Fall 2022

The Impacts of Framework Mutations on Steady-State Kinetics of an Aldolase Abzyme

Charles Andrew Mettler

LOYOLA UNIVERSITY CHICAGO

THE IMPACTS OF FRAMEWORK MUTATIONS ON STEADY-STATE KINETICS OF AN ALDOLASE ABZYME

A THESIS SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

MASTER OF SCIENCE

PROGRAM IN CHEMISTRY

BY

CHARLES A. METTLER

CHICAGO, IL

AUGUST 2022

ACKNOWLEDGMENTS

I would like to thank the Loyola Chemistry Department for their support and access to their expertise, especially Drs. Joerg Zimmermann, Graham Moran, Pengfei Li, and the laboratories of Drs. Liu, Ballicora, and Devery. I would also like to thank the graduate and undergraduate members of the Zimmermann lab especially Claire Baxter, Ellie Sharpe, Alex Wenner, Tyler Galardy, Thomas Hansen, Jonathan Anglade, and Lucas Manuszak. Finally, I would like to thank my family and close friends for their relentless support during this unique era of my life.

To William H. Riffle

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION TO ANTIBODY STUCTURE AND ABZYMES

Antibody Structure and Assembly

Immunoglobulins (antibodies) are molecular recognition proteins produced by B cells that function to recognize pathogenic and other non-self molecular entities. Their diversity is conferred from a combinatorial gene assembly process (i.e., V(D)J recombination) followed by a phase of elevated single-point mutation rates during somatic hypermutation (SHM). High-end estimates of the size of the human antibody repertoire are on the order of ~10^{15} before SHM [Rees *et al.* 2020]. With a genetic and mutational landscape this vast, an antibody can be elicited against virtually any ligand presented, natural or synthetic.

Of the five major classes of antibodies (IgM, IgD, IgG, IgA, and IgE named for their heavy chain type), IgG comprises the bulk (~75%) of soluble immunoglobulin protein [Alberts *et al.* 2002]. IgGs are tetrameric proteins composed of four separate peptide chains: two identical heavy chains and two identical light chains held together by intra- and interchain disulfide bonds (heavy and light refers to the relative lengths). Within each heavy-light chain pair, there are two distinct beta sheet domains, the heavy chain constant region (*Fc*) and the antigen-binding fragment (*Fab*, Figure 1). The *Fc* domain is sequence invariant and has little contact with light-chain residues, but forms the contact interface with the *Fc* of an identical heavy-light chain pair (via two additional disulfide bonds), thereby rendering the Ig bivalent. The two identical *Fab* domains are formed by heavy-light chain pairs that are covalently bound via interchain disulfide

bonds in the peptide chain. The *Fc* region acts mainly as a structural scaffold and recognition signal: this region anchors Igs to B-cell membranes, and provides recognition motifs to various immune cells [Alberts *et al.* 2002].

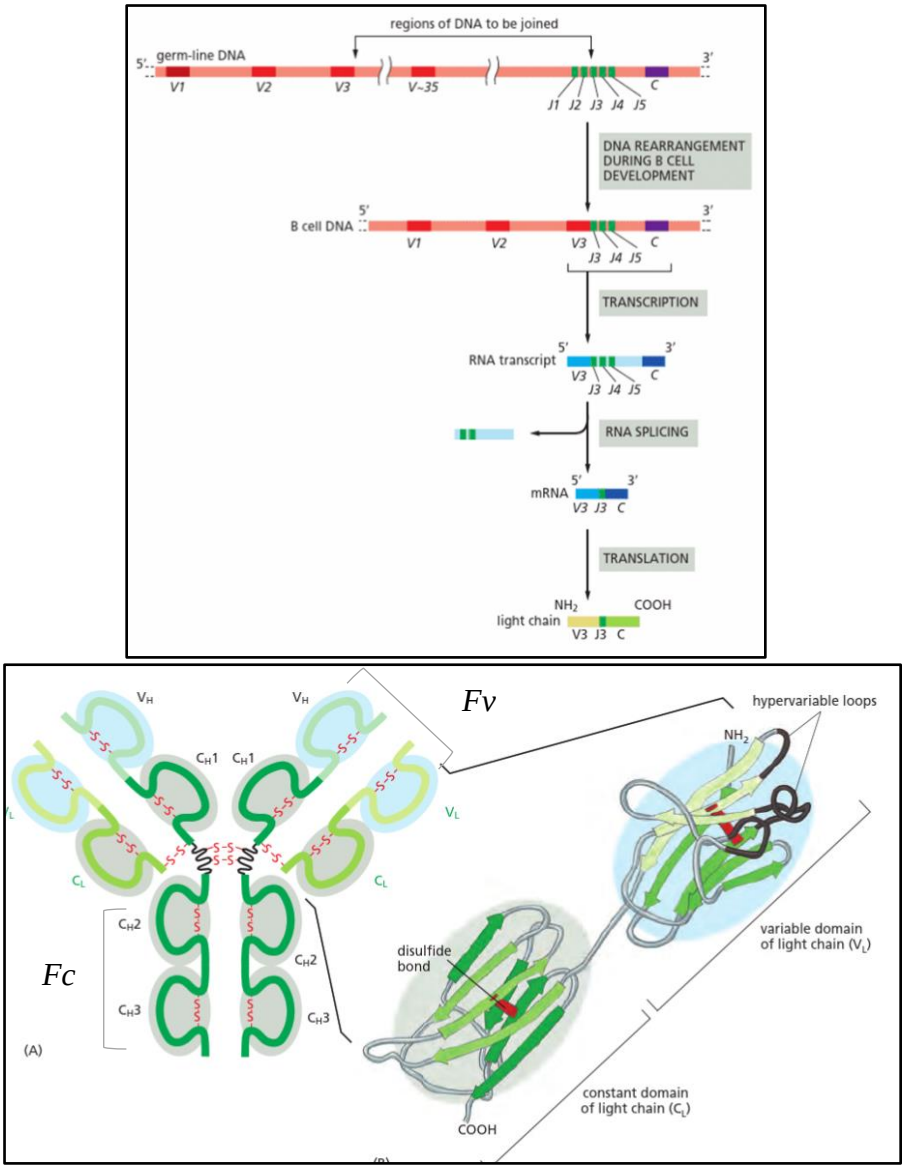


Figure 1. Antibody Genetic Recombination and Tertiary Structure. Top panel: Scheme showing the V(D)J recombination process for a light chain. *V* prefixes refer to *V* genes, likewise for *J* and *C*. Bottom panel: Bottom panel: Tertiary structure of an antibody protein showing relative positions of FWR and CDRs [from Alberts *et al.* 2002]

The *Fab* fragment is composed of two interchain beta sheet domains called the constant and variable domain. The constant domains of the heavy and the light chain (C^H and C^L, respectively) form an interchain disulfide bond near the C-terminal end of the domain. The variable region is formed by the *N*-terminal part of heavy and light chain (V^L and V^H, respectively), again in the form of beta sheet dimers. The V^L and V^H domains together are termed *Fv*, and account for the majority of the sequence diversity and all of the ligand binding capacity of an Ig protein. It is here that the binding site is formed between inter-strand loops of both V^L and V^H. The V^L and V^H domains are further divided into complementarity determining regions (CDRs, roughly corresponding to regions of ligand contact) and framework regions (FWRs, roughly corresponding to regions of no ligand contact) [Lerner *et al.* 1991]. *Fv* domain genes are generated via V(D)J recombination and thus highly diverse, but retain a relatively constant tertiary structure since the relatively short and hypervariable CDRs are largely responsible for binding ligand (Figure 1).

The assembly process of an Ig gene is essentially the same across all classes. The genes encoding heavy and light chains are organized into separate genetic loci with V^H gene segments adjacent to D (for "diversity") and J^H (for "heavy joining") segments, and V^L gene segments adjacent to corresponding J^L segments. V^L and V^H gene loci are distinguished by conjugation to a C (for "constant") gene segment. The $(VJ)^L$ gene locus is proximal to κ or λ type constant regions, while the $(VDJ)^H$ locus is adjacent to all C^H genes corresponding to each Ig class (IgM, IgD, IgG, IgA, or IgE). The Ig class is thus determined at the $(VDJ)^H$ locus. As the name suggests, V(D)J recombination fuses a single V^H gene to a D^H and J^H gene segment to generate V^H, and a single V^L gene to a J^L gene segment to generate V^L, giving rise to the *Fv* domain in the

final protein. The combinatorial ligation of V(D)J genes provides the primary Ig repertoire which, based upon the numbers of V, D, and J genes in the human genome, can theoretically generate ~10^6 unique *Fv* domains [Alberts *et al.* 2002].

Beyond V(D)J recombination, two additional mechanisms are responsible for the effectively limitless sequence diversity and consequent binding specificity of Ig proteins: junctional diversification and somatic hypermutation. The ligation of V to (D) to J gene segments often incurs a loss or gain of nucleotides at the junctional regions; these sections of a heavy or light chain mRNA script are consequently among the most highly variable in sequence [Alberts *et al.* 2002]. V(D)J recombination and concurrent junctional diversification provides the genetic raw material for subsequent mutations accrued during SHM [Alberts *et al.* 2002].

Once a ligand binds to an Ig produced from an unaltered V(D)J gene product, SHM is induced in the host B cell. Typically, 10 to 15 point mutations have been introduced in the V genes by the end of the SHM process. Due to high selection pressure during an immune response, nearly all amino acid substitutions resulting from SHM are thought to be adaptive, i.e. they provide *some* enhancement or serve some functional role in binding affinity or specificity [Sela-Culang *et al.* 2013; Sheng *et al.* 2017] . The nature of an affinity enhancement derived from CDR and FWR amino acid substitutions can often be related to newly formed stabilizing intermolecular contacts with the ligand, geometry modulation for optimal spatial complementarity [MacCallum *et al.* 1996], and, as is becoming increasingly clear, nonlocal dynamical effects associated with conformational changes [Zimmermann *et al.* 2006; Rini *et al.* 1992; Sela-Culang *et al.* 2013]

Role of CDR and FWR

Designation of CDRs and FWRs has chiefly been made on sequence positional and structural grounds [Wu *et al.* 1970]. CDRs are typically located in the inter-strand loops of the *Fv* beta sheets, permitting significant sequence and thus structural variability, while FWR regions comprise the non-CDR residues of the *Fv*. Intuitively enough, CDRs are the most variable regions of an Ig protein because CDR residues are forming intermolecular contacts with the antigen ("contact residues"—operationally defined by Padlan *et al.* [1995] as residues containing atoms that are separated from an atom of the antigen by less than the sum of their van der Waals radii plus 0.5 Å [Gelin *et al.* 1979]). Contact residues are hypervariable and are found preferentially in the third CDR in both heavy and light chains [Padlan *et al.* 1995]. CDR3 of both light and heavy chain contain the V(D)J junction and thus are directly affected by junctional diversification during V(D)J recombination. The hypervariability of CDRs is easily rationalized as the underlying functional basis for Ig selection during SHM; their amino acid composition and consequent structure are therefore said to be ligand-driven. On average however, only 20-30% of CDR residues are contact residues [Padlan *et al.* 1994]. This and the increasing availability of X-ray crystallographic data has led to attempts to refine CDRs into contact and noncontact regions based upon explicit identification of contact residues and subsequent alignments centered on structural motifs surrounding the ligand contact area [Ofran *et al.* 2008]. Such analyses have revealed that only four amino acids (tyrosine, tryptophan, serine, and asparagine) account for over half (~55%) of all residues present in CDRs (interestingly, tyrosine and tryptophan are present in every known CDR] [Ofran *et al.* 2008]). Apparently, selection pressures shape exquisitely specific CDRs but will generate them from a relatively small amino acid repertoire.

Summarily then, CDRs are more accurately described as being composed of two subregions: contact and noncontact. While the role of CDR contact residues is obvious, the functional interpretation of noncontact residues is not as straightforward. While they certainly contribute indirectly to shape-complementarity by allowing contact residue to assume the proper geometry for antigen binding, perhaps they are also selected for increasing backbone stability and conformational rigidity [Sela-Culang *et al.* 2013].

Compared to CDRs, the functional significance of FWR residues has received very little attention. The long-standing paradigm has been that FWR residues and mutations therein are responsible for maintaining proper folding and for affording conformational stability to the CDR loops [Stevenson *et al.* 2000; Zimmermann *et al.* 2006; Li *et al.* 2015] Work directed towards antibody-based drug design has identified key FW positions for retaining ligand binding upon CDR grafting to human *Fc* regions (i.e., chimeras from "antibody humanization" [Foote *et al.* 1992; Haji-Ghassemi *et al.* 2018]). Residues near the CDR-FWR interface (either in space or sequence) appear to have the strongest impact upon binding and CDR conformation [Foote *et al.* 1992; Sela-Culang *et al.* 2013; Sheng *et al.* 2017] especially in anomalously short CDR loops [Foote *et al.* 1992; Tramontano *et al.* 1990]. While the results of such studies thus reveal FW residues apparently essential for retaining or recovering binding affinity in specific instances [Tramontano *et al.* 1990], a general mechanism underlying the link between FWR residues and binding affinity and specificity remains unclear.

FWR and CDR Conformational Heterogeneity

It has been suggested that antibody evolution is partially a rigidification process whereby key residues in the binding site are locked into optimal configurations for ligand interaction;

especially in the complementary determining regions (CDRs) [Stanfield *et al.* 1993]. Therefore, Ig proteins have been conventionally defined as lock-and-key binders in their mature state. Purely V(D)J-recombined antibodies in the early stages of somatic hypermutation (SHM) are thought to be more conformationally heterogeneous and exhibit low binding specificity or affinity [Zimmermann *et al.* 2006] While this description is likely true on average, structural evidence for dynamic conformational changes upon ligand binding are abundant [Stanfield *et al.* 1993; Kodandapani *et al.* 1998; Rini *et al.* 1992; Guddat *et al.* 1994; Lin *et al.* 2013]. For example, while framework region (FWR) residues near the CDR-FWR junction contribute to CDR loop conformation, a comparative structural study famously found that an equally important determinant for a residue's contribution to protein flexibility was its presence in the V^L-V^H interface, with greater V^L-V^H contact surface area constraining any pivot between heavy and light CDR loops [Stanfield *et al.* 1993]. Similarly, structural changes upon binding of their haptens have been demonstrated for abzymes, largely occurring only after SHM [Rini *et al.* 1992; Wedemayer *et al.* 1997; Lindner *et al.* 1999]

An emerging view of structural evolution in antibodies holds that a balanced redistribution of conformational flexibility from binding site residues to distal sections of the peptide backbone is more common [Li *et al.* 2015]. The mechanics underlying the reconfiguration of hydrogen bond, ion pair, π-stacking, and cation- π networks are often highly context-dependent, and the residues responsible are found at both highly localized and distal sites from where the functional-dynamical effect is realized [Li *et al.* 2015]. Despite this, a few common themes do appear in conformational studies of antibodies, especially concerning the reorientation of H-bond networks and π- π interactions [Dalkas *et al.* 2014]. A major consensus

finding of both crystallographic and molecular dynamic studies is that the heavy chain's third and most hypervariable CDR rigidifies during SHM [Foote *et al.* 1992] while light chain CDRs tend to become more flexible [Li *et al.* 2015]. In this way, a form of entropy-enthalpy tradeoff may be at work between CDR and FWRs: entropic losses incurred by rigidification occurring in distal FWR and CDR loops are compensated for with commensurate improvements in ligand association energies [Li *et al.* 2015; Wang *et al.* 2013].

Catalytic Antibodies – Development and Scope

Antibodies can be evolved to provide finely tuned electrostatic and geometric complementarity to practically any molecule. The catalytic potential of such specificity and affinity was first realized in 1986 by Pollack *et al.* By eliciting monoclonal antibodies to a phosphonate analog of a carbonate hydrolysis transition state structure (TS), Pollack *et al.* generated catalytic antibodies (termed "abzymes") capable of hydrolyzing a broad range of carbonate esters. Conceptually, the rate-enhancement in abzymes mimics enzymes: reduction of activation barriers via TS stabilization. In this way, they effectively provide a macromolecular scaffold to catalytic residues in an optimal configuration to bind a TS analog. Since their initial development, highly stereospecific abzymes with rate enhancements on the order of 10^4 have been produced employing a wide array of TS analogs corresponding to simple unimolecular, bimolecular [Lerner *et al.* 1991; Stevenson *et al.* 2000], and even several cofactor-dependent reactions including metal centers [Ishikawa *et al.* 2013; Nicholas *et al.* 2002].

Due to the way they are generated, abzyme rate enhancement usually stems from stabilizing a *single* TSS along a reaction coordinate. However, many abzymes have also displayed enzymatic mechanisms necessary for and complementary to TS stabilization. Evidence

for geometric strain [Yin *et al.* 2003], acid-base catalysis [Debler *et al.* 2009], entropic trapping [Wagner *et al.* 1995], and proximity and reactive conformation effects [Wagner *et al.* 1995] have all been found in abzymes.

Yet, while abzymes are among the most proficient artificial protein catalysts currently available, providing significant rate enhancement over a background rate or compared to that afforded by small organic molecules, they are far below the catalytic power of enzymes. Abzymes also commonly suffer from product inhibition, further impeding their efficiency. Therefore, and despite their relative efficiency, high stereospecificity, and ease of manipulation, abzymes have limited actual uses beyond niche reactions [Kuah *et al.* 2016]. Undoubtedly, TS stabilization is necessary but not sufficient to achieve typical enzymatic rates, and the factors accounting for this disparity are a topic of considerable theoretic and practical interest [Lerner *et al.* 1991]. An obvious reason for this disparity are the contrasting selection pressures during the evolution of enzymes and abzymes; enzymes are selected for *reactivity*, while abzyme are selected for *binding affinity* to a designed structural and/or electronic analog to a desired reaction (a strategy sometimes referred to as "reactive immunization" [Lerner *et al.* 1991] or "immune-directed" evolution). Abzymes are in fact improvable catalysts: both rational, site-directed mutagenic [Tanaka *et al.* 2000] and directed evolution approaches [Benkovic *et al.* 1992] have been moderately successful in improving catalytic geometries and turnover efficiencies. Better TSS analog design has also predictably improved many abzymes [Benkovic *et al.* 1992].

Another possible source of disparity may lie in the extent of protein dynamics in abzymes and enzymes. It is well-established that antibodies undergo significant backbone rigidification over the course of somatic hypermutation [Zimmermann *et al.* 2006; Li *et al.* 2015]. The

resulting mature antibody binds substrate via a lock-and-key mechanism, i.e., its binding site is conformationally constrained, capable of assuming only a small number of low-energy conformations that are mostly confined to minor rearrangements of the CDR loops [Foote *et al.* 1992; Haji-Ghassemi *et al.* 2018; van den Elsen *et al.* 1999]. Conversely, enzymes are thought to bind via an induced fit, induced switch, or conformational selection mechanism whereby conformational rearrangements are transmitted from the active site [Vogt *et al.* 2014] It is these directed conformational shifts that allow enzymes to desolvate and respond dynamically to substrate binding, subsequently stabilizing *multiple* TS along a reaction coordinate, and releasing product. Lock-and-key binding in antibodies may function well for a single step reaction, however these are rare. It seems likely that the accessibility of multiple, directed conformations accounts for at least some of the enzyme-abzyme rate gap. Due to their inherently constrained dynamics, abzymes may in contrast provide optimal stabilization for mainly a single, ideally rate-limiting step. As a result, any TSS along the reaction path deviating significantly from the epitope may either be destabilized or only partially stabilized, which partially lowers the net catalytic potential of an abzyme.

Framework Mutations and Catalysis— Hypotheses and Rationale

Antibody framework regions can be functionally characterized as structural supports to the binding site, holding contact residues in optimal geometric or electronic complementarity to that of the ligand [Sela-Culang *et al.* 2013]. During affinity maturation, the range of thermodynamically accessible conformations to an antibody has been shown to decrease [Zimmermann *et al.* 2006; Li *et al.* 2015] and mutations occurring in framework regions may play a central role in this rigidification. Because abzymes are the result of affinity maturation and

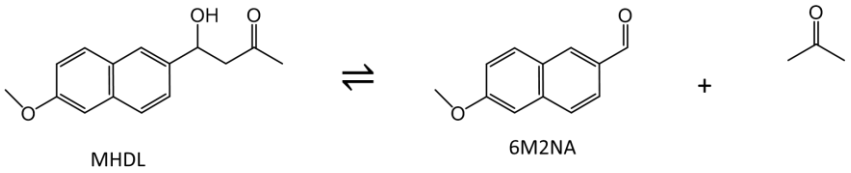
thus likely acquired rigidifying mutations, we hypothesize that mutating mature framework residues to their germline precursor may restore some degree of conformational freedom and perhaps improve catalysis. In other words, by undoing rigidifying FWR mutations while maintaining catalysis-enabling CDR mutations, the abzyme may become less of a lock-and-key binder and more of an induced fit/conformational selection binder, thus more closely resembling an enzyme. Because multiple accessible and induced conformations provide a kinetic advantage, framework mutants of abzymes may be more effective catalysts. We will test this hypothesis by characterizing abzyme framework mutants derived from the abzymes somatic hypermutation profile as well as commonly occurring FWR mutations which we have identified through comparison of mature and germline sequences of a large number of published antibody sequences (see Chapter 3).

Aldolase Abzymes and the 93F3 Model System

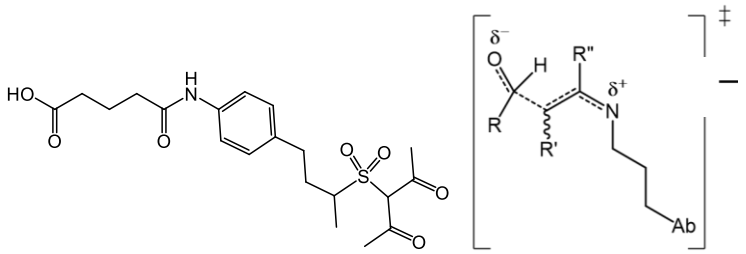
Aldol reactions are commonly employed in C-C bond-forming steps, and catalysts yielding enantiopure products are especially valued. Enantioselective abzymes catalyzing aldol and *retro*-aldol reactions were first developed in 1995 by Wagner *et al.* by raising antibodies to aromatic β-diketones in an effort create a deprotonated, catalytic enamine-forming lysine by trapping an antibody lysine's N_{ε} in a stable, cyclic enaminone [Wagner *et al.* 1995]. This approach is inspired by Class I aldolase enzymes, where enamine formation is the defining feature and initial phase of the aldol reaction cycle [Zhu *et al.* 2004]. While immunizing to β-diketones does generate abzymes capable of cleaving aldol bonds, enaminone formation as a selection criterion yields binding sites mainly proficient in stabilizing the initial proton-shuffling steps involved in carbinolamine formation and ketone dehydration [Zhu *et al.* 2004]. Ena*minone*

formation also models the amine attack and proton transfers leading to en*amine* formation, the species capable of forming the aldol C-C bond. The C-C bond forming step proceeds through a tetrahedral carbon-centered oxyanion, which was mimicked in later TS analog designs containing tetrahedral sulfoxides adjacent to the β-diketone [Zhu *et al.* 2004]. However, aldolase abzymes are generally not efficient with aldol *formation* and tend to be biased towards C-C cleavage [Zhong *et al.* 1999] largely due to inhibition by a highly electrophilic aldehyde product [Zeymer *et al.* 2017].

