deaminase and aberrant somatic hypermutation. *Leukemia*, *20*, 1089-95.
- Revy, P., Muto, T., Levy, Y., Geissman, F., Plebani, A., & Sanal, O. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell*, *102*, 565-75.
- Robbiani, D. F., Bothmer, A., Callen, E., Bernardo, R.-S.-M., Dorsett, Y., & Difilippantonio, S. (2008). AID is required for the chromosomal breaks in c-Myc that lead to c-Myc/iGH translocations. *Cell*, *135*, 1028-38.
- Rogozin, I. B., & Kolchanov, N. A. (1992). Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochimica et Biophysica Acta*, *1171*, 11-8.
- Ronai, D., Iglesias-Ussel, M. D., Fan, M., Li, Z., Martin, A., & Scharff, M. D. (2007). Detection of chromatin-associated single-stranded DNA in regions targeted for somatic hypermutation. *The Journal of Experimental Medicine*, *204*, 181-90.
- Shen, H. M., & Storb, U. (2004). Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 12997-3002.
- Shen, H. M., Peters, A., Baron, B., Zhu, X., & Storb, U. (1998). Mutation of BCL-6 mutations in normal B cells by the process of somatic hypermutation of Ig genes. *Science*, *280*, 1750-2.
- Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., & Trainer, J. A. (1996). A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature*, *382*, 87-92.
- Sohail, A., Klapacz, J., Samaranyake, M., Ullah, A., & Bhagwat, A. S. (2003). Human activation-induced cytidine deaminase causes transcription-dependent, strand biased C to U deaminations. *Nucleic Acids Research*, *31*, 2990-4.

- Stephens, O. M., Yi-Brunozzi, H. Y., & Beal, P. A. (2000). Analysis of the RNA-editing reaction of ADAR2 with structural and fluorescent analogues of the GluR-B R/G editing site. *Biochemistry*, *39*, 12243-51.
- Storb, U., Peters, A., Klotz, E., Kim, N., Shen, H. M., & Kage, K. (1998). Somatic hypermutation of immunoglobulin genes is linked to transcription. *Current Topics in Microbiology and Immunology*, *229*, 11-9.
- Ta, V.-T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., et al. (2003). AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nature Immunology*, *4*, 843-8.
- Thomas, M., White, R. L., & Davis, R. W. (1976, July). Hybridization of RNA to double-stranded DNA: Formation of R-loops. *Proceedings of the National Academy of Sciences of the United States of America*, *2294-2298*.
- Vallur, A. C., Yabuki, M., Larson, E. D., & Maizels, N. (2007). AID in antibody perfection. *Cellular and Molecular Life Sciences*, *64*, 555-65.
- Wang, C. L., Ryan, A. H., & Wabl, M. (2004). Genome-wide somatic hypermutation. *Proceedings of the National Academy of Sciences*, *101*, 7352-6.
- Wang, M., Rada, C., & Neuberger, M. S. (2010). Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *The Journal of Experimental Medicine*, *207*, 141-53.
- Wang, M., Yang, Z., Rada, C., & Neuberger, M. S. (2009). AID upmutants isolated using a high-throughput screen highlight the immunity/cancer balance limiting DNA deaminase activity. *Nature Structural & Molecular Biology*, *16*, 769-76.
- Wang, M., Yang, Z., Rada, C., & Neuberger, M. S. (2009). AID upmutants isolated using a high-throughput screen highlight the immunity/cancer balance limiting DNA deaminase activity. *Nature Structural & Molecular Biology*, *16*, 769-76.
- Warrington, R., Wade, W., Kim, H. L., & Antonetti, F. R. (2011). An introduction to immunology and immunopathology. *Allergy Asthma Clin Immunol*, *2011* (7 Supp 1), S1.
- Willis, T. G., & Dyer, M. J. (2000). The role of Immunoglobulin translocations in the pathogenesis of B cell malignancies. *Blood*, *96*, 808-22.
- Xue, K., Rada, C., & Neuberger, M. S. (2006). The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in *msh2*<sup>-/-</sup> *ung*<sup>-/-</sup> mice. *The Journal of Experimental Medicine*, *203*, 2085-94.
- Yi-Brunozzi, H. Y., Stephens, O. M., & Beal, P. A. (2001). Conformational changes that occur during an RNA-editing adenosine deamination reaction. *The Journal of Biological Chemistry*, *276*, 37827-33.

- Yoshikawa, K., Okazaki, I. M., Kinoshita, K., Muramatsu, M., & Nagaoka, H. (2002). AID enzyme-induced hypermutation in an actively transcribed gene in fibroblast. *Science*, 296, 2033-2036.
- Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., & Lieber, M. R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature Immunology*, 4, 442-51.
- Yu, K., Huang, F.-T., & Michael, R. L. (2004). DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *The Journal of Biological Chemistry*, 279, 6496-500.
- Yu, K., Roy, D., Bayramyan, M., Haworth, I. S., & Lieber, M. R. (2005). Fine-structure analysis of activation-induced deaminase accessibility to class switch region R-loops. *Molecular and Cellular Biology*, 25, 1730-6.
- Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., et al. (2004). Single-stranded specificity of the APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nature Structural & Molecular Biology*, 11, 435-42.