Antibody 93F3 is among the best characterized aldolase abzymes to date, with crystal structures and detailed binding site analyses available [Zhu *et al.* 2004]. Its optimal substrates are benzyl- and naphthyl-aldols which yield fluorescent aldehydes upon aldol cleavage, providing an easily trackable signal for monitoring reaction progress. 93F3 is highly enantiospecific with enantiomeric excess of more than 99% [Zhong *et al.* 1999]. 93F3 catalyzes the cleavage of methodol (MHDL) into 6-methoxy-2-napthyldehyde (6M2NA Figure 2) and acetone with a approximately 0.6 min^{-1} turnover rate [Zhu *et al.* 2004]. Because 93F3 is a structurally well-characterized abzyme catalyzing a simple, spectroscopically tractable reaction, we chose it as a model system to explore the influence of framework mutations, and by extension conformational heterogeneity, upon catalytic performance.



(**a**)



(**b**)

Figure 2. Study Reaction. Scheme of the model reaction (a) and corresponding hapten with a generalized aldol TSS (b**)**

CHAPTER 2

CONSERVED FRAMEWORK MUTATIONS

Background and Rationale

The majority of work to date regarding the mutability of antibody framework regions and their roles in structural evolution has been focused towards identifying key positions, residues, and relative orientations that modulate binding affinity in specific antibodies [Sheng *et al.* 2022; Koenig *et al.* 2017]. While this paradigm effectively describes structurally relevant mutations, the connection between high-frequency individual mutations and conformational flexibility remains difficult to interpret. Indeed, the impact of a CDR-distal framework mutation is far from purely deterministic as the selective context for nearly all somatic mutations lies in state of the antibody *at the time of selection*. Conventional supposition proposes that CDR mutations are selected first with nearly all substitutions conferring a direct affinity advantage, followed by overall rigidification via framework mutations [Allen *et al.* 1987]. To identify patterns between specific framework mutations and conformational flexibility (contrasted with the emergence of specific structural features), perhaps the most informative and comprehensive starting point lies in V gene specific substitution profiles (GSSPs) developed by Sheng *et al.* [Sheng *et al.* 2017]. A GSSP shows the most frequent amino acid substitutions along a given V gene, giving a straightforward reference for an antibody of interest. Common substitutions in a given V gene are largely due to inherent codon biases in the replication machinery and their preservation would suggest some sort of functional advantage [Sheng *et al.* 2017]. Interestingly, even a

particular serine codon can display different substitution preferences between different V genes further emphasizing the gene-dependence of framework mutations [Sheng *et al.* 2017]. A possible interpretation of this is that common framework mutations *are* in fact independent of ligand insofar as a V gene is a germline precursor for various "ligand types" (e.g., peptide, small organic, polysaccharide, etc.) on an equal basis. Common mutations could then represent a step towards convergent structural evolution unique to the V gene. These mutations seem to be the most likely to narrow the conformational landscape, since the majority of observed framework mutations are unique to the particular antibody (i.e., likely serving a ligand-specific purpose). Furthermore, the generalizability of the database approach to parsing out conformationally significant framework mutations is improved by the fact that V gene repertoires are also highly similar in sequence [Ehrenmann *et al.* 2010].

Assignment of Germline V Genes

To determine the likely germline V gene for a mature sequence, a FASTA batch file containing the antibody sequences derived from Sheng *et al.* [Sheng *et al.* 2022] was submitted to IMGT's DomainGapAlign tool [Ehrenmann *et al.* 2010] for amino acid sequences and the part of the variable region corresponding to the V gene was parsed out from the IMGT output using a Python script. The search was restricted to *Mus musculus* V genes ("V domain") using default settings (i.e., Smith-Waterman scores above zero, alignment E-value equal to 200, gap penalty for query equal to -5, gap penalty for reference equal to -20). The germline gene with the highest percent sequence identity to the mature sequence was assigned to be the likely germline precursor of the mature antibody and was used to identify the likely somatic mutations that occurred during affinity maturation of the mature antibody.

Filtering of Mature Antibody Sequences

A *Mus musculus* antibody sequence list was obtained from Sheng *et al.* [Sheng *et al.* 2022]. The database contained 966 full mouse-derived antibody sequences (V^H-V^L pairs). To exclude closely related antibodies that would bias the frequency analysis for somatic mutations, sequences with more than 90% homology were concatenated. The FASTA file then contained 741 V^H and 628 V^L sequences corresponding to 628 V^H-V^L pairs. 113 V^H sequences were therefore not paired with a V^L sequence.

Determination of Frequent Somatic Mutations

Likely germline V genes for each sequence were determined using the IMGT's DomainGapAlign tool [Ehrenmann *et al.* 2010] for amino acid sequences as described above. The two FASTA files containing the V^H and V^L sequences were each submitted at once to DomainGapAlign, which outputs the result as a single website. This website was saved locally in HTML format and a Python script was used to extract the likely germline gene and a list of the corresponding somatic mutations for each mature sequence and save this information in text files. 10,550 somatic mutations were identified via this process, 6,806 in V^H sequences and 3,744 in V^L sequences. This list was then split into heavy and kappa / lambda (light) V genes and further into IMGT-defined FWRs. Figure 3 shows the frequency histograms for the number of somatic mutations occurring in V^H and V^L.

The sequences were then further filtered using a combination of Excel and Python scripts to detect and eliminate sequences with identical somatic mutation patterns (i.e. two sequences for which all somatic mutations are the same) position ± 1 and germline/mature identity of the amino acid substitution. Finally, to exclude artefacts stemming from the assignment of an incorrect germline gene, in which case a large number of apparent mutations may be expected,

V^L sequences with more than 11 and V^H sequences with more than 18 somatic mutations were eliminated. The final data set included 520 V^L (containing 2,636 somatic mutations) and 641 V^H sequences (containing 5,395 somatic mutations).

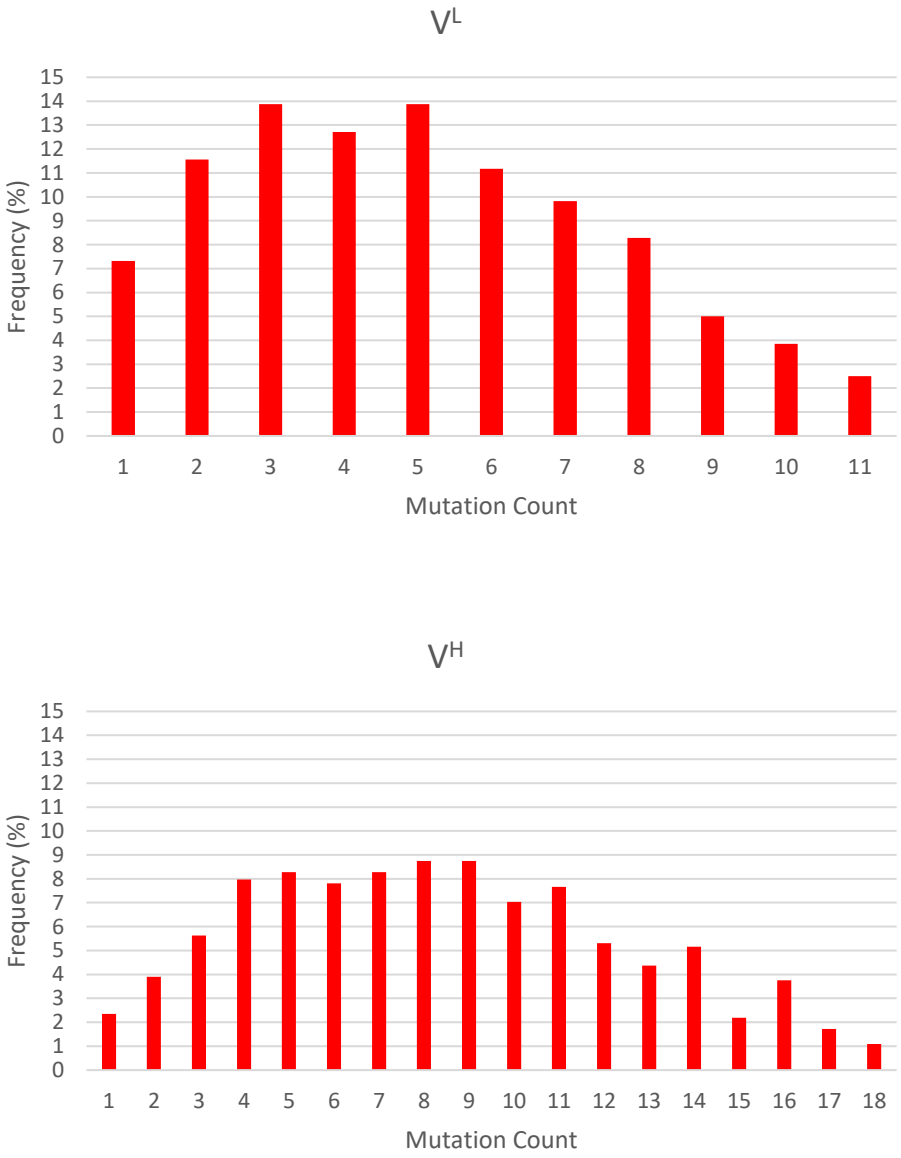


Figure 3. Histograms for number of somatic mutations in the V region of light (top) and heavy chain (bottom).

Somatic mutations in 93F3

We first confirmed the assignment of somatic mutations in 93F3 made by Wang *et al.* [2013] using IMGT. In all, 93F3 accrued 12 V^H and 9 V^L somatic mutations. Figure 4 shows the IMGT output for 93F3 with annotated CDRs and FWRs.

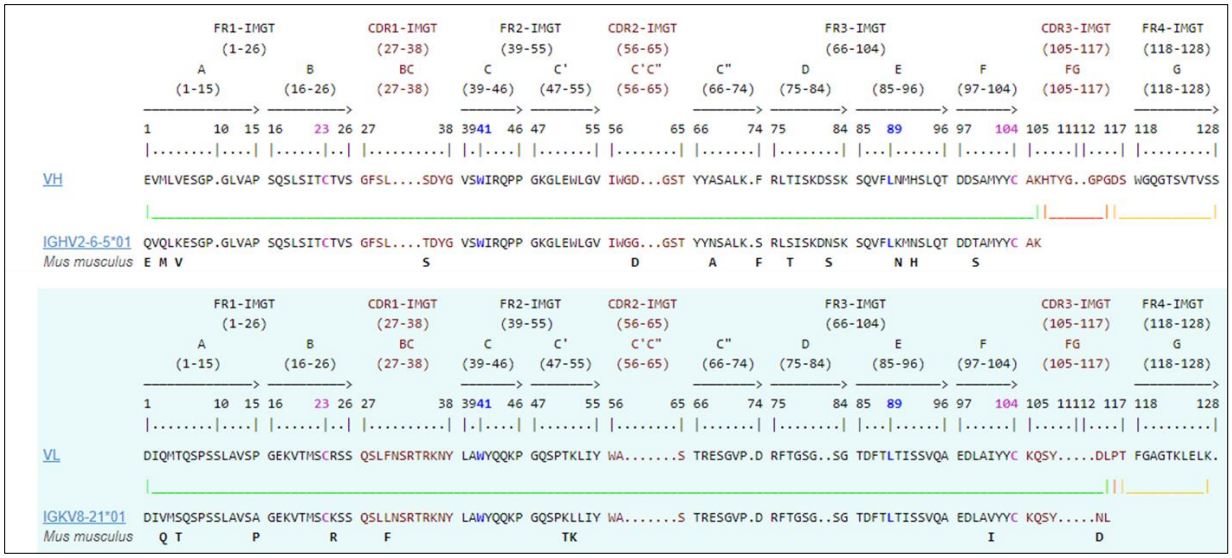


Figure 4. IMGT V gene assignment and somatic mutations for 93F3. The V gene sequence (IGHV2-6-5*01 for V^H and IGKV8-21*01 for V^L) is listed with substitutions in 93F3 represented as bolded amino acid codes below.

High-frequency somatic mutation results

High frequency mutations are summarized in Table 1 and Figure 4 show the most frequent mutations and most frequently mutated *positions* along an antibody sequence, respectively. Overall, mutations appearing in more than ~2% of the sequences were found to be switches between residues of similar character, and germline (GL) to mature changes were found to not generally follow any directionality in chemical structure or properties. Therefore, we decided to include position matches of the mature 93F3 sequence with either the germline or mature residue of a high-frequency mutation to identify candidates for single-point mutations in the 93F3 sequence. Our rationale was that if high-frequency mutations increase rigidity of the

binding site, reversing them will decrease it, and either way the effect on catalytic power can be captured.

Table 1. Somatic mutations occurring with frequencies larger than 2% within their respective FWR (exception: V^H FWR3).

	V_H		V_L	
	mutation	frequency [%]	mutation	frequency [%]
FWR1	Q^H6E	4.81	M^L4L	5.23
	Q^H3K	3.81	T^L7S	4.05
	Q^H1E	3.01	V^L2I	3.04
	K^H20R	2.41	D^L1E	2.70
	L^H12V	2.31	S^L7T	2.02
	L^H21I	2.21	L^L4M	2.02
FWR2	M^H39I	7.78	Y^L42F	4.56
	V^H42I	2.78	Y^L55F	3.49
			Y^L55H	3.49
			K^L45R	2.95
			K^L51R	2.95
			M^L39I	2.68
			L^L52V	2.41
			P^L46S	2.41
			K^L51Q	2.14
FWR3	G^H74D	1.98	V^L101I	3.73
	S^H85N	1.82	S^L92N	3.11
			Y^L103F	2.02

Table 2. Top doubly and triply co-occurring mutations in each chain. Frequency of individual mutation (w.r.t. total number of sequences) is shown in parenthesis with their product (P).

V_H				V_L			
mutations(f [%])		P [%]	frequency [%]	mutations(f [%])		P [%]	frequency [%]
L^H21I(3.4)	I^H101V(3.7)	0.13	2.81	V^L2L(3.5)	D^L1E(3.1)	0.11	1.35
T^H96S(3.1)	I^H101V	0.12	2.81	N^L114H(1.3)	A^L74V(1.7)	0.02	1.15
T^H96S	L^H21I	0.11	2.81	V^L2I	T^L7S(4.7)	0.16	0.96
I^H101V	T^H35S(5.9)	0.22	2.65				
T^H96S	T^H35S	0.18	2.50				
L^H21I	T^H35S	0.20	2.34				
L^H21I	I^H101V	T^H96S	2.81	A^L57V	K^L51R	S^L107I	0.77
I^H101V	T^H35S	L^H21I	2.34	T^L7S	M^L4L	V^L2I	0.58
I^H101V	T^H35S	T^H96S	2.34	S^L28T	V^L101I	V^L114F	0.58
L^H21I	T^H35S	T^H96S	2.34				
V^H80A	G^H74D	R^H47G	1.56				
K^H72R	M^H39I	V^H80A	1.25				

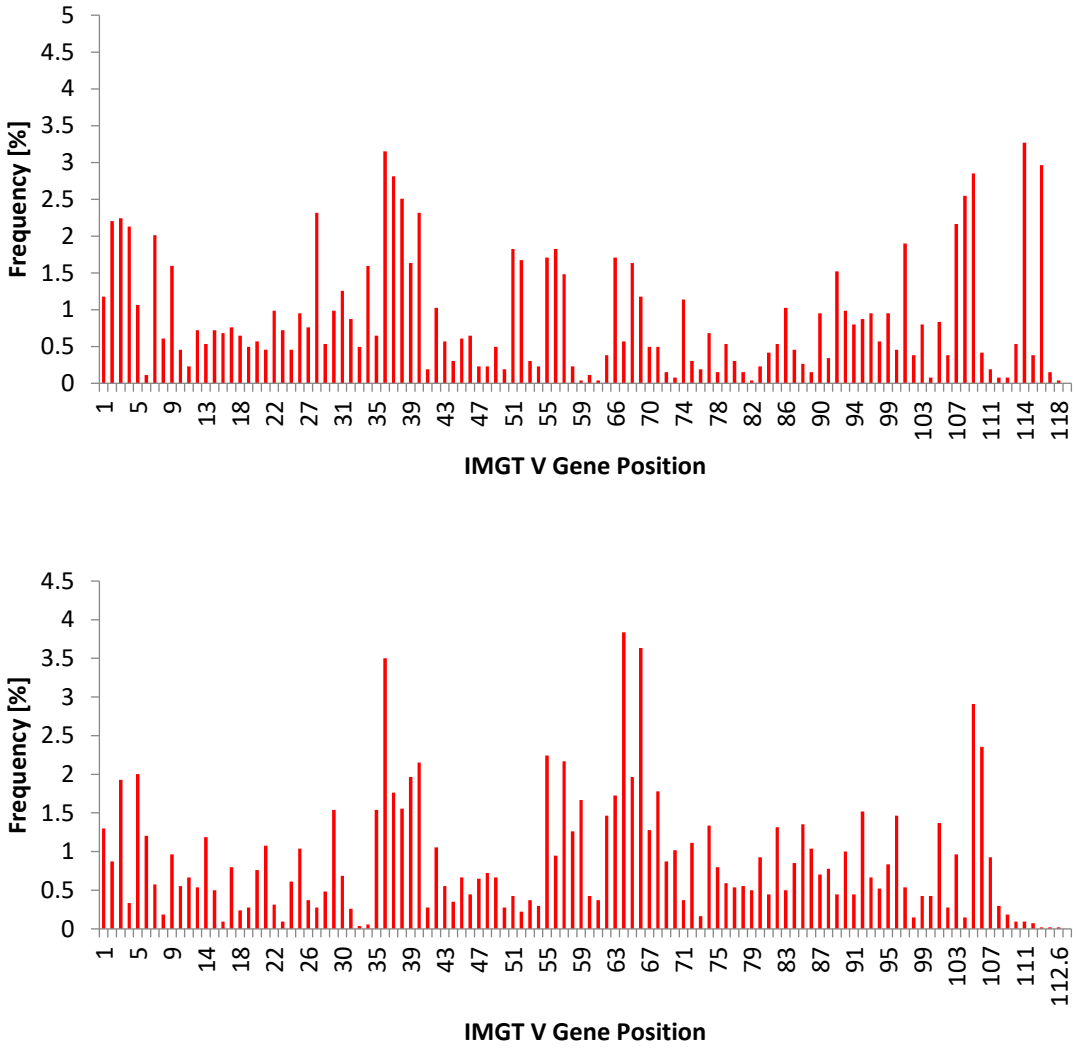


Figure 5. Frequency of mutations as a function of IMGT residue index for each chain (top panel, V_H; bottom panel V_L)

Significance of Conserved Framework Mutations as a Reference Database

The most frequently observed mutations in our nonredundant sequence database are shown in Table 1. The recurrence of some mutations on the order of 5% frequency within its framework region likely represents an interplay between convergent functional roles and the inherent mismatch and mutagenic bias of the DNA replication machinery [Sheng *et al.* 2017, 2022] In line with the latter possibility, V gene identity is known to be a significant predictor for

both position and amino acid substitution for framework somatic mutations [Sheng *et al.* 2017]. Therefore, frequent, "conserved" mutations will arise in framework regions *independent* of antigen identity due to functional or nucleotide motif biases [Sheng *et al.* 2017]. Of course, these are not mutually exclusive origins for common framework mutations.

Furthermore, our database results show basically no pattern or preference in amino acid exchange with respect to specific physiochemical character a consistent finding in the literature [Sheng *et al.* 2017; Koenig *et al.* 2017]). As an example, there is no clear preference for hydrophobic to charged residue mutations. If anything, the most common mutations trend towards switches between residues of similar character, with the top V^H switches being Q>E, M>I, and G>D and V^L with M>L, Y>F, and V>I. This suggests some degree of chemical preservation that complicates a simple physiochemical switch interpretation. Position, however, appears to be a reliable determinant of mutability (Figure 5) which may be expected because of the conserved location of secondary-structure elements and common interdomain-modulating residues such as the elbow angle between the V and C domains [Sheng *et al.* 2022; Koenig *et al.* 2017]. Functional significance is therefore unlikely to be easily inferred from inspection of substitution type alone, but is instead predictable largely on the basis of V gene origin and position in the antibody primary through tertiary structure. Indeed, our double- and triple-correlation analysis (Table 2) revealed that while there are a surprising number of these correlated mutations, all arose from the same V gene and the corresponding antibodies often bound similar antigens. While confidently assigning functional significance to common framework mutations remains a challenge, the high degree of similarity between V genes and the

FWRs giving rise to correspondingly higher antibody structural similarity facilitates comparison [Sela-Culang *et al.* 2013].

The double and triple-correlated mutations show a high degree of overlap in that the top three double mutations, occurring at a frequency of 2.81%, compose the top triple mutations given its identical frequency to the top three doubles. This indicates these three mutations (L^H21I, I^H101V, and T^H96S) always occur together and not as a part of an independent double mutation. This overlap is likely to occur throughout our correlation analysis, which significantly complicates identifying *purely* double, triple, or quadruple sets of correlated mutations. However, in some instances a double mutation appears at a slightly higher frequency than any combination of triple mutation containing those two doubles. An example is I^H101V and T^H35S that are contained in the 2nd and 3rd most frequent triple mutations in Table 2. It would seem the total occurrence of the double mutation could be obtained by summing the frequencies of the two triples. This does not occur in this instance because the 2nd and 3rd most frequent triple mutations together form a quadruple correlated mutation. Therefore, to obtain the fraction of the I^H101V-T^H35S frequency due to purely doubles, the frequency of the quadruples can be subtracted from that of the doubles. This procedure can be iterated out to higher order levels of correlation (quintuple, sextuple, etc.) to determine the fraction of lower order correlations that do occur as a part of a higher order correlation. Because this analysis and procedure is beyond the scope of the present work, we have listed the results of our correlation analysis up to triple mutations including frequencies and total counts in Appendix A for future work.

Because of the high degree of V gene similarity, a matching of mature residues, both by amino acid type and sequence position, of an antibody of interest ‘lacking’ the mutation to the

mutational database should to a reasonable degree reveal similar protein microenvironments between the antibodies contributing to the database count and the antibody of interest. Insofar as they are similar, this suggests common functional roles for a frequent residue-position pair. For the antibody of interest lacking this mutation and thus preserving the germline residue, an interpretation based upon selection arguments is that this unmodified residue-position pair is either beneficial to binding or is inconsequential, thus those antibodies with the mutation are preserved. Using a rigidifying classification for framework residues, this reasoning suggests convergent antibody evolution towards the most proximal maximum in sequence-functional space [Sheng *et al.* 2022].

The veracity of a reference database, even a gene-specific one [Sela-Culang *et al.* 2013], for identifying position-residue pairs conferring enhanced binding (or some other feature) is challenged by the consistent observation (including in our database) that a substantial majority of mutations are unique [Sheng *et al.* 2017, 2022]. This highlights the sensitivity of CDR-epitope interactions to distant framework residues and thus the exquisite context-dependency of the affinity maturation process [Burnett *et al.* 2020]. Perhaps then the most efficient approach to framework engineering should begin with an analysis of frequent mutations within the relevant V gene, followed by systematic functional evaluation of the "missing" frequent mutations via simulation or calculation. An extension of this approach for selecting additional mutations may involve matching sequence and/or structural motifs surrounding the residue or position of interest to *other* V genes or structurally-related proteins, which has been in part implemented algorithmically [Baran *et al.* 2017; Adolf-Bryfogle *et al.* 2018].

Objective of Mutational Strategy: Selection of Mutations for Analysis

With the general aim to increase backbone flexibility, we employed two approaches for selecting sites for site-directed mutagenesis in the 93F3 framework region: (1) computationally assigning a germline V gene (using IMGT) and determining likely somatic mutations acquired by 93F3 (SH), and (2) create a database of frequently occurring framework mutations and comparing it to 93F3's sequence. The motivation for the latter approach is that if mutations across disparate antibodies exist, this suggests a common functional role *independent of the specific antigen.* This functional role may be related to conformational stability and/or flexibility. From these two approaches, we selected 5 SH mutations and 5 DB mutations (Table 3) predicted to reduce the number of intermolecular interactions and thus increase the flexibility of the binding site and improve the catalytic power of the 93F3 abzyme. SH mutations were selected with the goal of eliminating or disrupting intermolecular associations with neighboring residues ($T^{H}87S$), to alter packing density ($Y^{H}102S$, $L^{L}46K$, $L^{L}27cF$), or purely on the dissimilarity of the substitution ($A^{L}15P$, $L^{L}27cF$) seen in the crystal structure. The remaining somatic mutations either resided on solvent exposed surface or were CDR residues and were considered too close to catalytic groups. Exploratory mutation $S^{H}87A$ sought to eliminate the H-bond of S87 (the somatic substitution is likely to retain H-bonding) while preserving a β-carbon with an alanine. K106 is involved in 2 cation-π interactions that appear to hold β-sheet inter-strand loops stable. Leucine was chosen as the substitution for its resemblance to a lysine residue lacking the ε-N and thus the cation involved in the interaction.

Table 3. 93F3 mutations selected for analysis. Rational mutations refer to exploratory mutants with residues involved in interesting and potentially rigidifying interaction

Source	Mutation	Description / justification
germline-derived (SH)	$T^{H}87S$	involved in H-bond network near catalytic site
	$Y^{H}102S$	S102 sits in tightly packed strand-loop interface
	$L^{L}27cF$	in loop near catalytic site, but role is unclear
	$A^{L}15P$	P>A change in distal loop
	$L^{L}46K$	within 5 Å of catalytic residues, but no clear interactions
database-derived	$I^{H}42V$	common somatic mutation in VH FWR, in catalytic site
	$R^{H}43K$	common somatic mutation in VH FWR, involved in H-bond network
	$Y^{H}67F$	common somatic mutation in VH FWR, just opposite catalytic site
	$K^{L}45R$	common somatic mutation in VH FWR, near catalytic site, but role is unclear
	$Y^{L}103F$	common somatic mutation in VH FWR, mutate to Phe removes inter-chain H-bond
exploratory	$S^{H}87A$	Eliminates H-bonds of S87
	$K^{H}106L$	K106 involved in interesting cation-π interaction at strand-loop interface

CHAPTER 3

PROTEIN EXPRESSION AND QUANTIFICATON

Background and Rationale

In Chapter 2, we identified twelve point mutations to be introduced into the mature 93F3 antibody to assess how they might change the catalytic power of the abzyme (Table 3). Since the *Fc* constant region does not contribute to catalysis, abzymes are usually expressed as *Fab* fragments (Figure 1). This also has the advantage of producing monovalent proteins as opposed to bivalent IgG constructs, which greatly simplifies the binding and kinetic assays used to characterize the abzymes.

Fab fragments are commonly expressed in *E. coli*, usually by secretion of the light chain and the heavy chain fragment containing V^H and C^H domains into the bacteria's periplasmic space because bacterial periplasm more closely approximates the oxidative conditions necessary for proper antibody folding. To facilitate secretion, the N-terminal ends of the V^H and V^L genes are fused to a periplasmic localization sequence (PPLS) which encodes for a signaling peptide that directs the cytoplasmic expressed V^H and V^L chains to preprotein conduction channels. Once in the periplasm, PPLS peptides are cleaved by signal peptidases, and the light and heavy chains combine to form Fab fragments.

We chose to express 93F3 Fabs using the pBAD expression system (Thermo Fisher). Briefly, pBAD is a bacterial promoter normally controlling the expression of the arabinose operon. Upstream of pBAD, the araC regulatory protein binds to a region controlling access to

pBAD by RNA polymerase [Schleif *et al.* 2010]. While araC is bound, pBAD is inaccessible to RNA polymerase and expression is repressed. In presence of arabinose, araC dimerizes and partially dissociates from the DNA, thereby permitting access to and expression of pBAD-promoted genes [Schleif *et al.* 2010]. Thus, expression of a pBAD-promoted gene can be induced by addition of arabinose in *E. coli* strains that are araBADC- (to suppress arabinose metabolism) and araEFGH+ (to allow arabinose transport). A generalized map of the pBAD plasmid used is shown in Figure 6.

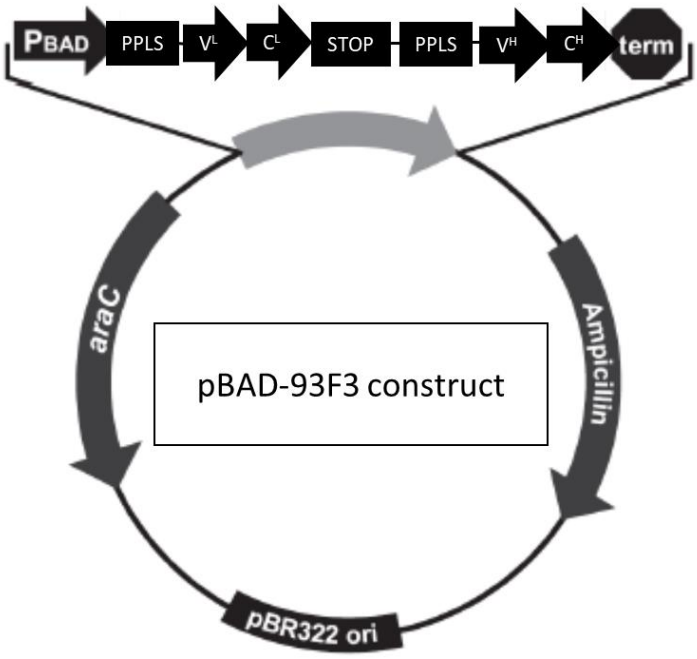


Figure 6. Map of 93F3 Fab expression plasmid. pBR322 ori is the origin of replication, Ampicillin is the amp^R gene, PPLS is the periplasmic localization sequence, and *araC* encodes the araC regulatory protein. Adapted from Thermo Fisher Publication #MAN0000049

Gene synthesis, subcloning and site-directed mutagenesis

Synthesis of the mature 93F3 V^H and V^L genes and subcloning into the pBad vector was carried out by Bio Basic Inc. The mature 93F3 plasmid was then mutated using primer mismatch techniques. Mutagenic primers were designed with Aligent's QuikChange Primer Design Tool [https://www.agilent.com/store/primerDesignProgram.jsp], synthesized by Integrated DNA Technologies, Inc., and single-nucleotide mutations were introduce into the mature 93F3 plasmid using the QuikChange mutagenesis kit (Agilent). Sequences were confirmed using cycle sequencing with fluorescent dyes as read-out terminators (Sequencing Core, University of Chicago). Due to the length of the genes, separate sequencing reactions were carried out for the V^L and V^H genes. The sequences of the mutagenic primers are shown in Table 4.

Protein expression

pBAD vectors carrying the genes of interest were transformed via heat-shock (48°C for 30 s) in presence of ampicillin into either DH10B or TOP10 cells (ThermoFisher Invitrogen DH10B catalog #EC011; TOP10 catalog #C404010) and plated on ampicillin-selection plates and incubated overnight at 37°C. Starter cultures were produced by suspending a single colony in 20 mL LB-Ampicillin and growing it to saturation. 0.9 mL of the starter culture was mixed with 80 % v./v. glycerol to a final concentration of 20 % v/v glycerol, frozen in liquid nitrogen and stored at -80°C. For subsequent expressions, stabs of the glycerol stock were used to inoculate the starter culture. The saturated starter culture was then inoculated in 2 L of 2YT-Ampicillin (100 µg / mL), shaken at 37°C, 225 rpm, and growth was monitored via OD_{600}. When OD_{600} reached 0.6, the flask was moved to a 20 – 25°C shaker (225 rpm), expression was induced by adding 0.2% arabinose, and the culture was incubated for approximately 24 hours.

Cells were pelleted at 6,000 rpm for 10 minutes and resuspended in periplasmic lysis buffer (50 mM Tris·HCl, pH 7.2, 0.53 mM EDTA). Resuspended cells were lysed via sonication for 3-5 minutes and the lysate was then filtered and purified on a protein G affinity column (GE Health or Cytiva). Eluates were buffer exchanged into PBS pH 7.4 and concentrated to a final concentration of ~10-15 μM using centrifugal filtration (EMD Millipore Amicon Ultra-15).

Table 4. Mutagenic primers for each mutant.

Source	Mutation Code	Primers
germline-derived	$T^{H}87S$	3'-tgtacgtgtcagacgtttgactactatgccggtacatgataaca-5' 5'-acatgcacagtctgcaaactgatgatacggccatgtactattgt-3'
	$Y^{H}102S$	3'-cgggccccctgataaccccagttcc-5' 5'-gcccgggggactattggggtcaagg-3'
	$L^{L}27cF$	3'-cgtctaggtcagtctcagacaatttgtcatcttgg-5' 5'-gcagatccagtcagagtctgttaaacagtagaacc-3'
	$A^{L}15P$	3'-gggaccgacacagtcgtcctctcttccag-5' 5'-ccctggctgtgtcagcaggagagaaggtc-3'
	$L^{L}46K$	3'-ttgggcctgtcagaggatgtgataactagatgacccgtaggtg-5' 5'-aacccggacagtctcctacactattgatctactgggcatccac-3'
database-derived	$I^{H}42V$	3’-tatggtgtaagctgggttcgccagcctccag-5’ 5’-ataccacattcgacccaagcggtcggaggtc-3’
	$R^{H}43K$	3’-actaagcgactatggtgtaagctggattaagcagcctccaggaaag-5’ 5’-tgattcgctgataccacattcgacctaattcgtcggaggtcctttc-3’
	$Y^{H}67F$	3’-ggggtgatggaagcacatactttgcctcagctc-5’ 5’-ccccactaccttcgtgtatgaaacggagtcgag-3’
	$K^{L}45R$	3’-ttggcttggtaccagcagagacccggacag-5’ 5’-aaccgaaccatggtcgtctctgggcctgtc-3’
	$Y^{L}103F$	3’-caggctgaggacctggcaatttatttttgtaagcaatcttat-5’ 5’-gtccgactcctggaccgttaaataaaaacattcgttagaata-3’
exploratory	$S^{H}87A$	3’-tgcacagtctgcaaactgatgatgcagccatgtacta-5’ 5’-acgtgtcagacgtttgactactacgtcggtacatgat-3’
	$K^{H}106L$	3’-ccatgtactattgtgccttacatacctacggcggcc-5’ 5’-ggtacatgataacacggaatgtatggatgccgccgg-3’

Briefly, 15 mL of eluate was reduced to ~1.5 - 2 mL via centrifugation, resuspended in 15 mL total volume of PBS pH 7.4, and again reduced to 1.5 mL. This process was repeated several times for concentration and buffer exchange into PBS.

Activity Assays / Protein Quantification

Extinction coefficients of the Fab fragments at 280 nm were calculated using the online Expansy ProtParam tool (https://web.expasy.org/protparam/), which predicted values of 72,310 and 72,810 M^{-1} cm^{-1} for completely reduced and completely free cystine residues, respectively. Using the average of the two values, protein concentrations were determined using an extinction coefficient of $\varepsilon_{280} = 0.0725$ $\mu M^{-1} cm^{-1}$.

6M2NA was purchased from TCI Chemicals. Since 6M2NA is not soluble in aqueous solutions at micromolar concentrations, stock solutions of 6M2NA in neat DMSO were prepared and diluted 20-fold in PBS pH 7.4 such that the final DMSO concentration was 5% v./v. All titrations were performed with PBS pH 7.4 / 5% DMSO as the solvent. Fluorescence spectra were collected at room temperature on a JASCO fluorometer. If not otherwise stated, photoexcitation occurred at 330 nm.

Results and Discussion

93F3 and its point mutants were expressed as chimeric Fabs in which the genes for 93F3 V^H and V^L, originating from *Mus musculus*, were fused to human C^H and C^L regions. Figure 7 shows the nucleotide and amino acid sequences of the 93F3 variable regions with Kabat and IMGT numbering. Expression yields (determined via OD_{280}) were in the range of 0.1 to 0.5 mg *Fab* protein per liter of expression culture.

(A) nucleotide sequences

>V^H
gaggtgatgctggtggagtcaggacctggcctggtggcgccctcgcagagcctatccatcacatgcactgtctcagg
gttctcactaagcgactatggtgtaagctggattcgccagcctccaggaaagggtctggagtggctgggagtaattt
ggggtgatggaagcacatactatgcctcagctctcaaattcagactgaccatcagcaaggacagctccaagagccaa
gttttcttaaacatgcacagtctgcaaactgatgattcagccatgtactattgtgccaaacatacctacggcggccc
gggggactcttggggtcaaggaacctcagtcaccgtctcctca

>V^L
gatattcagatgacacagtctccatcctccctggctgtgtcaccaggagagaaggtcactatgagctgcagatccag
tcagagtctgttcaacagtagaacccgaaagaactacttggcttggtaccagcagaaacccggacagtctcctacaa
agttgatctactgggcatccactcgggaatctggggtccctgatcgcttcacaggcagtggatctgggacagatttc
actctcaccatcagcagtgtgcaggctgaggacctggcaatttattattgtaagcaatcttatgatcttcccacgtt
cggtgctgggaccaagctggagctgaaa

(B) IMGT numbering

```
   1        10        20        30        40        50        60
VH EVMLVESGP-GLVAPSQSLSITCTVS**GFSL----SDYG**VSWIRQPPGKGLEWLGVI**WGD---GST**YYAS

   70        80        90        100       110       120
VH ALK-FRLTISKDSSKSQVFLNMHSLQTDDSAMYYC**AKHTYG—GPGDSW**GQGTSVTVSS

   1        10        20        30        40        50        60
VL DIQMTQSPSSLAVSPGEKVTMSCRSS**QSLFNSRTRKNY**LAWYQQKPGQSPTKLIY**WA-------S**TRES

   70        80        90        100       110       120
VL GVP-DRFTGSG--SGTDFTLTISSVQAEDLAIYYC**KQSY-----DLPT**FGAGTKLELK
```

(C) Kabat numbering

```
   1        10        20        30        40        50        60
VH EVMLVESGPGLVAPSQSLSITCTVSGFSLS**DYGVS**WIRQPPGKGLEWLG**VIWGDGSTYYASALKF**RLTI

                                    100
   70        80 abc       90        |a         110
VH SKDSSKSQVFLNMHSLQTDDSAMYYCAK**HTYGGPGDS**WGQGTSVTVSS

   1        10        20      abcdef  30        40        50
VL DIQMTQSPSSLAVSPGEKVTMSC**RSSQSLFNSRTRKNYLA**WYQQKPGQSPTKLIY**WASTRES**GVP

   60        70        80        90        100
VL DRFTGSGSGTDFTLTISSVQAEDLAIYYC**KQSYDLPTF**GAGTKLELK
```

Figure 7. (A) 93F3 V^H and V^L (A) nucleotide sequences in FASTA format. (B,C) Amino acid sequences in IMGT (B) and Kabat (C) numbering. CDRs according to the IMGT or Kabat numbering schemes are shown in bold font.

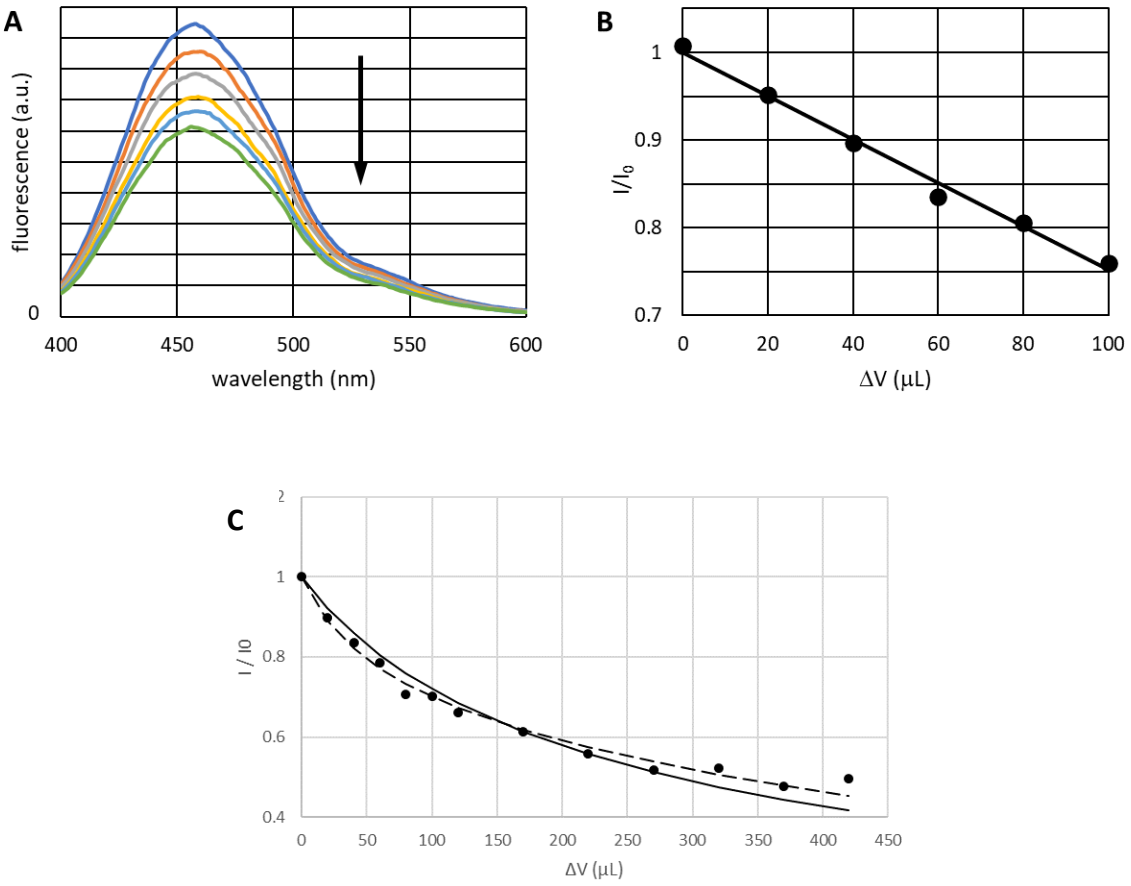


Figure 8. Change in fluorescence of 6M2NA upon addition of mature 93F3. (A) Fluorescence spectra for titration of 30 μM 6M2NA with mature 93F3 (the arrow shows the direction of change). (B) Normalized fluorescence intensity for titration of 30 μM 6M2NA with increasing volume of ~ 60 μM mature 93F3 stock (the line shows a linear interpolation of the data). (C) Normalized fluorescence intensity for titration of 1 μM 6M2NA with increasing volume of ~ 30 μM mature 93F3 stock (the dashed line shows the best fit using Equation 5 with freely varied K_D and I_∞/I_0 and solid line to quench factor fixed to 50%). The excitation wavelength for all experiments was 330 nm. For panel B and C, fluorescence intensity was integrated between 455-465 nm.

To estimate the concentration of active binding sites, the reaction of 93F3 Fab with 6M2NA was analyzed. The fluorescence of 6M2NA is quenched when it is bound to 93F3 [Zeymer *et al.* 2017]. Thus, the concentration of *active* catalytic sites was determined by titrating

30 µM 6M2NA stock solutions with protein and monitoring the decrease in 460 nm fluorescence (Figure 8A). If the concentration of protein is smaller than 6M2NA, nearly all binding sites will be occupied assuming a K_D for 6M2NA binding of ~1 µM [Zhu *et al*. 2004]. Therefore, early in a titration curve the increase in bound ligand (i.e., a decrease in 460 nm fluorescence) will be approximately linearly dependent on the volume of 93F3 stock solution added, which allows to determine the catalytic-site concentration according to

$$\frac{V_0 + \Delta V}{V_0} I = I_0 - (I_0 - I_\infty) \cdot \frac{[93\text{F3}]_0}{[6\text{M2NA}]_0} \cdot \frac{\Delta V}{V_0} \quad (1)$$

where V_0, I_0, and $[6\text{M2NA}]_0$ are the initial volume, initial fluorescence intensity, and concentration of the 6M2NA solution, respectively, ΔV is the volume of 93F3 stock solution added, I the observed fluorescence intensity after addition of ΔV, I_∞ the fluorescence intensity of a 6M2NA solution with a concentration of $[6\text{M2NA}]_0$ in which all 6MNA is bound to 93F3, and $[93\text{F3}]_0$ the concentration of active catalytic sites in the 93F3 stock solution. For improved signal-to-noise ratios, all fluorescence intensities were integrated from 455-465 nm. The slope S of dilution-corrected fluorescence intensity vs. ΔV is thus

$$S = \frac{d}{d(\Delta V)} \left(\frac{V_0 + \Delta V}{V_0} I \right) = -\frac{I_0 - I_\infty}{V_0} \cdot \frac{[93\text{F3}]_0}{[6\text{M2NA}]_0} \quad (2)$$

Rearranging for $[93\text{F3}]_0$ gives

$$[93\text{F3}]_0 = -S \cdot \frac{V_0}{I_0 - I_\infty} \cdot [6\text{M2NA}]_0 \quad (3)$$

In order to determine the active catalytic site concentrations in this manner, the quench factor (I_∞ / I_0) needs to be determined, which can be done by saturating low concentrations of 6M2NA (we chose ~1 µM) with 93F3. In addition, for the derivation of Equation 3 it was

assumed that $[6M2NA]_0 \gg K_D$, the dissociation constant of 6M2NA-93F3 complex. K_D is approximately known for mature 93F3 (~1 μM), but for the single mutants its value is likely to be different. Determination of K_D for the product the single mutants was beyond the scope of this work as we focused upon kinetics of the opposing reaction, so we made the assumption that it remains in the low micromolar range for all single mutants, but this should be verified in future work.

A related problem is the trade-off between quenching and K_D. An example for a titration with $[6M2NA]_0$ = 1 μM is shown in Figure 8C. For $\Delta V > 350\ \mu L$, apparently near 100% of 6M2NA is bound to protein since further increase of protein concentration yield negligible changes in signal. Therefore, the quench factor can be estimated from the ratio of asymptotic and initial intensity, i.e., $I_\infty/I_0 \approx 0.5$. Fitting the fluorescence intensities to a bimolecular binding curve,

$$[6M2NA]_b = \frac{1}{2}\left([6M2NA]_0 + [93F3]_0 + K_D - \ldots \ldots \sqrt{([6M2NA]_0 + [93F3]_0 + K_D)^2 - 4[6M2NA]_0[93F3]_0}\right) \quad (4)$$

$$I_{fit} = ([6M2NA]_0 - [6M2NA]_b) \cdot \left(\frac{I_0}{[6M2NA]_0}\right) + [6M2NA]_b \cdot \left(\frac{I_\infty}{I_0}\right) \cdot \left(\frac{I_0}{[6M2NA]_0}\right) \quad (5)$$

where $[6M2NA]_b$, and $[6M2NA]_0$are the bound and total concentration of 6M2NA respectively and [93F3] is 93F3 concentration. yields a K_D = 8 μM for $I_\infty/I_0 = 0.5$. Alternatively, if the quench factor is included as a fit parameter along with K_D, a least square fit of the data yields $K_D \approx 1\ \mu M$ and $I_\infty/I_0 = 0.76$. This example demonstrates that there is a possible trade-off between K_D and I_∞/I_0, changing one can be compensated for by changing the other without significantly affecting fit quality. For fits with freely varying K_D and I_∞/I_0, K_D values were

found to fall in the range of 1-8 µM, quench factors in the range of 0.25 to 0.5 for mature 93F3. Therefore, we concluded that this method of determining active catalytic site concentrations only gives reasonable results if the quench factor is fixed to 0.5, the value observed for saturation of 6M2NA with 93F3. In agreement with this conclusion, using the protein concentration determined via OD_{280} for $[93F3]_0$ in Equation 3 also gives quench factors of approximately 0.5. Since the protein concentration determined via OD_{280} is an upper limit for the active catalytic site concentration, 0.5 is a lower limit for the quench factor, confirming our approach. Table 5 shows a comparison of protein concentrations determined via OD_{280} and 93F3 titration. Where both could be determined, we found a good agreement between the two, showing that most of the protein is capable of binding 6M2NA. We thus conclude that OD_{280} is a reliable way to determine active catalytic site concentrations.

Table 5. Exemplary comparison of OD_{280} and titration-determined protein concentrations. Data assuming a 50% fluorescence quenching factor are shown.

Tube ID	**OD_{280} – determined (µM)**	**Titration-determined (µM)**
t8	18.6	17.0
tJT	63.4	58.0

CHAPTER 4

STEADY STATE KINETICS OF FRAMEWORK MUTANTS

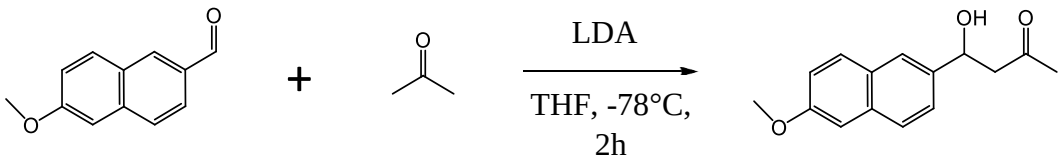
Introduction

Abzymes are a valuable comparative model for evaluating thermodynamic and kinetic factors contributing to enzymatic rate enhancement, and perhaps even to the nature of catalyst free energy landscapes generally [Bugg 2012]. This is in part due to the relatively simple potential-energy landscape of antibodies where the bound form is thought to exist as a single conformation in a narrow global energetic minimum. In contrast, enzymes are often modeled with relatively wide bound state minima (i.e., a "free energy basin" [Li *et al.* 2015]) with multiple fine-scale wells corresponding to the mechanochemical coupling between substrate and enzyme conformational transitions [Richard *et al.* 2019; Boehr *et al.* 2006].

Steady-state kinetic parameters have long been the standard for evaluating enzyme catalysis and their mimics, and their utilization comes with the advantage of facile execution, replicability, analysis, and comparison to prior work. A drawback is the lack of mechanistic information in comparison to single-turnover experiments, rendering a conformational interpretation difficult [Johnson *et al.* 2019]. Despite this, our central hypothesis that more conformationally heterogeneous abzymes should be altogether more proficient catalysts (all else equal) can be tested straightforwardly from steady-state data coupled with measures of protein flexibility. Our rationale for this hypothesis and its assessment is based upon several general findings about the

relationship of conformational heterogeneity and function in antibodies (and by extension, abzymes):

- Antibodies generally adhere to lock-and-key binding models [Vogt *et al.* 2014; Braden *et al.* 1995].
- However, there are several examples of ligand-induced conformational shifts in antibodies [Stanfield *et al.* 1993; Kodandapani *et al.* 1998; Rini *et al.* 1992; Guddat *et al.* 1994; Lin *et al.* 2013; Lindner *et al.* 1999; Kabir *et al.* 2021] though it is unknown whether this is widespread [Debler *et al.* 2008].
- Abzymes, just as enzymes, are often functionally dependent upon such conformational shifts [Lindner *et al.* 1999; Otten *et al.* 2020; Debler *et al.* 2008].
- Antibody conformational distributions (i.e., as a measure of flexibility) can be modified in a general but predictable way using the natural affinity evolution process [Zimmermann *et al.* 2006; Li *et al.* 2015], in contrast to rational design protocols employed in enzyme engineering.



6M2NA **MHDL**

Figure 9. Synthesis of MHDL and General Scheme of the Study Reaction.

Materials and Experimental Methods

6M2NA was purchased from TCI Chemicals at the highest availability purity (98%). MHDL was kindly provided by the laboratory of Professor James Devery (Loyola University Chicago). Briefly, THF and diisopropylamine (1.5 equiv) were added to a flame dried flask in a dry ice/acetone bath. Then n-BuLi (1.5 equiv) was added and allowed to stir for 30 min, after which acetone dried over K_2CO_3 was added dropwise and allowed to stir for 30 min. A solution of carbaldehyde in THF was added (Figure 9). The solution was allowed to stir for 1 hour until completed on TLC. The reaction was quenched with NH_4Cl, extracted with diethyl ether, washed with NaCl brine, and dried over Na_2SO_4. Solvent was removed under reduced pressure and product was purified via flash chromatography using EtOAc/Hexanes and confirmed via ^1H NMR. The overall reaction yield was 72%.

MHDL and 6M2NA are not soluble in aqueous solution at micromolar concentrations. Therefore, stock solutions of both were prepared in DMSO and diluted into PBS buffer, pH 7.4 to the desired concentration such that the resulting solutions always contained 5 % v/v DMSO. All fluorescence experiments were performed on a JASCO FP-80 Series Fluorometer. For the steady-state kinetic experiments, solutions of 1-150 µM MHDL in PBS pH 7.4 containing 5% DMSO were individually prepared in 500 µL cuvettes, and sufficient protein stock was added to achieve ~0.1 µM active sites. Reaction solutions were mixed thoroughly via pipetting before initiating kinetic experiment in which the fluorescence intensity of the solution at 460 nm was recorded on the same spectrometer for 90 seconds at 5 seconds time intervals with excitation at 330 nm. Each MHDL concentration was repeated at least in triplicate, with outliners discarded using Grubb's tests.

Photophysical properties of MHDL and 6M2NA

The normalized absorption and emission spectra of MHDL and 6M2NA in PBS pH 7.4/5% DMSO are shown in Figure 10A. The normalized fluorescence spectra of both are virtually identical, but fluorescence intensities vary by a factor of 64.6 for MHDL and 6M2NA solutions of equal molarity when photoexcited at 330 nm (Figure 10B). To exclude the possibility that the weak fluorescence for MHDL originates from an impurity, fluorescence excitation spectra were collected and found to closely adhere to the respective absorption spectra, demonstrating that the observed weaker fluorescence indeed stems from MHDL (Figure 10C).

While 6M2NA was found to be photostable under the experimental conditions applied, MHDL solutions exhibit a small but significant amount of bleaching of the 460 nm fluorescence intensity upon continuous irradiation with 330 nm light (Figure 11). During the first 90 s (the temporal range of the steady-state kinetic experiments) the bleaching is approximately linear, and the corresponding zeroth-order rate constant is approximately concentration-independent (Figure 11). Therefore, under these conditions, bleaching of an MHDL solution approximately follows the relationship

$$I_{obs} = I_0 \cdot (1 - b \cdot t) \tag{6}$$

where $I_{obs}(t)$ is the experimentally observed intensity reading at time *t*, and I_0 is initial, unbleached intensity. The rate constant *b* was determined to be $(1.23 \pm 0.13) \times 10^{-4}\ s^{-1}$ for the specific settings of the fluorometer used in the experiments, which were kept constant for all fluorescence and steady-state kinetics experiments.

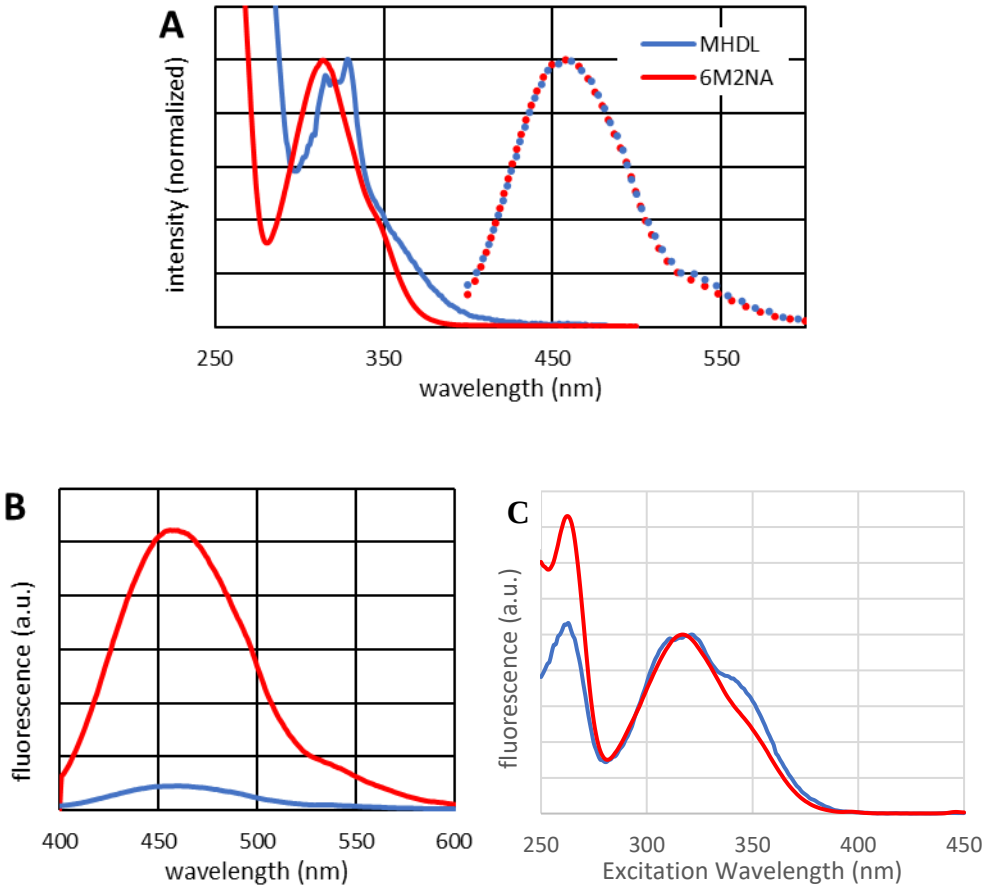


Figure 10. 6M2NA and MHDL Spectra. (A) Normalized absorption (solid lines) and fluorescence (dashed lines) spectra of MHDL (blue) and 6M2NA (red). (B) Emission spectra of MHDL (blue) and 6M2NA (red), both at a concentration of 83 μM and collected with identical instrument settings. (C) Fluorescence excitation spectra of MHDL (blue) and 6M2NA (red) for emission at 460 nm. (All solutions were in PBS pH 7.4/5% DMSO. Excitation wavelength for all fluorescence experiments was 330 nm).

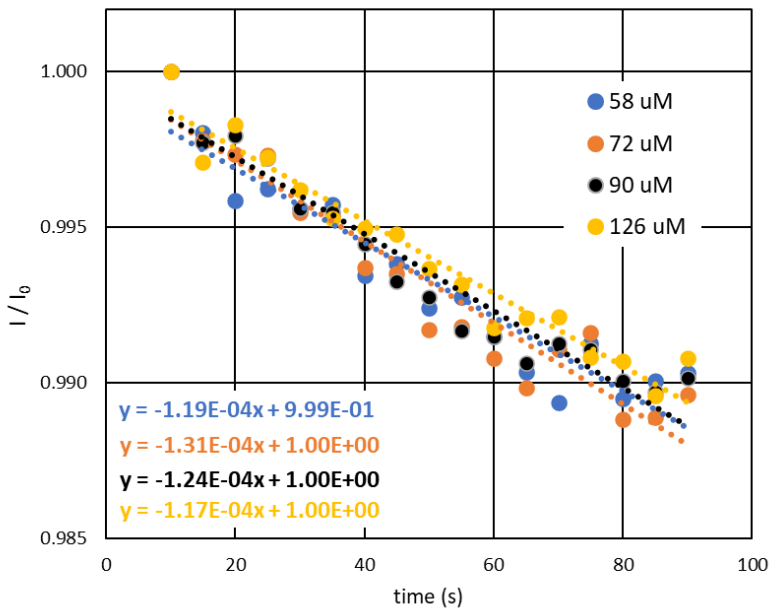


Figure 11. MHDL Photobleaching. Decay of fluorescence intensity for various concentrations of MHDL in PBS pH 7.4/5% DMSO upon continuous excitation at 330 nm (blue, 58 μM; orange, 72 μM; black, 90 μM; yellow, 126 μM). Intensities are normalized at 10 s irradiation time. Dotted lines are linear trendlines for the different concentrations. Trendline equations and trendlines are shown in the same color as the corresponding data points.

Kinetic Experiments

Because of the dramatic increase in fluorescence intensity upon transformation of MHDL into 6M2NA (Figure 10B), this relatively simple reaction provides a spectroscopically tractable probe for monitoring reaction progress and thus catalytic performance. As mentioned above, MHDL photobleaches according to Equation 6, while 6M2NA is photostable. The experimentally observed fluorescence intensity, $I_{obs}^{ss}(t)$, which is a superposition of MHDL and 6MN2A fluorescence, can be corrected for MHDL bleach via

$$I_{corr}^{ss}(t) = I_{obs}^{ss}(t) + I_0 \cdot b \cdot t \qquad (7)$$

assuming that the concentration of the MHDL substrate is approximately constant over the course of the experiment, and where $I_{corr}^{ss}(t)$ is the bleach-corrected intensity, I_0 is the initial intensity at $t = 0$, and $b = 1.23 \times 10^{-4}\ s^{-1}$. For the steady-state kinetics experiments, which are run with an excess of the MHDL substrate and initially no 6M2NA product present, accounting for the MHDL bleach according to Equation 7 resulted in small but significant corrections of the steady-state rates.

For all but one abzyme (L^L46K), an increase in fluorescence intensity was observed after the reaction was initiated by adding ~0.1 μM protein to solutions of varying concentrations of the MHDL substrate. The concentration of the 6M2NA reaction product was determined from the bleach-corrected time-dependent increase in fluorescence (Equation 7) via

$$[P]_t = \left(\frac{I_{corr}^{ss}(t)}{I_0} - 1\right) \cdot \frac{[S]_0}{64.6} \tag{8}$$

where $[S]_t$ and $[P]_t$ are the time-dependent concentrations of substrate (MHDL) and product (6M2NA), respectively, and $[S]_0$ is the initial substrate concentration. I_0 again is the initial fluorescence intensity at the start of the reaction, and it is assumed that the substrate concentration is approximately constant over the observation time.

Equations 7 and 8 thus allow to convert time-resolved fluorescence data, $I_{obs}^{ss}(t)$, to time-dependent product concentrations, $[P]_t$ (Figure 9). Steady-state reaction rates were then determined from the latter via linear regression of $[P]_t$ during the first 90 seconds after initiation of the reaction. There was no noticeable slowing of the reaction during this time, and the product concentration at 90 seconds was well below 1 % of the initial substrate concentration, justifying this approach.

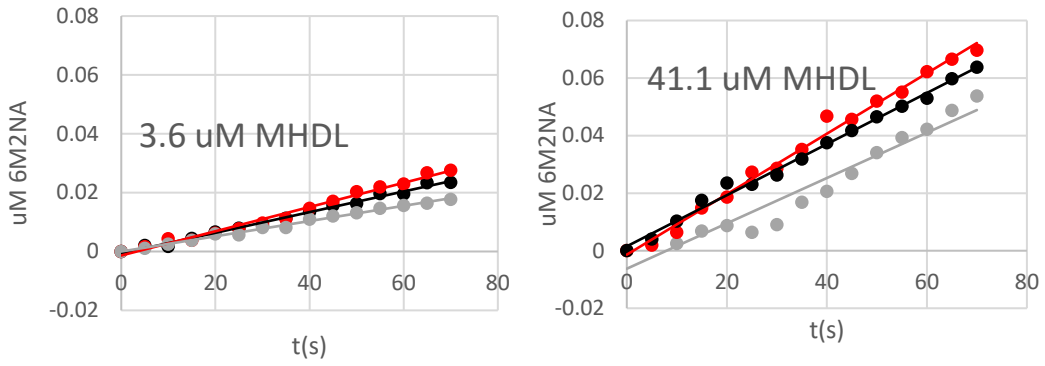


Figure 12. Reaction progress traces. Shown for a steady-state kinetics experiment at two different substrate concentrations. Linear regression (lines shown in matching colors) of these data yield the initial reaction velocities (v_0) to be fit to the Michaelis-Menten equation. Steady-state rates v_0 were determined for $[S]_0$ between 1 μM and 120 μM (Figure 9).

Most plots showed the typical behavior expected according to the Michaelis-Menten equation, with rates increasing at low $[S]_0$ before leveling off and reaching a constant maximum value v_{max}. However, for wild-type 93F3 and some mutants, a decrease in steady-state rates was observed at large $[S]_0$, which typically is attributed to substrate inhibition [Kokkonen *et al.* 2021]. Therefore, steady-state rates were fit to the Michaelis-Menten equation including substrate inhibition,

$$v_0 = \frac{v_{max} \cdot [S]_0}{K_M + [S]_0 \cdot \left(1 + \frac{[S]_0}{K_i}\right)} \tag{9}$$

where K_M is the Michaelis constant (approximately equal to the equilibrium constant of the substrate binding step), and K_i is the equilibrium constant of the substrate-inhibition step according to the following reaction scheme

$$\begin{aligned} E + S &\underset{}{\overset{K_M}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P \\ ES + S &\overset{K_i}{\rightleftharpoons} ES_2 \end{aligned} \tag{10}$$

where k_{cat} is the catalyst rate constant for product turnover, which is related to v_{max} and enzyme concentration $[E]$ via

$$k_{cat} = v_{max} \,/\, [E] \tag{11}$$

k_{cat}, K_M, and K_i values were determined for wild-type 93F3 and each of the single mutants by combined fitting of all experimentally determined steady-state rates for a given abzyme to Equation 9. Confidence intervals for all fitted parameters were estimated by evaluating the range over which summed squared residuals (SSR) increased by 5% (i.e., the range at which $SSR_{perturbed}$ / $SSR_{best\ fit}$ = 95%) [Johnson 2019]. Fit results are summarized in Figure 10 and Table 6.

Discussion

We were able to successfully follow the retro-aldol reaction of MHDL to 6M2NA and acetone via time-resolved fluorescence spectroscopy and to determine steady-state kinetic parameters of the abzyme-catalyzed reaction. We chose to observe the retro-aldol reaction (i.e., methodol cleavage) opposed to the aldol (C-C bond forming) direction due to the difficulty in saturating 93F3 (and aldolase antibodies in general) with a ketone thereby holding the catalytic lysine in the enamine form; in this case high concentrations of acetone would be required. Additional obstacles to tracking aldol synthesis are the strong overlap of 6M2NA and MHDL absorbance spectra and the high affinity of the catalytic lysine for 6M2NA, further complicating accurate saturations of the catalytic site since both production of MHDL and binding of 6M2NA

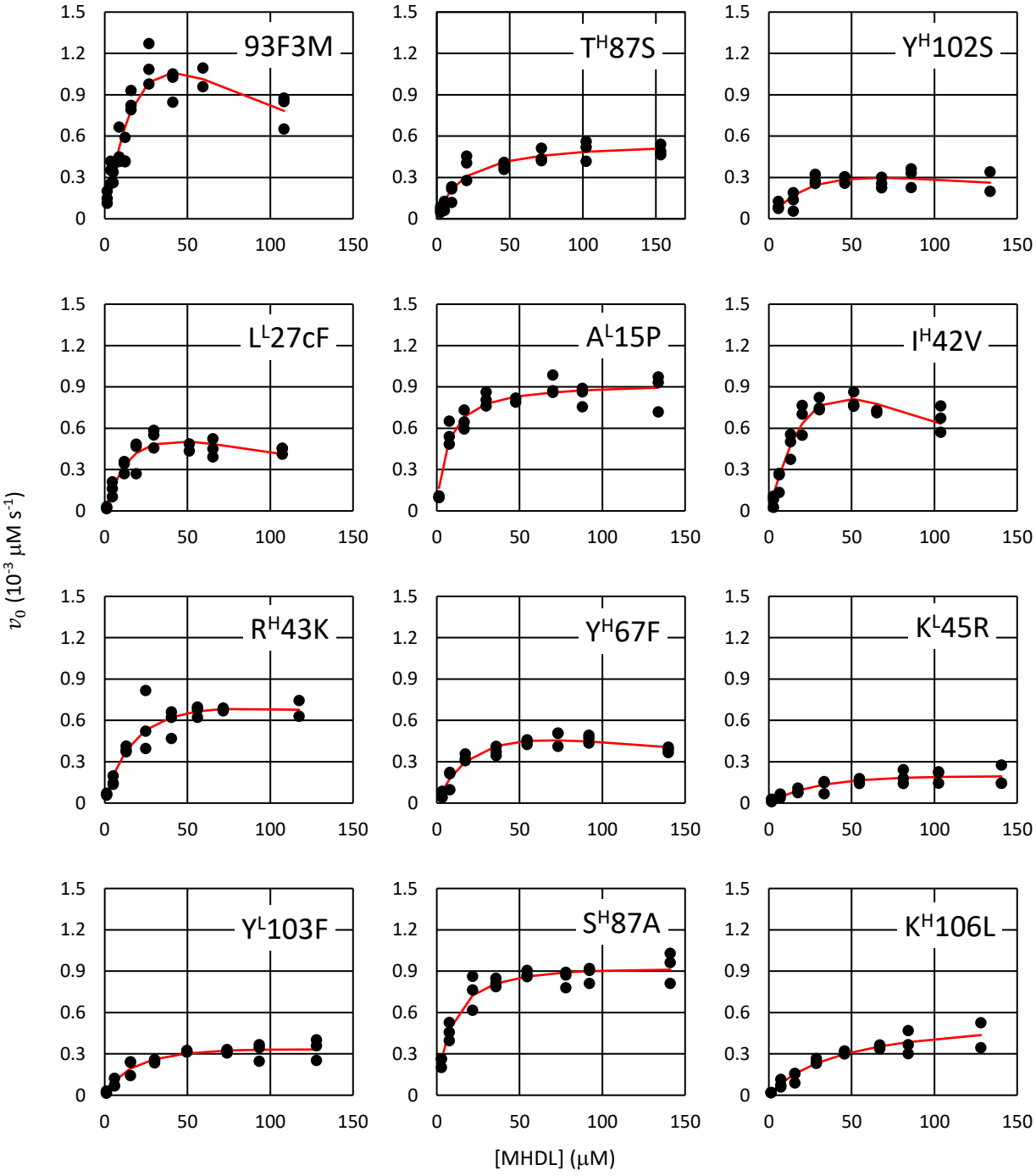


Figure 13. Steady-state rate vs. substrate concentration plots. Shown for MHDL cleavage for mature 93F3 and each mutant (black circles, data; red lines, best fit to Equation 9).

Table 6. Summary of fitted steady-state kinetic parameters. n.r. indicates no reaction.GL is germline / hypermutation, DB is database, and EX is exploratory derived

Mutation origin		K_M (μM)	k_{cat} (min^{-1})	K_i (μM)	k_{cat}/K_M (10^{-3} $min^{-1}\cdot\mu M^{-1}$)
GL	$93F3^{M}$	51 ± 4	2.15 ± 0.08	34 ± 4	42 ± 4
	$T^{H}87S$	17 ± 2	0.31 ± 0.01	$>10^6$	18 ± 2
	$Y^{H}102S$	41 ± 5	0.37 ± 0.02	110 ± 20	9 ± 1
	$L^{L}27cF$	26 ± 2	0.60 ± 0.02	77 ± 9	23 ± 2
	$A^{L}15P$	6.1 ± 0.8	0.52 ± 0.01	$>10^6$	90 ± 10
	$L^{L}46K$	n.r.	n.r.	n.r.	n.r.
DB	$I^{H}42V$	97 ± 5	2.51 ± 0.07	21 ± 1	26 ± 2
	$R^{H}43K$	21 ± 2	0.58 ± 0.02	400 ± 200	28 ± 3
	$Y^{H}67F$	30. ± 2	0.47 ± 0.01	160 ± 20	16 ± 1
	$K^{L}45R$	41 ± 7	0.18 ± 0.01	500 ± 300	4 ± 1
	$Y^{L}103F$	20. ± 2	0.25 ± 0.01	700 ± 400	13 ± 2
EX	$S^{H}87A$	8.5 ± 0.8	0.46 ± 0.01	$>10^3$	54 ± 5
	$K^{H}106L$	44 ± 4	0.29 ± 0.01	$>10^6$	7 ± 1

would lead to a decrease in 460 nm emission. To properly fit the data, photobleaching of MHDL and substrate inhibition of the reaction by MHDL needed to be taken into account.

The steady-state kinetic parameters vary significantly among the wild type 93F3 and the single-point framework mutants (Table 6). Almost without exception and irrespective of the

amino acid exchange, modification of residues in the framework region reduced the catalyst rate constant k_{cat}, and decreasing the Michaelis constant K_M.

Predicting the Energetic and Catalytic Effect of Framework Mutations

Steady-state kinetics are inescapably composite measures of catalysis, which necessitates careful interpretation of the parameters obtained from this analysis. As such, we describe the following generalized model for evaluating changes in k_{cat}, K_M, (k_{cat}/K_M), and K_i as they may relate to the protein energy landscape. Figure 11 shows the proposed energy landscape used in our interpretation.

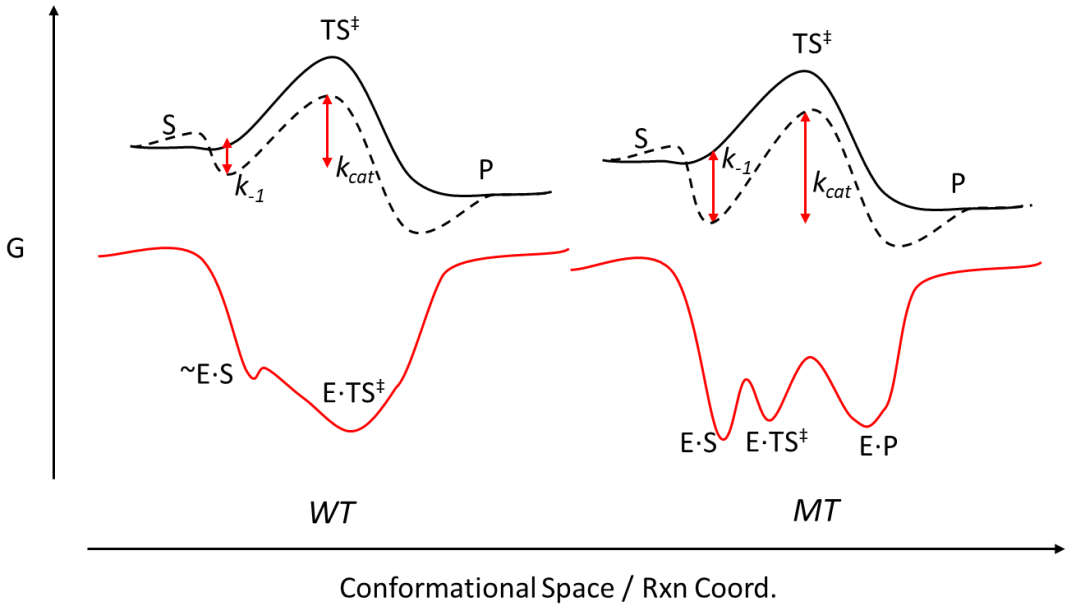


Figure 14. Energy Landscape Interpretation of the Steady-State Kinetic Results. The solid black line corresponds to an uncatalyzed reaction, while the dashed line is the catalyzed version resulting from contributions from the protein energy landscape (red). ~E·S corresponds to a conformation *approximately* conducive to substrate binding. WT and MT refer to the wild-type and mutants, respectively. Double-headed red arrows demonstrate changes in energetic gaps between WT and MT.

The turnover number (k_{cat}) is perhaps the most straightforward and important to interpret. This effectively summarizes catalytic capacity of the average binding site including all steps

following initial substrate binding, including product release, substrate dissociation, or formation of the probe species [Johnson 2019]. While a useful measure of per-site catalytic turnover, a connection to conformational shifts is difficult to clearly discern; though crystallographic and kinetic studies point unsurprisingly to a requirement for conformations with rigidly oriented catalytic residues in transition state-stabilizing configurations [Richard 2019]. The protein will bind an intermediate in a transition state-stabilizing configuration, using this slightly imperfect binding geometry (and thus weaker binding energy) to "guide" the intermediate towards the transition state [Richard 2019].

Because of this, the protein potential energy surface *past* substrate binding steps should consist of a series of narrow energy wells corresponding to conformations associated to the chemical events, which are separated by relatively small kinetic barriers. Because abzymes lack the evolutionarily-tuned chemical-conformational coupling seen in enzymes, they may lack the ability to accommodate multiple TS geometries. Given the approximate nature of the hapten as a TS analog used to generate an abzyme, its resemblance in some cases to substrate and product may impede substrate reactivity or product release, since the abzyme will also acquire some affinity for both. This is due to thermodynamic stabilization that increases the energetic gap to the transition state (substrate inhibition) or to the free product and abzyme (product inhibition). Therefore, reasonable criteria for catalytically beneficial (i.e., improved k_{cat}) framework mutations may be (1) weaker ligand-induced protein stabilization allowing for kinetically feasible substrate reactivity and product release and (2) TS-stabilizing conformations ("catalytically active" or bound states) are preserved such that the basic alignment of catalytic residues in their

optimal and fixed orientations is retained within the mean residence time of the substrate or intermediate [Richard 2019].

K_M is usually described as an estimate for the substrate dissociation constant (K_d). However, complications arise with this approximation without underlying knowledge of the relative magnitudes of intrinsic rate constants since K_M is a function of rate constants [Johnson 2019]. Our interpretation of K_M follows the minimal model conventional definition: $K_M = (k_{-1} + k_{cat})/k_1$ which is approximately equal to K_d when $k_{-1} \gg k_{cat}$, i.e. when substrate dissociation is much faster than the rate of the chemical step. Because both our observed k_{cat}values and reliable estimates of antibody dissociation rate constants (k_{-1}) for cognate ligands are in the 10^{-3} s^{-1} range [Zimmermann *et al.* 2006], this approximation is likely to be invalid. While we did not measure binding rate constants, it seems reasonable that the association rate (k_1) will be nearly constant as binding rates are theorized to be insensitive to mutations, even those perturbing electrostatic attraction between binding site and ligand [Pang *et al.* 2017]. Therefore, we chose to interpret changes in K_M as arising mainly from changes in the relative magnitudes of k_{-1} and k_{cat}. Because we primarily observed decreases in K_M, there are two main non-mutually exclusive possibilities accounting for this: (1) framework mutations alter catalytic residue orientations such that the TS is less stabilized, thus increasing the energy gap between the substrate-bound state (ES) to TS and leading to a decreased k_{cat}, or (2) k_{-1} decreases as a result of tighter ES binding, concurrently increasing the energy gap to TS, assuming TS energy remains fixed.

An energy landscape interpretation of the kinetic data may rely upon assuming that the framework mutations (in general) confer access to new and/or previously unattainable conformational states. If the framework mutations generally confer greater conformational heterogeneity,

it is more likely that a conformation conducive to MHDL binding is more stable in a framework mutant compared to the mature 93F3's rigid active site that was evolved to bind the TS. If such conformations become accessible to substrate, the changes in K_M and k_{cat} may be interrelated as tighter MHDL binding (lower k_{-1} and K_M) increases the kinetic barrier to TS (lower k_{cat}). This again assumes that free energy of TS remains constant in the mutant proteins. If mutations lead to misaligned catalytic residues, the TS barrier corresponding to k_{cat} may rise and thus lower turnover independently of the changes in the free energy of the ES minimum. This possibility seems less likely, however, since residue alignment for TS stabilization is to some extent related to binding affinity, especially in antibodies where binding sites are mostly static.

Interestingly, this trend is not seen in the substrate inhibition constant, K_i. If conformations conducive for binding MHDL have emerged and simultaneously lose the ability to detectably bind a second MHDL molecule, it would seem the MHDL-binding conformations are more specific for MHDL and perhaps generally more isolated from the bulk solvent environment, assuming the second MHDL molecule binds at the abzyme active site. The consequence of enhanced substrate binding is clear in the k_{cat}results, where all mutants (with a single exception) saw more than a twofold decline. Coupled with the K_M results, this is consistent with substrate inhibition stemming from a novel free conformation of the abzyme with higher substrate affinity, and thus an increased kinetic barrier to moving out of the substrate-abzyme complex towards product release (Figure 11).

Substrate inhibition may proceed under a variety of mechanisms [Kokkonen *et al.* 2021; Reed *et al.* 2010] Binding to an allosteric site is among the most common [Kokkonen *et al.* 2021], but this is an unlikely scenario in an abzyme *Fab* fragment since antibodies are not

commonly regulated by small hydrophobic molecules [Alberts *et al*. 2002]. Therefore, substrate inhibition in 93F3 likely arises from the sequential binding of a second MHDL molecule in the binding site with π-stacking between MHDL naphthyl rings and contacts with binding site residues providing the driving force for the association. This complex is assumed to be catalytically inactive, which accounts for the decline in reaction velocity at higher MHDL concentrations. Alternatively, the second MHDL molecule may bind to any other bound intermediate in the reaction with the same effect [Kokkonen *et al.* 2021]. One explanation for the near-complete loss of substrate inhibition in most mutant proteins may be the emergence of conformational gating [Pang *et al.* 2017; Zhou *et al.* 1998], whereby MHDL binds to a conducive active site that subsequently transitions to a "closed" conformation (i.e., an initial conformational selection followed by an induced fit). The function of distinct closed enzyme conformations is in part to provide desolvation allowing for controlled access to active site residues, such that doubly bound substrate complexes are less likely [Zhou *et al.* 1998]. The stronger MHDL-binding conformers may therefore also be more gated / closed conformations that weaken the association energy for a second substrate molecule. In all, this explanation requires that framework mutations enhance conformation heterogeneity in the free state such that substrate encounters and binds to binding-conducive conformations, while the bound state itself remains closed, rigid, and catalytically impeded in a deeper thermodynamic minimum.

A single exception to the general trends is the databased-derived mutant $I^{H}42V$, which exceeded the mature k_{cat} and K_M value but *gained* some association energy for the substrate-inhibited complex. 42V is in close proximity to the binding site, and may contribute to both catalysis and substrate binding. The rather similar nature of Ile and Val apparently leaves catalytic

turnover slightly enhanced, but its effect upon K_M shows that it plays a role in binding. Because this mutant also showed stronger substrate inhibition this may indicate a potentially reconfigured binding site that favorably accommodates a second MHDL molecule. A fairly simple explanation could be the slight decrease in volume from Ile to Val in the binding site which could provide the necessary volume (a gain of ~ 26 $Å^3$) for a second MHDL molecule.

Although K_M has limited kinetic meaning and can be difficult to interpret without independent knowledge of underlying rate constants, it remains useful for defining (k_{cat}/K_M). This ratio is perhaps the most important parameter for evaluating catalytic *proficiency*. k_{cat}/K_M is equivalent to the product of the intrinsic rate constant for substrate binding (k_1) and the probability of bound substrate reacting to form products ("commitment factor") [Johnson 2019]. If k_1 remains approximately constant across framework mutants [Pang *et al*. 2017], changes in k_{cat}/K_M may largely be due to changes in the commitment factor, defined as $k_{cat}/(k_{-1} + k_{cat})$ [Johnson 2019]. Clearly, improved binding affinities leading to lower k_{-1} and k_{cat} will have no effect upon k_{cat}/K_Mif the TS energy remains fixed since k_{-1} and k_{cat} will decrease by the same amount. If the k_{cat} barrier instead increased (i.e., k_{cat} decreased) by more than that accounted for by the enhanced binding energy, the commitment factor will fall as less substrate is processed. This same mechanism (poorer TS stabilization) will of course also yield a lower commitment factor in the absence of enhanced substrate binding affinity, which avoids the need to invoke the emergence of new protein conformations.

CHAPTER 5

METRICS FOR PROTEIN FLEXIBILITY

Background & Methods– Protein 3D Spectra and Excitation-Dependent Stokes Shift

The conventional hypothesis that framework residues (and the somatic hypermutation process in general) maintain structural rigidity and limit an antibody's conformational landscape is possible to evaluate spectroscopically. Tryptophan (Trp) and tyrosine (Tyr) fluorescence are reliable probes of their local electrostatic environment (Trp more so than Tyr). We collected 3-dimensional fluorescence spectra for all 93F3 mutants as a general but quantitative measure of the distribution in fluorophore-solvent environments [Kwok *et al.* 2021]. In a protein, the relevant fluorophore is Trp, and its immediate electrostatic environment formed by neighboring side chains and the protein interior surrounding the Trp residue can be thought of as the "solvent".

The extent of excited state stabilization by the solvent is governed primarily by the relaxation time of the solvent relative to the fluorophore's excited state lifetime (i.e., the well-known Stokes shift: if the solvent can reorganize fast enough to stabilize the excited state, the energy level from which emission occurs will be closer to the ground state and emission wavelength will be longer) [Demchenko *et al.* 2002]. For Trp in a protein population, the particularly slow solvent relaxation times (relative to the fluorescence lifetime) and thus the extent of excited state stabilization can be probed by varying the excitation wavelength [Kwok *et al.* 2021; Mukherjee *et al.* 1995]. The extent to which Trp emission shifts as a function of

excitation wavelength is related to the equilibrium distribution of Trp excited state environments as outlined by Kwok *et al.* [Kwok *et al.* 2021]. In brief, by gradually lowering excitation energy, subpopulations of Trp microenvironments in configurations more closely resembling environments equilibrated with the Trp excited state are selectively excited and fluorescence occurs from these lower-energy excited state, all under the time for complete solvent relaxation [Kwok *et al.* 2021]. If more intermediate Trp-solvent states are available between the two stabilization extremes (i.e., environments fully equilibrated with the Trp ground state vs. with the Trp excited state), this suggests a more flexible protein and is reflected in a more pronounced excitation dependence of the emission shift.

We collected emission spectra (300-450 nm) at excitation wavelengths between 240 nm and 300 nm to quantify changes in emission centers of spectral mass (CSM), a sensitive measure of emission shifts, defined as [Kwok *et al.* 2021]:

$$CSM(\lambda_{ex}) = \frac{\sum I(\lambda_{ex}, \lambda_{em,i}) \cdot \lambda_{em,i}}{\sum I(\lambda_{ex}, \lambda_{em,i})} \tag{12}$$

where $I(\lambda_{ex}, \lambda_{em,i})$ is emission intensity at wavelength $\lambda_{em,i}$ for excitation at wavelength λ_{ex}. The dependency of CSM on the excitation wavelength (λ_{ex}) can be modeled (assuming a two-state model and a relatively long solvent relaxation time) as a sigmoidal function described by

$$CSM(\lambda_{ex}) = \frac{CSM_{FC} + CSM_R \cdot e^{m(\lambda_{ex} - \lambda_{ex,50\%})/RT}}{1 + e^{m(\lambda_{ex} - \lambda_{ex,50\%})/RT}} \tag{13}$$

where CSM_{FC} is the CSM corresponding to "Franck-Condon absorption", i.e. micro-environments in equilibrium with the Trp ground state, selectively excited at shorter λ_{ex}, CSM_R is the CSM corresponding to micro-environments in equilibrium with the Trp excited state, selectively excited at longer λ_{ex}, m is related to the width of the transition, $\lambda_{ex,50\%}$ is the λ_{ex} at

which half-maximal emission shift occurs, *R* is the universal gas constant, and *T* is temperature in K [Kwok *et al.* 2021]. Kwok *et al.* provide interpretations for changes in the above parameters as they relate to the distribution of Trp-solvent environments. ΔCSM, the difference between CSM_R and CSM_{FC}, should report on the *range* of populated Trp-solvent environments with a smaller difference corresponding to a tighter distribution in Trp-solvent energies. In this case, the extremes in conformation will be more similar in energy; which suggests a more rigid protein and vice versa [Kwok *et al.* 2021]. The slope at the midpoint, *m*, is a useful measure of the spectral distribution and number of intermediate emissive states: a greater *m* (i.e., a shaper transition) reflects fewer accessible states separated by a larger energy gap and vice versa [Kwok *et al.* 2021]. Example protein 3D spectra and plots of CSM vs. λ_{ex} with fits to Equation 12 is shown in Figure 15.

Of course, interpretation of ΔCSM and *m* values is most straightforward in a single Trp protein. Because 93F3 contains many Trp residues, each with its own unique microenvironment, the interpretation of the excitation-dependent emission shifts is significantly complicated. Therefore, a major caveat of our analysis is that we are unable to distinguish between spectral shifts for Trp with identical microenvironments (in which case we can apply the theory explained above) and spectral shifts due to excitation of Trp in nonuniform microenvironments and therefore heterogeneous emission spectra. Fitted ΔCSM and *m* values are shown in Table 7.

Table 7. Excitation-dependent stokes shift results.

		m (kcal nm)	ΔCSM (nm)
GL	$93F3^{M}$	761.01	2.08
	$T^{H}87S$	1028.07	1.90
	$Y^{H}102S$	974.95	2.27
	$L^{L}27cF$	912.17	3.73
	$A^{L}15P$	947.43	2.00
	$L^{L}46K$	1093.45	1.76
DB	$I^{H}42V$	761.00	2.07
	$R^{H}43K$	1018.83	2.11
	$Y^{H}67F$	1176.05	1.91
	$K^{L}45R$	1030.27	2.60
	$Y^{L}103F$	964.85	2.26
EX	$S^{H}87A$	967.12	2.12
	$K^{H}106L$	757.05	2.65

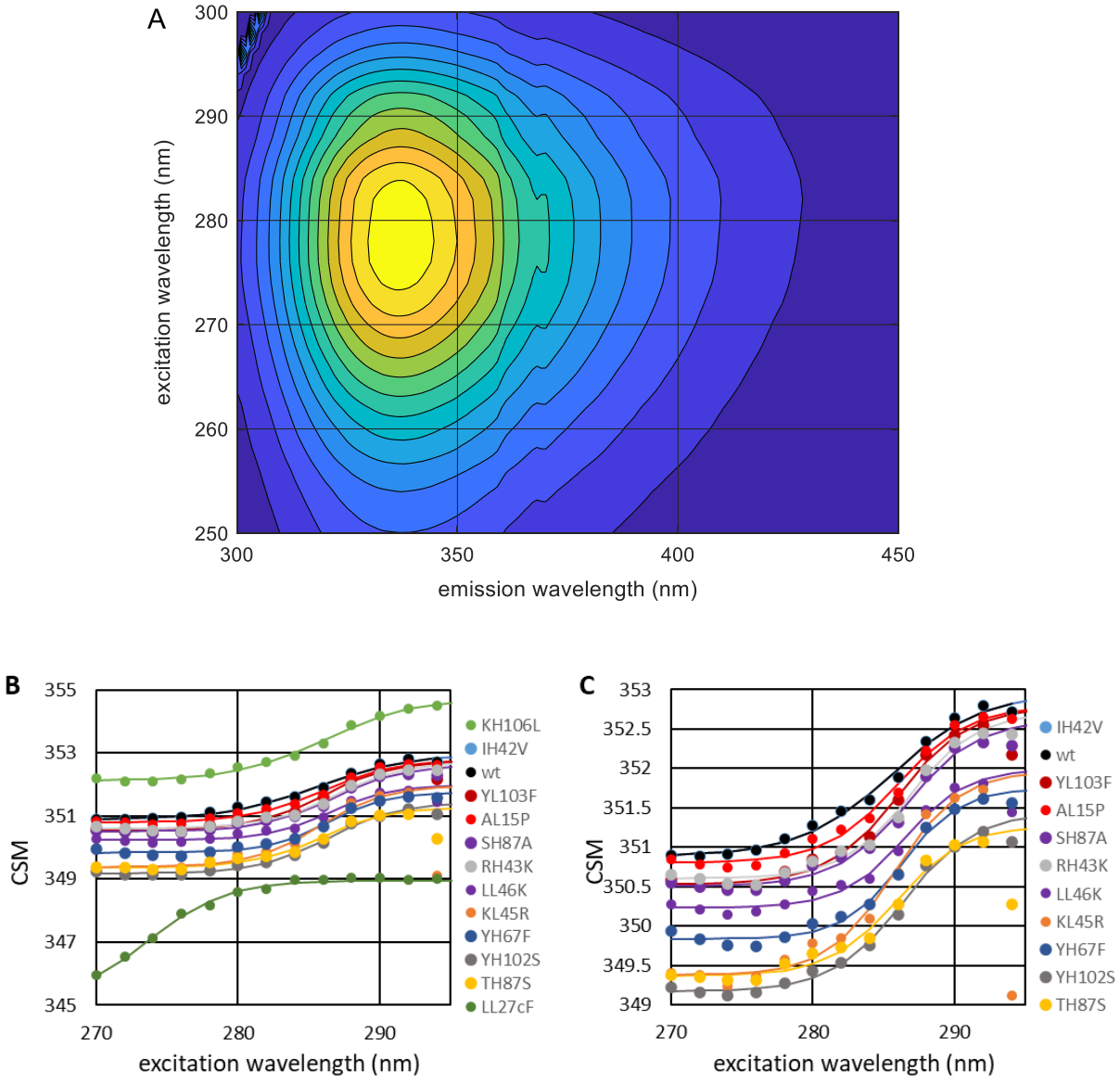


Figure 15. Excitation-dependent Stokes shift analysis. (A) Excitation-emission matrix for $93F3^{M}$. (B) CSM as a function of λ_{ex}. (C) Zoom-in of panel (B) with the data for $K^{H}106L$ and $L^{L}27cF$ removed.

Computational Determination of ΔG and ΔS_{vib}

To assess changes in protein flexibility and dynamics generally, we calculated the change in folding free energy ($\Delta\Delta G$) and vibrational entropy (ΔS_{vib}) between $93F3^{M}$ and each

framework mutant. To this end, we submitted all mutants to DynaMut, a web server that calculates a range of protein thermodynamic parameters using normal mode analysis [Rodrigues *et al.* 2018, 2021]. Positive ΔG or ΔS_{vib} suggest an increase in the flexibility of the protein.

While ΔG is a measure of the overall protein stability and in some cases is related to internal vibrational fluctuations [Karshikoff *et al.* 2015], vibrational entropy is related to the less stringently-defined concept of protein *flexibility*. Flexibility is often associated with *configurational* entropy which is composed of both *conformational* entropy, a function of the number of energetically distinct well minima in the free energy landscape, and *vibrational* entropy which is linked to vibrations within these wells [Goethe *et al.* 2015]. We hypothesize that conformational and vibrational entropy are interrelated; a mor flexible protein will populate more conformations in broader wells. . A positive ΔS_{vib} will therefore be interpreted as an increase in flexibility of the framework mutants, i.e. changes in the weighted sum of bond angles and vibrational distributions that permit increased fluctuation about conformational minima [Ohkubo *et al.* 2006].

Thermal Protein Stability

Protein denaturation temperatures are a useful measure of the net thermal stability, and to a good extent, conformational rigidity [Vihinen *et al.* 1987]. Framework residues are thought to serve primarily structural roles, and so their reversion from mature to germline identity may destabilize antibodies and therefore lower denaturation temperatures. We collected denaturation temperatures in order to discern any relationship between abzyme stability and catalytic parameters.

We used SYPRO Orange as a reporter dye (10-40X) in PBS pH 7.4 (DMSO 0.2-0.8% v/v Dye and protein solutions were mixed in equal volumes in 96-well plates and fluorescence at 620 nm was monitored as a function of temperature using a QuantStudio 3 qPCR instrument (Thermo Fisher). QuantStudio software was used for calculating first derivatives of fluorescence vs. temperature traces. We refined the transition temperature by evaluating third derivatives from this data using a Savitzkye-Golay algorithm as described in Bhayani *et al.* 2022.

Table 8. T_m results for 93F3 and the single mutants.

		T_m (°C)
GL	$93F3^{M}$	78.08 ± 0.08
	$T^{H}87S$	77.71 ± 0.14
	$Y^{H}102S$	77.46 ± 0.35
	$L^{L}27cF$	77.63 ± 0.08
	$A^{L}15P$	77.51 ± 0.08
	$L^{L}46K$	79.06[a]
DB	$I^{H}42V$	77.06 ± 0.20
	$R^{H}43K$	70.95 ± 0.30
	$Y^{H}67F$	72.40 ± 0.08
	$K^{L}45R$	78.40[a]
	$Y^{L}103F$	76.84 ± 0.20
EX	$S^{H}87A$	71.36[a]
	$K^{H}106L$	77.65 ± 0.08

[a] no replicated were obtained due to low protein concentration

Thermal denaturation temperatures (T_m) for all 93F3 mutants are shown in Table 8.

Correlations Between Kinetic Parameters and Protein Flexibility Metrics

Turnover rates for 93F3 varied between framework mutants, with all but one mutant (I^H42V) showing reduced values for k_{cat} compared to the native protein. In contrast, all mutants (exception: I^H42V) showed improved K_M values; however only two (A^L15P and S^H87A) led to enhanced catalytic proficiencies (k_{cat}/K_M). K_i values almost unanimously increased; most beyond a measurable extent. Mutant thermal denaturation points were lower than $93F3^M$ (with a few exceptions), confirming the general findings that framework mutations may provide stability and confer fully binding capacity [Sela-Culang *et al.* 2013]. However, denaturation temperatures showed no discernable relationship to any catalytic parameters or flexibility metrics.

In general, then, framework residues appear to be required for preserving turnover rate, which seems predictable from stability arguments alone: even slight variations in catalytic geometry transmitted from distant points in the protein could perturb effective TS stabilization [Richard 2019]. This is especially true if the selected framework mutations facilitate conformational transitions within the bound state basin. For our interpretation of flexibility trends, we focus on potential changes to the conformational landscape effected by framework residues that may explain the observed trends. Furthermore, because ΔS_{vib} calculations are necessarily carried out on $93F3^M$'s PDB free form (i.e., un-ligated) crystal structure, and excitation-dependent emission spectra were measured for solutions of the free protein, our discussion will be mostly restricted to changes occurring in the conformational landscape of the free form.

Interestingly, K_M values were negatively associated with gains in vibrational entropy (i.e., more vibrationally heterogeneous mutants showed "improved" K_M's). This indirectly supports our central hypothesis that framework mutations increase conformational sampling by the free abzyme. As explained above, we hypothesize that vibrational and conformational entropies are positively correlated. An increase in vibrational entropy therefore suggests an increase in conformational entropy, which will increase the probability that the substrate encounters a binding site with favorable geometry, thereby lowering K_M values (see Chapter 4). While the novel conformations and their associated S_{vib} are unlikely to be functionally coupled to the binding or catalytic processes (an effect known as "vibrational steering" in enzymes [Niessen *et al.* 2017; Pang *et al.* 2017]), binding processes and the lifetimes of binding-conducive conformational substates may be on comparable timescales and sufficiently correlated [Ishikawa *et al.* 2008; Finkelstein *et al.* 2007].

Higher ΔS_{vib} was also positively related to k_{cat} and k_{cat} / K_M, which is especially pronounced when the $93F3^M$ and I^H42V outliers are excluded. Whether ligand binding alters a conformational landscape and in what way is highly context-dependent [Moritsugu *et al.* 2010; Turton *et al.* 2014; Kabir *et al.* 2021], complicating an interpretation of the k_{cat} and k_{cat} / K_M values given their dependence on the bound state conformational landscape. However, if S_{vib} is essentially preserved in the bound state, it would seem reasonable that a greater range of vibrational substates surveyed by the catalytic residues would increase the likelihood of transiently assuming an optimal TS-stabilizing substate, leading to higher turnover.

As an additional estimate to protein conformational distributions, we tracked spectroscopically distinct Trp-solvent environments using excitation-dependent emission spectra.

We observed that fewer energetically discrete Trp-solvent environments (as determined from higher ΔCSM and *m* values) [Kwok *et al.* 2021] is associated with a lower K_M as seen in Figure 16, in somewhat of a disagreement with the ΔS_{vib} results. Insofar as distinct Trp-solvent environments are representative of conformational minima, this trend is consistent with a decrease in free state conformational entropy but interestingly also a simultaneous increase in S_{vib}. This may indicate the K_M benefit conferred by framework mutations is more closely tied to changes in allowed vibrational substates (S_{vib}) and not conformational distributions per se. ΔCSM however showed essentially no relationship to K_M values, indicating changes in the *distribution* or occupancy of these discrete Trp-solvent environments were unrelated to binding.

These spectroscopic measures of flexibility (ΔCSM and *m*) were negatively associated with turnover. This indicates that (1) framework mutations increasing conformational entropy (i.e., more Trp-solvent states) confer higher turnover, and (2) more energetic similarity across Trp-solvent states (conformations) is also linked with catalytic proficiency.

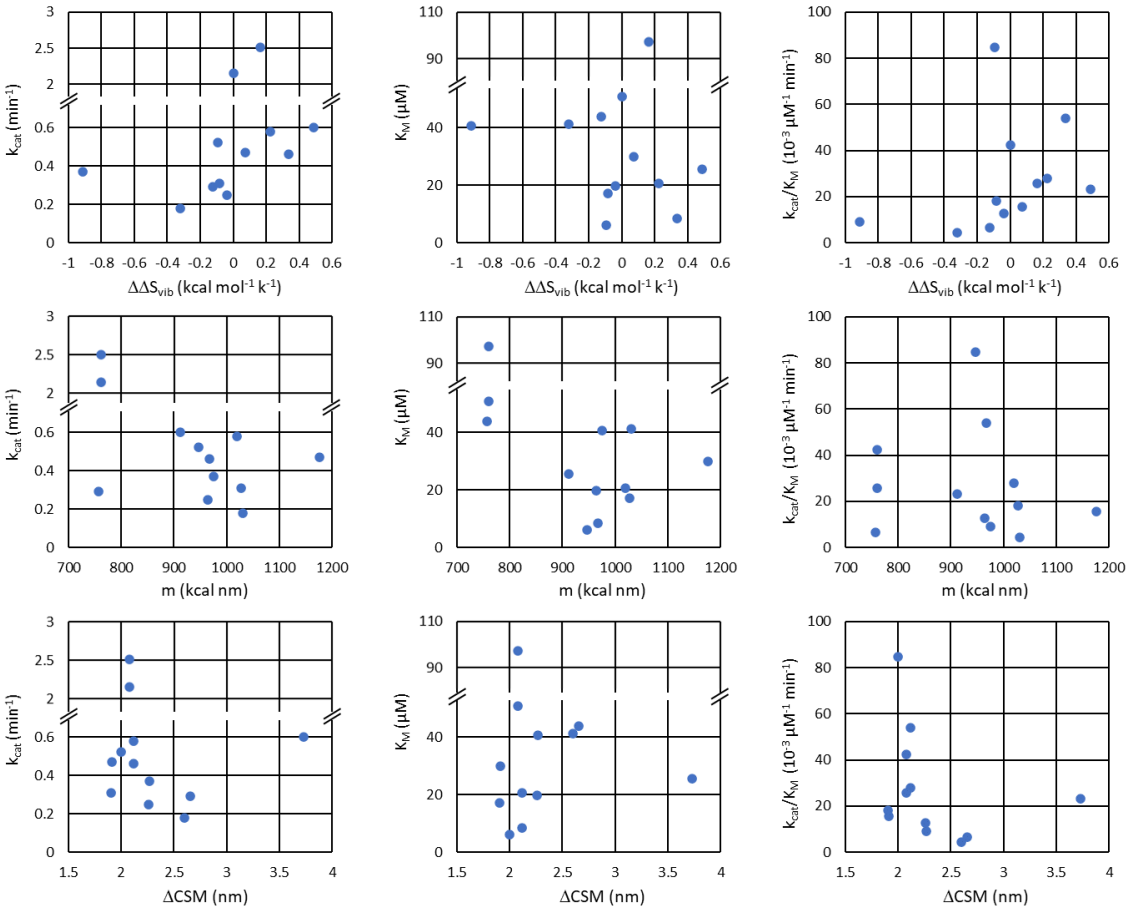


Figure 16. Catalytic parameters plotted against flexibility metrics

CHAPTER 6

SUMMARY

During affinity maturation, antibodies acquire point mutations that lead to enhanced binding strength to a particular ligand while simultaneously becoming more conformationally restricted as they become increasingly complementary to their ligand's electrostatics and geometry. This property has catalytic potential in transition-state (TS) stabilization whereby kinetic barriers (i.e., the energetic gap between the substrate and its TS) are reduced and reaction rates are enhanced. This basic strategy is thought to underlie the majority of the rate enhancement observed for enzymes. Catalytic antibodies ("abzymes") combine antibody binding specificity and affinity with the enzymatic catalytic strategy. Abzymes are designed to bind a single, ideally rate-limiting TS of a given organic reaction. While abzymes are an impressive verification of enzymatic TS stabilization as a catalytic strategy, abzymes fall short with regard to their catalytic rate enhancements. A largely overlooked explanation for this pertains to the differences in conformational heterogeneity between abzymes and enzymes. Abzymes, like most mature antibodies, are likely rigid lock-and-key binders with little to no conformational arrangements. Enzymes, by contrast, are flexible proteins subject to directed conformational shifts that are mechanochemically coupled to the substrate.

In the present work, we attempted to restore conformational heterogeneity to a model abzyme to examine its effect upon catalysis. Because our goal was to retain catalytic activity in our model abzyme, mutations were selected in the antibody framework region (FWR), which is

thought to be responsible for structurally scaffolding the complementary determining regions (CDRs) where ligand binds. We do this using the natural maturation process of antibodies, somatic hypermutation. We identified residues potentially contributing to rigidity both via somatic hypermutation and through cross-reference to a database of common mutations we compiled. We evaluated catalytic performance with steady-state kinetics, and flexibility with spectroscopic and computational techniques.

We observed significant decreases in turnover rates, but improvements in the Michealis constant (K_M). We interpret this general trend as improved substrate binding as a result of increased conformational sampling, but while also raising the barrier to the TS, thus lowering turnover rates. Perturbation of binding site residue alignment could be a reasonable explanation for lower turnover in framework mutants. We utilized computationally-determined changes in vibrational entropy as an approximate metric to conformational distributions. Vibrational entropy was positively related to turnover and improved K_M, supporting our general hypothesis that conformational sampling confers enhanced catalysis. We found only weak relationships between catalytic parameters and Trp fluorescence shifts, a spectroscopic measure of flexibility, indicating either that the method failed to capture significant changes in conformational distributions or that conformational heterogeneity is unrelated to the loss of catalytic capacity in our selected mutations.

Our results emphasize the requirement for *directed* conformational transitions in protein catalysts, the complex relationship between conformational and reaction energy landscapes, and the functional contrast between antibodies and enzymes generally. While we may have been successful in restoring conformational flexibility in the abzyme, this was apparently effective

only at improving K_M values; a difficult parameter to interpret. This result highlights the near-complete reliance of catalytic antibodies on TS-stabilization for catalysis, and their exquisite sensitivity to framework integrity. Antibody function is comparatively simple with respect to enzymes, and abzyme improvement via directed evolution may resemble to some degree the natural conformational evolution of many enzymes. In this way, abzyme improvement may serve as a model for understanding the functional evolution of natural enzymes.

To explore the evolution of conformational landscapes of protein catalysts and the corresponding catalytic impact, future work should rely on the mutationally stochastic but highly controlled selective strategy used in directed evolution protocols. Along the evolutionary trajectory of an abzyme (where catalysis may be improved via conformational flexibility), a multifaceted spectroscopic approach to characterizing the conformational landscape of free and bound states could be a particularly insightful study design.

APPENDIX A

CORRELATED MUTATIONS

Correlated Mutations. Shown are the total counts (C) and frequencies (f(%)) of sequences containing the indicated co-occurring mutations. An arbitrary cut-off was applied at approximately half of the top mutations' frequency for double and triples.

Table S1. List of V^H top correlated mutations.

Double correlated mutations

Mutations			C	f(%)
L21I	I101V		18	2.81
T96S	I101V		18	2.81
T96S	L21I		18	2.81
I101V	T35S		17	2.65
T96S	T35S		16	2.50
L21I	T35S		15	2.34
Q3K	Q6E		12	1.87
S96T	N40S		12	1.87
K72R	M39I		11	1.72
V80A	G74D		11	1.72
G74D	M39I		10	1.56
S36T	Q6E		10	1.56
R47G	G74D		10	1.56
R47G	V80A		10	1.56
T35S	M39I		9	1.40
A9P	M39I		9	1.40
V80A	M39I		9	1.40
Q5L	Q6E		9	1.40

Triple correlated mutations

Mutations			C	f(%)
L21I	I101V	T96S	18	2.81
I101V	T35S	L21I	15	2.34
I101V	T35S	T96S	15	2.34
L21I	T35S	T96S	15	2.34
V80A	G74D	R47G	10	1.56
K72R	M39I	V80A	8	1.25
G74D	M39I	K72R	7	1.09
G74D	M39I	V80A	7	1.09
G74D	M39I	R47G	7	1.09
G74D	M39I	A55V	7	1.09
G74D	M39I	R14T	7	1.09
G74D	M39I	F103H	7	1.09
K72R	M39I	R47G	7	1.09
K72R	M39I	A55V	7	1.09
K72R	M39I	R14T	7	1.09
K72R	M39I	F103H	7	1.09
V80A	M39I	R47G	7	1.09
V80A	M39I	A55V	7	1.09
V80A	M39I	R14T	7	1.09
V80A	M39I	F103H	7	1.09
R47G	M39I	A55V	7	1.09
R47G	M39I	R14T	7	1.09
R47G	M39I	F103H	7	1.09
A55V	M39I	R14T	7	1.09
A55V	M39I	F103H	7	1.09
R14T	M39I	F103H	7	1.09
K72R	G74D	V80A	7	1.09
K72R	G74D	R47G	7	1.09
K72R	G74D	A55V	7	1.09
K72R	G74D	R14T	7	1.09
K72R	G74D	F103H	7	1.09
V80A	G74D	A55V	7	1.09
V80A	G74D	R14T	7	1.09
V80A	G74D	F103H	7	1.09
R47G	G74D	A55V	7	1.09
R47G	G74D	R14T	7	1.09
R47G	G74D	F103H	7	1.09
A55V	G74D	R14T	7	1.09
A55V	G74D	F103H	7	1.09
R14T	G74D	F103H	7	1.09
V80A	K72R	R47G	7	1.09
V80A	K72R	A55V	7	1.09
V80A	K72R	R14T	7	1.09
V80A	K72R	F103H	7	1.09
R47G	K72R	A55V	7	1.09
R47G	K72R	R14T	7	1.09
R47G	K72R	F103H	7	1.09
A55V	K72R	R14T	7	1.09
A55V	K72R	F103H	7	1.09
R14T	K72R	F103H	7	1.09
L21I	I101V	G36S	7	1.09
T96S	I101V	G36S	7	1.09
T96S	L21I	G36S	7	1.09
N40S	I56V	R106N	7	1.09
N40S	I56V	K49R	7	1.09
R106N	I56V	K49R	7	1.09
R47G	V80A	A55V	7	1.09

R47G	V80A	R14T		7	1.09
R47G	V80A	F103H		7	1.09
A55V	V80A	R14T		7	1.09
A55V	V80A	F103H		7	1.09
R14T	V80A	F103H		7	1.09
R106N	N40S	K49R		7	1.09
T100A	D36S	T68N		7	1.09
T100A	D36S	L80S		7	1.09
T100A	D36S	D90E		7	1.09
T68N	D36S	L80S		7	1.09
T68N	D36S	D90E		7	1.09
L80S	D36S	D90E		7	1.09
A55V	R47G	R14T		7	1.09
A55V	R47G	F103H		7	1.09
R14T	R47G	F103H		7	1.09
T68N	T100A	L80S		7	1.09
T68N	T100A	D90E		7	1.09
L80S	T100A	D90E		7	1.09
R14T	A55V	F103H		7	1.09
L80S	T68N	D90E		7	1.09

Quadruple correlated mutations

Mutations				C	f(%)
T35S	I101V	L21I	T96S	15	2.34
M39I	G74D	K72R	V80A	7	1.09
M39I	G74D	K72R	R47G	7	1.09
M39I	G74D	K72R	A55V	7	1.09
M39I	G74D	K72R	R14T	7	1.09
M39I	G74D	K72R	F103H	7	1.09
M39I	G74D	V80A	R47G	7	1.09
M39I	G74D	V80A	A55V	7	1.09
M39I	G74D	V80A	R14T	7	1.09
M39I	G74D	V80A	F103H	7	1.09
M39I	G74D	R47G	A55V	7	1.09
M39I	G74D	R47G	R14T	7	1.09
M39I	G74D	R47G	F103H	7	1.09
M39I	G74D	A55V	R14T	7	1.09
M39I	G74D	A55V	F103H	7	1.09
M39I	G74D	R14T	F103H	7	1.09
M39I	K72R	V80A	R47G	7	1.09
M39I	K72R	V80A	A55V	7	1.09
M39I	K72R	V80A	R14T	7	1.09
M39I	K72R	V80A	F103H	7	1.09
M39I	K72R	R47G	A55V	7	1.09
M39I	K72R	R47G	R14T	7	1.09
M39I	K72R	R47G	F103H	7	1.09
M39I	K72R	A55V	R14T	7	1.09
M39I	K72R	A55V	F103H	7	1.09
M39I	K72R	R14T	F103H	7	1.09
M39I	V80A	R47G	A55V	7	1.09
M39I	V80A	R47G	R14T	7	1.09
M39I	V80A	R47G	F103H	7	1.09
M39I	V80A	A55V	R14T	7	1.09
M39I	V80A	A55V	F103H	7	1.09
M39I	V80A	R14T	F103H	7	1.09
M39I	R47G	A55V	R14T	7	1.09
M39I	R47G	A55V	F103H	7	1.09
M39I	R47G	R14T	F103H	7	1.09
M39I	A55V	R14T	F103H	7	1.09
G74D	K72R	V80A	R47G	7	1.09
G74D	K72R	V80A	A55V	7	1.09
G74D	K72R	V80A	R14T	7	1.09
G74D	K72R	V80A	F103H	7	1.09
G74D	K72R	R47G	A55V	7	1.09
G74D	K72R	R47G	R14T	7	1.09
G74D	K72R	R47G	F103H	7	1.09
G74D	K72R	A55V	R14T	7	1.09
G74D	K72R	A55V	F103H	7	1.09
G74D	K72R	R14T	F103H	7	1.09
G74D	V80A	R47G	A55V	7	1.09
G74D	V80A	R47G	R14T	7	1.09
G74D	V80A	R47G	F103H	7	1.09
G74D	V80A	A55V	R14T	7	1.09
G74D	V80A	A55V	F103H	7	1.09
G74D	V80A	R14T	F103H	7	1.09
G74D	R47G	A55V	R14T	7	1.09
G74D	R47G	A55V	F103H	7	1.09
G74D	R47G	R14T	F103H	7	1.09
G74D	A55V	R14T	F103H	7	1.09
K72R	V80A	R47G	A55V	7	1.09
K72R	V80A	R47G	R14T	7	1.09
K72R	V80A	R47G	F103H	7	1.09
K72R	V80A	A55V	R14T	7	1.09
K72R	V80A	A55V	F103H	7	1.09
K72R	V80A	R14T	F103H	7	1.09
K72R	R47G	A55V	R14T	7	1.09
K72R	R47G	A55V	F103H	7	1.09
K72R	R47G	R14T	F103H	7	1.09
K72R	A55V	R14T	F103H	7	1.09
I101V	L21I	T96S	G36S	7	1.09
V80A	R47G	A55V	R14T	7	1.09
V80A	R47G	A55V	F103H	7	1.09
V80A	R47G	R14T	F103H	7	1.09
V80A	A55V	R14T	F103H	7	1.09
D36S	T100A	T68N	L80S	7	1.09
D36S	T100A	T68N	D90E	7	1.09
D36S	T100A	L80S	D90E	7	1.09

D36S	T68N	L80S	D90E		7	1.09
R47G	A55V	R14T	F103H		7	1.09
T100A	T68N	L80S	D90E		7	1.09

Quintuple correlated mutations

Mutations					C	f(%)
M39I	G74D	K72R	V80A	R47G	7	1.09
M39I	G74D	K72R	V80A	A55V	7	1.09
M39I	G74D	K72R	V80A	R14T	7	1.09
M39I	G74D	K72R	V80A	F103H	7	1.09
M39I	G74D	K72R	R47G	V80A	7	1.09
M39I	G74D	K72R	R47G	A55V	7	1.09
M39I	G74D	K72R	R47G	R14T	7	1.09
M39I	G74D	K72R	R47G	F103H	7	1.09
M39I	G74D	K72R	A55V	V80A	7	1.09
M39I	G74D	K72R	A55V	R47G	7	1.09
M39I	G74D	K72R	A55V	R14T	7	1.09
M39I	G74D	K72R	A55V	F103H	7	1.09
M39I	G74D	K72R	R14T	V80A	7	1.09
M39I	G74D	K72R	R14T	R47G	7	1.09
M39I	G74D	K72R	R14T	A55V	7	1.09
M39I	G74D	K72R	R14T	F103H	7	1.09
M39I	G74D	K72R	F103H	V80A	7	1.09
M39I	G74D	K72R	F103H	R47G	7	1.09
M39I	G74D	K72R	F103H	A55V	7	1.09
M39I	G74D	K72R	F103H	R14T	7	1.09
M39I	G74D	V80A	R47G	K72R	7	1.09
M39I	G74D	V80A	R47G	A55V	7	1.09
M39I	G74D	V80A	R47G	R14T	7	1.09
M39I	G74D	V80A	R47G	F103H	7	1.09
M39I	G74D	V80A	A55V	K72R	7	1.09
M39I	G74D	V80A	A55V	R47G	7	1.09
M39I	G74D	V80A	A55V	R14T	7	1.09
M39I	G74D	V80A	A55V	F103H	7	1.09
M39I	G74D	V80A	R14T	K72R	7	1.09
M39I	G74D	V80A	R14T	R47G	7	1.09
M39I	G74D	V80A	R14T	A55V	7	1.09
M39I	G74D	V80A	R14T	F103H	7	1.09
M39I	G74D	V80A	F103H	K72R	7	1.09
M39I	G74D	V80A	F103H	R47G	7	1.09
M39I	G74D	V80A	F103H	A55V	7	1.09
M39I	G74D	V80A	F103H	R14T	7	1.09
M39I	G74D	R47G	A55V	K72R	7	1.09
M39I	G74D	R47G	A55V	V80A	7	1.09
M39I	G74D	R47G	A55V	R14T	7	1.09
M39I	G74D	R47G	A55V	F103H	7	1.09
M39I	G74D	R47G	R14T	K72R	7	1.09
M39I	G74D	R47G	R14T	V80A	7	1.09
M39I	G74D	R47G	R14T	A55V	7	1.09
M39I	G74D	R47G	R14T	F103H	7	1.09
M39I	G74D	R47G	F103H	K72R	7	1.09
M39I	G74D	R47G	F103H	V80A	7	1.09
M39I	G74D	R47G	F103H	A55V	7	1.09
M39I	G74D	R47G	F103H	R14T	7	1.09
M39I	G74D	A55V	R14T	K72R	7	1.09
M39I	G74D	A55V	R14T	V80A	7	1.09
M39I	G74D	A55V	R14T	R47G	7	1.09
M39I	G74D	A55V	R14T	F103H	7	1.09
M39I	G74D	A55V	F103H	K72R	7	1.09
M39I	G74D	A55V	F103H	V80A	7	1.09
M39I	G74D	A55V	F103H	R47G	7	1.09
M39I	G74D	A55V	F103H	R14T	7	1.09
M39I	G74D	R14T	F103H	K72R	7	1.09
M39I	G74D	R14T	F103H	V80A	7	1.09
M39I	G74D	R14T	F103H	R47G	7	1.09
M39I	G74D	R14T	F103H	A55V	7	1.09
M39I	K72R	V80A	R47G	G74D	7	1.09
M39I	K72R	V80A	R47G	A55V	7	1.09
M39I	K72R	V80A	R47G	R14T	7	1.09
M39I	K72R	V80A	R47G	F103H	7	1.09
M39I	K72R	V80A	A55V	G74D	7	1.09
M39I	K72R	V80A	A55V	R47G	7	1.09
M39I	K72R	V80A	A55V	R14T	7	1.09
M39I	K72R	V80A	A55V	F103H	7	1.09
M39I	K72R	V80A	R14T	G74D	7	1.09
M39I	K72R	V80A	R14T	R47G	7	1.09
M39I	K72R	V80A	R14T	A55V	7	1.09
M39I	K72R	V80A	R14T	F103H	7	1.09
M39I	K72R	V80A	F103H	G74D	7	1.09
M39I	K72R	V80A	F103H	R47G	7	1.09
M39I	K72R	V80A	F103H	A55V	7	1.09
M39I	K72R	V80A	F103H	R14T	7	1.09
M39I	K72R	R47G	A55V	G74D	7	1.09
M39I	K72R	R47G	A55V	V80A	7	1.09
M39I	K72R	R47G	A55V	R14T	7	1.09
M39I	K72R	R47G	A55V	F103H	7	1.09
M39I	K72R	R47G	R14T	G74D	7	1.09
M39I	K72R	R47G	R14T	V80A	7	1.09
M39I	K72R	R47G	R14T	A55V	7	1.09
M39I	K72R	R47G	R14T	F103H	7	1.09
M39I	K72R	R47G	F103H	G74D	7	1.09
M39I	K72R	R47G	F103H	V80A	7	1.09
M39I	K72R	R47G	F103H	A55V	7	1.09
M39I	K72R	R47G	F103H	R14T	7	1.09
M39I	K72R	A55V	R14T	G74D	7	1.09
M39I	K72R	A55V	R14T	V80A	7	1.09
M39I	K72R	A55V	R14T	R47G	7	1.09

M39I	K72R	A55V	R14T	F103H	7	1.09	G74D	K72R	V80A	R47G	M39I	7	1.09
M39I	K72R	A55V	F103H	G74D	7	1.09	G74D	K72R	V80A	R47G	A55V	7	1.09
M39I	K72R	A55V	F103H	V80A	7	1.09	G74D	K72R	V80A	R47G	R14T	7	1.09
M39I	K72R	A55V	F103H	R47G	7	1.09	G74D	K72R	V80A	R47G	F103H	7	1.09
M39I	K72R	A55V	F103H	R14T	7	1.09	G74D	K72R	V80A	A55V	M39I	7	1.09
M39I	K72R	R14T	F103H	G74D	7	1.09	G74D	K72R	V80A	A55V	R47G	7	1.09
M39I	K72R	R14T	F103H	V80A	7	1.09	G74D	K72R	V80A	A55V	R14T	7	1.09
M39I	K72R	R14T	F103H	R47G	7	1.09	G74D	K72R	V80A	A55V	F103H	7	1.09
M39I	K72R	R14T	F103H	A55V	7	1.09	G74D	K72R	V80A	R14T	M39I	7	1.09
M39I	V80A	R47G	A55V	G74D	7	1.09	G74D	K72R	V80A	R14T	R47G	7	1.09
M39I	V80A	R47G	A55V	K72R	7	1.09	G74D	K72R	V80A	R14T	A55V	7	1.09
M39I	V80A	R47G	A55V	R14T	7	1.09	G74D	K72R	V80A	R14T	F103H	7	1.09
M39I	V80A	R47G	A55V	F103H	7	1.09	G74D	K72R	V80A	F103H	M39I	7	1.09
M39I	V80A	R47G	R14T	G74D	7	1.09	G74D	K72R	V80A	F103H	R47G	7	1.09
M39I	V80A	R47G	R14T	K72R	7	1.09	G74D	K72R	V80A	F103H	A55V	7	1.09
M39I	V80A	R47G	R14T	A55V	7	1.09	G74D	K72R	V80A	F103H	R14T	7	1.09
M39I	V80A	R47G	R14T	F103H	7	1.09	G74D	K72R	R47G	A55V	M39I	7	1.09
M39I	V80A	R47G	F103H	G74D	7	1.09	G74D	K72R	R47G	A55V	V80A	7	1.09
M39I	V80A	R47G	F103H	K72R	7	1.09	G74D	K72R	R47G	A55V	R14T	7	1.09
M39I	V80A	R47G	F103H	A55V	7	1.09	G74D	K72R	R47G	A55V	F103H	7	1.09
M39I	V80A	R47G	F103H	R14T	7	1.09	G74D	K72R	R47G	R14T	M39I	7	1.09
M39I	V80A	A55V	R14T	G74D	7	1.09	G74D	K72R	R47G	R14T	V80A	7	1.09
M39I	V80A	A55V	R14T	K72R	7	1.09	G74D	K72R	R47G	R14T	A55V	7	1.09
M39I	V80A	A55V	R14T	R47G	7	1.09	G74D	K72R	R47G	R14T	F103H	7	1.09
M39I	V80A	A55V	R14T	F103H	7	1.09	G74D	K72R	R47G	F103H	M39I	7	1.09
M39I	V80A	A55V	F103H	G74D	7	1.09	G74D	K72R	R47G	F103H	V80A	7	1.09
M39I	V80A	A55V	F103H	K72R	7	1.09	G74D	K72R	R47G	F103H	A55V	7	1.09
M39I	V80A	A55V	F103H	R47G	7	1.09	G74D	K72R	R47G	F103H	R14T	7	1.09
M39I	V80A	A55V	F103H	R14T	7	1.09	G74D	K72R	A55V	R14T	M39I	7	1.09
M39I	V80A	R14T	F103H	G74D	7	1.09	G74D	K72R	A55V	R14T	V80A	7	1.09
M39I	V80A	R14T	F103H	K72R	7	1.09	G74D	K72R	A55V	R14T	R47G	7	1.09
M39I	V80A	R14T	F103H	R47G	7	1.09	G74D	K72R	A55V	R14T	F103H	7	1.09
M39I	V80A	R14T	F103H	A55V	7	1.09	G74D	K72R	A55V	F103H	M39I	7	1.09
M39I	R47G	A55V	R14T	G74D	7	1.09	G74D	K72R	A55V	F103H	V80A	7	1.09
M39I	R47G	A55V	R14T	K72R	7	1.09	G74D	K72R	A55V	F103H	R47G	7	1.09
M39I	R47G	A55V	R14T	V80A	7	1.09	G74D	K72R	A55V	F103H	R14T	7	1.09
M39I	R47G	A55V	R14T	F103H	7	1.09	G74D	K72R	R14T	F103H	M39I	7	1.09
M39I	R47G	A55V	F103H	G74D	7	1.09	G74D	K72R	R14T	F103H	V80A	7	1.09
M39I	R47G	A55V	F103H	K72R	7	1.09	G74D	K72R	R14T	F103H	R47G	7	1.09
M39I	R47G	A55V	F103H	V80A	7	1.09	G74D	K72R	R14T	F103H	A55V	7	1.09
M39I	R47G	A55V	F103H	R14T	7	1.09	G74D	V80A	R47G	A55V	M39I	7	1.09
M39I	R47G	R14T	F103H	G74D	7	1.09	G74D	V80A	R47G	A55V	K72R	7	1.09
M39I	R47G	R14T	F103H	K72R	7	1.09	G74D	V80A	R47G	A55V	R14T	7	1.09
M39I	R47G	R14T	F103H	V80A	7	1.09	G74D	V80A	R47G	A55V	F103H	7	1.09
M39I	R47G	R14T	F103H	A55V	7	1.09	G74D	V80A	R47G	R14T	M39I	7	1.09
M39I	A55V	R14T	F103H	G74D	7	1.09	G74D	V80A	R47G	R14T	K72R	7	1.09
M39I	A55V	R14T	F103H	K72R	7	1.09	G74D	V80A	R47G	R14T	A55V	7	1.09
M39I	A55V	R14T	F103H	V80A	7	1.09	G74D	V80A	R47G	R14T	F103H	7	1.09
M39I	A55V	R14T	F103H	R47G	7	1.09	G74D	V80A	R47G	F103H	M39I	7	1.09

G74D	V80A	R47G	F103H	K72R	7	1.09
G74D	V80A	R47G	F103H	A55V	7	1.09
G74D	V80A	R47G	F103H	R14T	7	1.09
G74D	V80A	A55V	R14T	M39I	7	1.09
G74D	V80A	A55V	R14T	K72R	7	1.09
G74D	V80A	A55V	R14T	R47G	7	1.09
G74D	V80A	A55V	R14T	F103H	7	1.09
G74D	V80A	A55V	F103H	M39I	7	1.09
G74D	V80A	A55V	F103H	K72R	7	1.09
G74D	V80A	A55V	F103H	R47G	7	1.09
G74D	V80A	A55V	F103H	R14T	7	1.09
G74D	V80A	R14T	F103H	M39I	7	1.09
G74D	V80A	R14T	F103H	K72R	7	1.09
G74D	V80A	R14T	F103H	R47G	7	1.09
G74D	V80A	R14T	F103H	A55V	7	1.09
G74D	R47G	A55V	R14T	M39I	7	1.09
G74D	R47G	A55V	R14T	K72R	7	1.09
G74D	R47G	A55V	R14T	V80A	7	1.09
G74D	R47G	A55V	R14T	F103H	7	1.09
G74D	R47G	A55V	F103H	M39I	7	1.09
G74D	R47G	A55V	F103H	K72R	7	1.09
G74D	R47G	A55V	F103H	V80A	7	1.09
G74D	R47G	A55V	F103H	R14T	7	1.09
G74D	R47G	R14T	F103H	M39I	7	1.09
G74D	R47G	R14T	F103H	K72R	7	1.09
G74D	R47G	R14T	F103H	V80A	7	1.09
G74D	R47G	R14T	F103H	A55V	7	1.09
G74D	A55V	R14T	F103H	M39I	7	1.09
G74D	A55V	R14T	F103H	K72R	7	1.09
G74D	A55V	R14T	F103H	V80A	7	1.09
G74D	A55V	R14T	F103H	R47G	7	1.09
K72R	V80A	R47G	A55V	M39I	7	1.09
K72R	V80A	R47G	A55V	G74D	7	1.09
K72R	V80A	R47G	A55V	R14T	7	1.09
K72R	V80A	R47G	A55V	F103H	7	1.09
K72R	V80A	R47G	R14T	M39I	7	1.09
K72R	V80A	R47G	R14T	G74D	7	1.09
K72R	V80A	R47G	R14T	A55V	7	1.09
K72R	V80A	R47G	R14T	F103H	7	1.09
K72R	V80A	R47G	F103H	M39I	7	1.09
K72R	V80A	R47G	F103H	G74D	7	1.09
K72R	V80A	R47G	F103H	A55V	7	1.09
K72R	V80A	R47G	F103H	R14T	7	1.09
K72R	V80A	A55V	R14T	M39I	7	1.09
K72R	V80A	A55V	R14T	G74D	7	1.09
K72R	V80A	A55V	R14T	R47G	7	1.09
K72R	V80A	A55V	R14T	F103H	7	1.09
K72R	V80A	A55V	F103H	M39I	7	1.09
K72R	V80A	A55V	F103H	G74D	7	1.09
K72R	V80A	A55V	F103H	R47G	7	1.09
K72R	V80A	A55V	F103H	R14T	7	1.09
K72R	V80A	R14T	F103H	M39I	7	1.09
K72R	V80A	R14T	F103H	G74D	7	1.09
K72R	V80A	R14T	F103H	K72R	7	1.09
K72R	V80A	R14T	F103H	A55V	7	1.09
K72R	V80A	R14T	F103H	R47G	7	1.09
K72R	R47G	A55V	R14T	M39I	7	1.09
K72R	R47G	A55V	R14T	G74D	7	1.09
K72R	R47G	A55V	R14T	V80A	7	1.09
K72R	R47G	A55V	R14T	F103H	7	1.09
K72R	R47G	A55V	F103H	M39I	7	1.09
K72R	R47G	A55V	F103H	G74D	7	1.09
K72R	R47G	A55V	F103H	V80A	7	1.09
K72R	R47G	A55V	F103H	R14T	7	1.09
K72R	R47G	R14T	F103H	M39I	7	1.09
K72R	R47G	R14T	F103H	G74D	7	1.09
K72R	R47G	R14T	F103H	V80A	7	1.09
K72R	R47G	R14T	F103H	A55V	7	1.09
K72R	A55V	R14T	F103H	M39I	7	1.09
K72R	A55V	R14T	F103H	G74D	7	1.09
K72R	A55V	R14T	F103H	V80A	7	1.09
K72R	A55V	R14T	F103H	R47G	7	1.09
V80A	R47G	A55V	R14T	M39I	7	1.09
V80A	R47G	A55V	R14T	G74D	7	1.09
V80A	R47G	A55V	R14T	K72R	7	1.09
V80A	R47G	A55V	R14T	F103H	7	1.09
V80A	R47G	A55V	F103H	M39I	7	1.09
V80A	R47G	A55V	F103H	G74D	7	1.09
V80A	R47G	A55V	F103H	K72R	7	1.09
V80A	R47G	A55V	F103H	R14T	7	1.09
V80A	R47G	R14T	F103H	M39I	7	1.09
V80A	R47G	R14T	F103H	G74D	7	1.09
V80A	R47G	R14T	F103H	K72R	7	1.09
V80A	R47G	R14T	F103H	A55V	7	1.09
V80A	A55V	R14T	F103H	M39I	7	1.09
V80A	A55V	R14T	F103H	G74D	7	1.09
V80A	A55V	R14T	F103H	K72R	7	1.09
V80A	A55V	R14T	F103H	R47G	7	1.09
D36S	T100A	T68N	L80S	D90E	7	1.09
D36S	T100A	T68N	D90E	L80S	7	1.09
D36S	T100A	L80S	D90E	T68N	7	1.09
D36S	T68N	L80S	D90E	T100A	7	1.09
R47G	A55V	R14T	F103H	M39I	7	1.09
R47G	A55V	R14T	F103H	G74D	7	1.09
R47G	A55V	R14T	F103H	K72R	7	1.09
R47G	A55V	R14T	F103H	V80A	7	1.09
T100A	T68N	L80S	D90E	D36S	7	1.09

Table S2. List of V^L top correlated mutations.Double correlated mutations

Mutations		C	f(%)
V2L	D1E	7	1.35
N114H	A74V	6	1.15
V2I	T7S	5	0.96
T7S	M4L	4	0.77
V2I	M4L	4	0.77
V114F	V101I	4	0.77
T108S	V101I	4	0.77
D1E	T7S	4	0.77
V2L	T7S	4	0.77
L3V	V2I	4	0.77
Y55H	S109N	4	0.77
N114H	Y55H	4	0.77
A56G	Y55H	4	0.77
A57V	K51R	4	0.77
S107I	K51R	4	0.77
S27R	*116Y	4	0.77
S107I	A57V	4	0.77
S107T	M4L	3	0.58
V3Q	M4L	3	0.58
Q3V	M4L	3	0.58
Q3E	M4L	3	0.58
T7S	V101I	3	0.58
S28T	V101I	3	0.58
V2I	V101I	3	0.58
S107T	V101I	3	0.58
Y103F	V101I	3	0.58
S32N	V101I	3	0.58
V114F	S28T	3	0.58
D1E	S92N	3	0.58
S26T	S92N	3	0.58
S107T	Y42F	3	0.58
L3V	D1E	3	0.58
L4M	S37N	3	0.58
A74V	S37N	3	0.58
Y87F	S37N	3	0.58
M94V	S107T	3	0.58
Y103F	S109N	3	0.58
A74V	S109N	3	0.58
Y87F	S109N	3	0.58
N114H	S109N	3	0.58
Y38N	S109N	3	0.58
A56G	S109N	3	0.58
I54L	S109N	3	0.58
Y55H	Y103F	3	0.58
A74V	Y103F	3	0.58
Y87F	Y103F	3	0.58
N114H	Y103F	3	0.58
Y38N	Y103F	3	0.58
A56G	Y103F	3	0.58
I54L	Y103F	3	0.58
A74V	Y55H	3	0.58
Y87F	Y55H	3	0.58
Y38N	Y55H	3	0.58
I54L	Y55H	3	0.58
H31Y	N34D	3	0.58
L4M	S7T	3	0.58
Q3V	S7T	3	0.58
I2L	L4M	3	0.58
T90S	S22T	3	0.58
L30F	S5T	3	0.58
V2L	L3V	3	0.58
Y87F	A74V	3	0.58
Y38N	A74V	3	0.58
A56G	A74V	3	0.58
I54L	A74V	3	0.58
S32N	K51Q	3	0.58
N114H	Y87F	3	0.58
Y38N	Y87F	3	0.58
A56G	Y87F	3	0.58
I54L	Y87F	3	0.58
R66K	N40S	3	0.58
Y38N	N114H	3	0.58
A56G	N114H	3	0.58
I54L	N114H	3	0.58
S14T	K3V	3	0.58
A56G	Y38N	3	0.58
I54L	Y38N	3	0.58
I54L	A56G	3	0.58

Triple correlated mutations

Mutations			C	f(%)
A57V	K51R	S107I	4	0.77
T7S	M4L	V2I	3	0.58
S28T	V101I	V114F	3	0.58
D1E	T7S	V2L	3	0.58
L3V	D1E	V2L	3	0.58
Y103F	S109N	Y55H	3	0.58
Y103F	S109N	A74V	3	0.58
Y103F	S109N	Y87F	3	0.58
Y103F	S109N	N114H	3	0.58
Y103F	S109N	Y38N	3	0.58
Y103F	S109N	A56G	3	0.58
Y103F	S109N	I54L	3	0.58

Y55H	S109N	A74V	3	0.58
Y55H	S109N	Y87F	3	0.58
Y55H	S109N	N114H	3	0.58
Y55H	S109N	Y38N	3	0.58
Y55H	S109N	A56G	3	0.58
Y55H	S109N	I54L	3	0.58
A74V	S109N	Y87F	3	0.58
A74V	S109N	N114H	3	0.58
A74V	S109N	Y38N	3	0.58
A74V	S109N	A56G	3	0.58
A74V	S109N	I54L	3	0.58
Y87F	S109N	N114H	3	0.58
Y87F	S109N	Y38N	3	0.58
Y87F	S109N	A56G	3	0.58
Y87F	S109N	I54L	3	0.58
N114H	S109N	Y38N	3	0.58
N114H	S109N	A56G	3	0.58
N114H	S109N	I54L	3	0.58
Y38N	S109N	A56G	3	0.58
Y38N	S109N	I54L	3	0.58
A56G	S109N	I54L	3	0.58
Y55H	Y103F	A74V	3	0.58
Y55H	Y103F	Y87F	3	0.58
Y55H	Y103F	N114H	3	0.58
Y55H	Y103F	Y38N	3	0.58
Y55H	Y103F	A56G	3	0.58
Y55H	Y103F	I54L	3	0.58
A74V	Y103F	Y87F	3	0.58
A74V	Y103F	N114H	3	0.58
A74V	Y103F	Y38N	3	0.58
A74V	Y103F	A56G	3	0.58
A74V	Y103F	I54L	3	0.58
Y87F	Y103F	N114H	3	0.58
Y87F	Y103F	Y38N	3	0.58
Y87F	Y103F	A56G	3	0.58
Y87F	Y103F	I54L	3	0.58
N114H	Y103F	Y38N	3	0.58
N114H	Y103F	A56G	3	0.58
N114H	Y103F	I54L	3	0.58
Y38N	Y103F	A56G	3	0.58
Y38N	Y103F	I54L	3	0.58
A56G	Y103F	I54L	3	0.58
A74V	Y55H	Y87F	3	0.58
A74V	Y55H	N114H	3	0.58
A74V	Y55H	Y38N	3	0.58
A74V	Y55H	A56G	3	0.58
A74V	Y55H	I54L	3	0.58
Y87F	Y55H	N114H	3	0.58
Y87F	Y55H	Y38N	3	0.58
Y87F	Y55H	A56G	3	0.58
Y87F	Y55H	I54L	3	0.58
N114H	Y55H	Y38N	3	0.58
N114H	Y55H	A56G	3	0.58
N114H	Y55H	I54L	3	0.58
Y38N	Y55H	A56G	3	0.58
Y38N	Y55H	I54L	3	0.58
A56G	Y55H	I54L	3	0.58
Y87F	A74V	N114H	3	0.58
Y87F	A74V	Y38N	3	0.58
Y87F	A74V	A56G	3	0.58
Y87F	A74V	I54L	3	0.58
N114H	A74V	Y38N	3	0.58
N114H	A74V	A56G	3	0.58
N114H	A74V	I54L	3	0.58
Y38N	A74V	A56G	3	0.58
Y38N	A74V	I54L	3	0.58
A56G	A74V	I54L	3	0.58
N114H	Y87F	Y38N	3	0.58
N114H	Y87F	A56G	3	0.58
N114H	Y87F	I54L	3	0.58
Y38N	Y87F	A56G	3	0.58
Y38N	Y87F	I54L	3	0.58
A56G	Y87F	I54L	3	0.58
Y38N	N114H	A56G	3	0.58
Y38N	N114H	I54L	3	0.58
A56G	N114H	I54L	3	0.58
A56G	Y38N	I54L	3	0.58
V2I	M4L	V3Q	2	0.38
V2I	M4L	R93S	2	0.38
V3Q	M4L	R93S	2	0.38
N36K	M4L	L116R	2	0.38
N36K	M4L	Q3E	2	0.38
N36K	M4L	S77R	2	0.38
L116R	M4L	Q3E	2	0.38
L116R	M4L	S77R	2	0.38
Q3E	M4L	S77R	2	0.38
T7S	V101I	V2L	2	0.38
T7S	V101I	T108S	2	0.38
S28T	V101I	K51Q	2	0.38
S28T	V101I	S32N	2	0.38
V2I	V101I	S32N	2	0.38
V2L	V101I	T108S	2	0.38
K51Q	V101I	S32N	2	0.38
K51Q	V101I	V114F	2	0.38
S32N	V101I	V114F	2	0.38
V2I	T7S	E97A	2	0.38
V2L	T7S	T108S	2	0.38
K51Q	S28T	S32N	2	0.38

K51Q	S28T	V114F	2	0.38
S32N	S28T	V114F	2	0.38
Y55F	S92N	S26T	2	0.38
Y55F	S92N	K51Q	2	0.38
Y55F	S92N	V15I	2	0.38
S26T	S92N	K51Q	2	0.38
S26T	S92N	V15I	2	0.38
S5T	S92N	L30F	2	0.38
S5T	S92N	S93G	2	0.38
L30F	S92N	S93G	2	0.38
K51Q	S92N	V15I	2	0.38
V3Q	V2I	R93S	2	0.38
V2L	D1E	S107G	2	0.38
S109N	S37N	Y103F	2	0.38
S109N	S37N	Y55H	2	0.38
S109N	S37N	A74V	2	0.38
S109N	S37N	Y87F	2	0.38
S109N	S37N	N114H	2	0.38
S109N	S37N	Y38N	2	0.38
S109N	S37N	A56G	2	0.38
S109N	S37N	I54L	2	0.38
Y103F	S37N	Y55H	2	0.38
Y103F	S37N	A74V	2	0.38
Y103F	S37N	Y87F	2	0.38
Y103F	S37N	N114H	2	0.38
Y103F	S37N	Y38N	2	0.38
Y103F	S37N	A56G	2	0.38
Y103F	S37N	I54L	2	0.38
Y55H	S37N	A74V	2	0.38
Y55H	S37N	Y87F	2	0.38
Y55H	S37N	N114H	2	0.38
Y55H	S37N	Y38N	2	0.38
Y55H	S37N	A56G	2	0.38
Y55H	S37N	I54L	2	0.38
A74V	S37N	Y87F	2	0.38
A74V	S37N	N114H	2	0.38
A74V	S37N	Y38N	2	0.38
A74V	S37N	A56G	2	0.38
A74V	S37N	I54L	2	0.38
Y87F	S37N	N114H	2	0.38
Y87F	S37N	Y38N	2	0.38
Y87F	S37N	A56G	2	0.38
Y87F	S37N	I54L	2	0.38
N114H	S37N	Y38N	2	0.38
N114H	S37N	A56G	2	0.38
N114H	S37N	I54L	2	0.38
Y38N	S37N	A56G	2	0.38
Y38N	S37N	I54L	2	0.38
A56G	S37N	I54L	2	0.38
S22T	S109N	S93G	2	0.38
S22T	S109N	T90S	2	0.38
S93G	S109N	T90S	2	0.38
S26T	Y55F	K51Q	2	0.38
S26T	Y55F	V15I	2	0.38
K51Q	Y55F	V15I	2	0.38
S109T	P116L	A100G	2	0.38
A74V	P116L	N114H	2	0.38
L4M	S7T	I2L	2	0.38
K51Q	S26T	V15I	2	0.38
S93G	S22T	T90S	2	0.38
L30F	S5T	S93G	2	0.38
I29L	V2L	N34S	2	0.38
N40S	Q3V	R66K	2	0.38
D86H	*116Y	S27R	2	0.38
L116R	N36K	Q3E	2	0.38
L116R	N36K	S77R	2	0.38
Q3E	N36K	S77R	2	0.38
S32N	K51Q	V114F	2	0.38
Q3E	L116R	S77R	2	0.38
S14T	K3V	N36Y	2	0.38
A57T	S28R	T5I	2	0.38

Quadruple correlated mutations

Mutations				C	f(%)
T35S	I101V	L21I	T96S	15	2.88
M39I	G74D	K72R	V80A	7	1.35
M39I	G74D	K72R	R47G	7	1.35
M39I	G74D	K72R	A55V	7	1.35
M39I	G74D	K72R	R14T	7	1.35
M39I	G74D	K72R	F103H	7	1.35
M39I	G74D	V80A	R47G	7	1.35
M39I	G74D	V80A	A55V	7	1.35
M39I	G74D	V80A	R14T	7	1.35
M39I	G74D	V80A	F103H	7	1.35
M39I	G74D	R47G	A55V	7	1.35
M39I	G74D	R47G	R14T	7	1.35
M39I	G74D	R47G	F103H	7	1.35
M39I	G74D	A55V	R14T	7	1.35
M39I	G74D	A55V	F103H	7	1.35
M39I	G74D	R14T	F103H	7	1.35
M39I	K72R	V80A	R47G	7	1.35
M39I	K72R	V80A	A55V	7	1.35
M39I	K72R	V80A	R14T	7	1.35
M39I	K72R	V80A	F103H	7	1.35
M39I	K72R	R47G	A55V	7	1.35
M39I	K72R	R47G	R14T	7	1.35
M39I	K72R	R47G	F103H	7	1.35

M39I	K72R	A55V	R14T	7	1.35
M39I	K72R	A55V	F103H	7	1.35
M39I	K72R	R14T	F103H	7	1.35
M39I	V80A	R47G	A55V	7	1.35
M39I	V80A	R47G	R14T	7	1.35
M39I	V80A	R47G	F103H	7	1.35
M39I	V80A	A55V	R14T	7	1.35
M39I	V80A	A55V	F103H	7	1.35
M39I	V80A	R14T	F103H	7	1.35
M39I	R47G	A55V	R14T	7	1.35
M39I	R47G	A55V	F103H	7	1.35
M39I	R47G	R14T	F103H	7	1.35
M39I	A55V	R14T	F103H	7	1.35
G74D	K72R	V80A	R47G	7	1.35
G74D	K72R	V80A	A55V	7	1.35
G74D	K72R	V80A	R14T	7	1.35
G74D	K72R	V80A	F103H	7	1.35
G74D	K72R	R47G	A55V	7	1.35
G74D	K72R	R47G	R14T	7	1.35
G74D	K72R	R47G	F103H	7	1.35
G74D	K72R	A55V	R14T	7	1.35
G74D	K72R	A55V	F103H	7	1.35
G74D	K72R	R14T	F103H	7	1.35
G74D	V80A	R47G	A55V	7	1.35
G74D	V80A	R47G	R14T	7	1.35
G74D	V80A	R47G	F103H	7	1.35
G74D	V80A	A55V	R14T	7	1.35
G74D	V80A	A55V	F103H	7	1.35
G74D	V80A	R14T	F103H	7	1.35
G74D	R47G	A55V	R14T	7	1.35
G74D	R47G	A55V	F103H	7	1.35
G74D	R47G	R14T	F103H	7	1.35
G74D	A55V	R14T	F103H	7	1.35
K72R	V80A	R47G	A55V	7	1.35
K72R	V80A	R47G	R14T	7	1.35
K72R	V80A	R47G	F103H	7	1.35
K72R	V80A	A55V	R14T	7	1.35
K72R	V80A	A55V	F103H	7	1.35
K72R	V80A	R14T	F103H	7	1.35
K72R	R47G	A55V	R14T	7	1.35
K72R	R47G	A55V	F103H	7	1.35
K72R	R47G	R14T	F103H	7	1.35
K72R	A55V	R14T	F103H	7	1.35
I101V	L21I	T96S	G36S	7	1.35
V80A	R47G	A55V	R14T	7	1.35
V80A	R47G	A55V	F103H	7	1.35
V80A	R47G	R14T	F103H	7	1.35
V80A	A55V	R14T	F103H	7	1.35
D36S	T100A	T68N	L80S	7	1.35
D36S	T100A	T68N	D90E	7	1.35
D36S	T100A	L80S	D90E	7	1.35
D36S	T68N	L80S	D90E	7	1.35
R47G	A55V	R14T	F103H	7	1.35

Quintuple correlated mutations

Mutations					C	f(%)
M39I	G74D	K72R	V80A	R47G	7	1.35
M39I	G74D	K72R	V80A	A55V	7	1.35
M39I	G74D	K72R	V80A	R14T	7	1.35
M39I	G74D	K72R	V80A	F103H	7	1.35
M39I	G74D	K72R	R47G	V80A	7	1.35
M39I	G74D	K72R	R47G	A55V	7	1.35
M39I	G74D	K72R	R47G	R14T	7	1.35
M39I	G74D	K72R	R47G	F103H	7	1.35
M39I	G74D	K72R	A55V	V80A	7	1.35
M39I	G74D	K72R	A55V	R47G	7	1.35
M39I	G74D	K72R	A55V	R14T	7	1.35
M39I	G74D	K72R	A55V	F103H	7	1.35
M39I	G74D	K72R	R14T	V80A	7	1.35
M39I	G74D	K72R	R14T	R47G	7	1.35
M39I	G74D	K72R	R14T	A55V	7	1.35
M39I	G74D	K72R	R14T	F103H	7	1.35
M39I	G74D	K72R	F103H	V80A	7	1.35
M39I	G74D	K72R	F103H	R47G	7	1.35
M39I	G74D	K72R	F103H	A55V	7	1.35
M39I	G74D	K72R	F103H	R14T	7	1.35
M39I	G74D	V80A	R47G	K72R	7	1.35
M39I	G74D	V80A	R47G	A55V	7	1.35
M39I	G74D	V80A	R47G	R14T	7	1.35
M39I	G74D	V80A	R47G	F103H	7	1.35
M39I	G74D	V80A	A55V	K72R	7	1.35
M39I	G74D	V80A	A55V	R47G	7	1.35
M39I	G74D	V80A	A55V	R14T	7	1.35
M39I	G74D	V80A	A55V	F103H	7	1.35
M39I	G74D	V80A	R14T	K72R	7	1.35
M39I	G74D	V80A	R14T	R47G	7	1.35
M39I	G74D	V80A	R14T	A55V	7	1.35
M39I	G74D	V80A	R14T	F103H	7	1.35
M39I	G74D	V80A	F103H	K72R	7	1.35
M39I	G74D	V80A	F103H	R47G	7	1.35
M39I	G74D	V80A	F103H	A55V	7	1.35
M39I	G74D	V80A	F103H	R14T	7	1.35
M39I	G74D	R47G	A55V	K72R	7	1.35
M39I	G74D	R47G	A55V	V80A	7	1.35
M39I	G74D	R47G	A55V	R14T	7	1.35
M39I	G74D	R47G	A55V	F103H	7	1.35
M39I	G74D	R47G	R14T	K72R	7	1.35

M39I	G74D	R47G	R14T	V80A	7	1.35	M39I	K72R	A55V	R14T	R47G	7	1.35
M39I	G74D	R47G	R14T	A55V	7	1.35	M39I	K72R	A55V	R14T	F103H	7	1.35
M39I	G74D	R47G	R14T	F103H	7	1.35	M39I	K72R	A55V	F103H	G74D	7	1.35
M39I	G74D	R47G	F103H	K72R	7	1.35	M39I	K72R	A55V	F103H	V80A	7	1.35
M39I	G74D	R47G	F103H	V80A	7	1.35	M39I	K72R	A55V	F103H	R47G	7	1.35
M39I	G74D	R47G	F103H	A55V	7	1.35	M39I	K72R	A55V	F103H	R14T	7	1.35
M39I	G74D	R47G	F103H	R14T	7	1.35	M39I	K72R	R14T	F103H	G74D	7	1.35
M39I	G74D	A55V	R14T	K72R	7	1.35	M39I	K72R	R14T	F103H	V80A	7	1.35
M39I	G74D	A55V	R14T	V80A	7	1.35	M39I	K72R	R14T	F103H	R47G	7	1.35
M39I	G74D	A55V	R14T	R47G	7	1.35	M39I	K72R	R14T	F103H	A55V	7	1.35
M39I	G74D	A55V	R14T	F103H	7	1.35	M39I	V80A	R47G	A55V	G74D	7	1.35
M39I	G74D	A55V	F103H	K72R	7	1.35	M39I	V80A	R47G	A55V	K72R	7	1.35
M39I	G74D	A55V	F103H	V80A	7	1.35	M39I	V80A	R47G	A55V	R14T	7	1.35
M39I	G74D	A55V	F103H	R47G	7	1.35	M39I	V80A	R47G	A55V	F103H	7	1.35
M39I	G74D	A55V	F103H	R14T	7	1.35	M39I	V80A	R47G	R14T	G74D	7	1.35
M39I	G74D	R14T	F103H	K72R	7	1.35	M39I	V80A	R47G	R14T	K72R	7	1.35
M39I	G74D	R14T	F103H	V80A	7	1.35	M39I	V80A	R47G	R14T	A55V	7	1.35
M39I	G74D	R14T	F103H	R47G	7	1.35	M39I	V80A	R47G	R14T	F103H	7	1.35
M39I	G74D	R14T	F103H	A55V	7	1.35	M39I	V80A	R47G	F103H	G74D	7	1.35
M39I	K72R	V80A	R47G	G74D	7	1.35	M39I	V80A	R47G	F103H	K72R	7	1.35
M39I	K72R	V80A	R47G	A55V	7	1.35	M39I	V80A	R47G	F103H	A55V	7	1.35
M39I	K72R	V80A	R47G	R14T	7	1.35	M39I	V80A	R47G	F103H	R14T	7	1.35
M39I	K72R	V80A	R47G	F103H	7	1.35	M39I	V80A	A55V	R14T	G74D	7	1.35
M39I	K72R	V80A	A55V	G74D	7	1.35	M39I	V80A	A55V	R14T	K72R	7	1.35
M39I	K72R	V80A	A55V	R47G	7	1.35	M39I	V80A	A55V	R14T	R47G	7	1.35
M39I	K72R	V80A	A55V	R14T	7	1.35	M39I	V80A	A55V	R14T	F103H	7	1.35
M39I	K72R	V80A	A55V	F103H	7	1.35	M39I	V80A	A55V	F103H	G74D	7	1.35
M39I	K72R	V80A	R14T	G74D	7	1.35	M39I	V80A	A55V	F103H	K72R	7	1.35
M39I	K72R	V80A	R14T	R47G	7	1.35	M39I	V80A	A55V	F103H	R47G	7	1.35
M39I	K72R	V80A	R14T	A55V	7	1.35	M39I	V80A	A55V	F103H	R14T	7	1.35
M39I	K72R	V80A	R14T	F103H	7	1.35	M39I	V80A	R14T	F103H	G74D	7	1.35
M39I	K72R	V80A	F103H	G74D	7	1.35	M39I	V80A	R14T	F103H	K72R	7	1.35
M39I	K72R	V80A	F103H	R47G	7	1.35	M39I	V80A	R14T	F103H	R47G	7	1.35
M39I	K72R	V80A	F103H	A55V	7	1.35	M39I	V80A	R14T	F103H	A55V	7	1.35
M39I	K72R	V80A	F103H	R14T	7	1.35	M39I	R47G	A55V	R14T	G74D	7	1.35
M39I	K72R	R47G	A55V	G74D	7	1.35	M39I	R47G	A55V	R14T	K72R	7	1.35
M39I	K72R	R47G	A55V	V80A	7	1.35	M39I	R47G	A55V	R14T	V80A	7	1.35
M39I	K72R	R47G	A55V	R14T	7	1.35	M39I	R47G	A55V	R14T	F103H	7	1.35
M39I	K72R	R47G	A55V	F103H	7	1.35	M39I	R47G	A55V	F103H	G74D	7	1.35
M39I	K72R	R47G	R14T	G74D	7	1.35	M39I	R47G	A55V	F103H	K72R	7	1.35
M39I	K72R	R47G	R14T	V80A	7	1.35	M39I	R47G	A55V	F103H	V80A	7	1.35
M39I	K72R	R47G	R14T	A55V	7	1.35	M39I	R47G	A55V	F103H	R14T	7	1.35
M39I	K72R	R47G	R14T	F103H	7	1.35	M39I	R47G	R14T	F103H	G74D	7	1.35
M39I	K72R	R47G	F103H	G74D	7	1.35	M39I	R47G	R14T	F103H	K72R	7	1.35
M39I	K72R	R47G	F103H	V80A	7	1.35	M39I	R47G	R14T	F103H	V80A	7	1.35
M39I	K72R	R47G	F103H	A55V	7	1.35	M39I	R47G	R14T	F103H	A55V	7	1.35
M39I	K72R	R47G	F103H	R14T	7	1.35	M39I	A55V	R14T	F103H	G74D	7	1.35
M39I	K72R	A55V	R14T	G74D	7	1.35	M39I	A55V	R14T	F103H	K72R	7	1.35
M39I	K72R	A55V	R14T	V80A	7	1.35	M39I	A55V	R14T	F103H	V80A	7	1.35

M39I	A55V	R14T	F103H	R47G	7	1.35
G74D	K72R	V80A	R47G	M39I	7	1.35
G74D	K72R	V80A	R47G	A55V	7	1.35
G74D	K72R	V80A	R47G	R14T	7	1.35
G74D	K72R	V80A	R47G	F103H	7	1.35
G74D	K72R	V80A	A55V	M39I	7	1.35
G74D	K72R	V80A	A55V	R47G	7	1.35
G74D	K72R	V80A	A55V	R14T	7	1.35
G74D	K72R	V80A	A55V	F103H	7	1.35
G74D	K72R	V80A	R14T	M39I	7	1.35
G74D	K72R	V80A	R14T	R47G	7	1.35
G74D	K72R	V80A	R14T	A55V	7	1.35
G74D	K72R	V80A	R14T	F103H	7	1.35
G74D	K72R	V80A	F103H	M39I	7	1.35
G74D	K72R	V80A	F103H	R47G	7	1.35
G74D	K72R	V80A	F103H	A55V	7	1.35
G74D	K72R	V80A	F103H	R14T	7	1.35
G74D	K72R	R47G	A55V	M39I	7	1.35
G74D	K72R	R47G	A55V	V80A	7	1.35
G74D	K72R	R47G	A55V	R14T	7	1.35
G74D	K72R	R47G	A55V	F103H	7	1.35
G74D	K72R	R47G	R14T	M39I	7	1.35
G74D	K72R	R47G	R14T	V80A	7	1.35
G74D	K72R	R47G	R14T	A55V	7	1.35
G74D	K72R	R47G	R14T	F103H	7	1.35
G74D	K72R	R47G	F103H	M39I	7	1.35
G74D	K72R	R47G	F103H	V80A	7	1.35
G74D	K72R	R47G	F103H	A55V	7	1.35
G74D	K72R	R47G	F103H	R14T	7	1.35
G74D	K72R	A55V	R14T	M39I	7	1.35
G74D	K72R	A55V	R14T	V80A	7	1.35
G74D	K72R	A55V	R14T	R47G	7	1.35
G74D	K72R	A55V	R14T	F103H	7	1.35
G74D	K72R	A55V	F103H	M39I	7	1.35
G74D	K72R	A55V	F103H	V80A	7	1.35
G74D	K72R	A55V	F103H	R47G	7	1.35
G74D	K72R	A55V	F103H	R14T	7	1.35
G74D	K72R	R14T	F103H	M39I	7	1.35
G74D	K72R	R14T	F103H	V80A	7	1.35
G74D	K72R	R14T	F103H	R47G	7	1.35
G74D	K72R	R14T	F103H	A55V	7	1.35
G74D	V80A	R47G	A55V	M39I	7	1.35
G74D	V80A	R47G	A55V	K72R	7	1.35
G74D	V80A	R47G	A55V	R14T	7	1.35
G74D	V80A	R47G	A55V	F103H	7	1.35
G74D	V80A	R47G	R14T	M39I	7	1.35
G74D	V80A	R47G	R14T	K72R	7	1.35
G74D	V80A	R47G	R14T	A55V	7	1.35
G74D	V80A	R47G	R14T	F103H	7	1.35
G74D	V80A	R47G	F103H	M39I	7	1.35
G74D	V80A	R47G	F103H	K72R	7	1.35
G74D	V80A	R47G	F103H	A55V	7	1.35
G74D	V80A	R47G	F103H	R14T	7	1.35
G74D	V80A	A55V	R14T	M39I	7	1.35
G74D	V80A	A55V	R14T	K72R	7	1.35
G74D	V80A	A55V	R14T	R47G	7	1.35
G74D	V80A	A55V	R14T	F103H	7	1.35
G74D	V80A	A55V	F103H	M39I	7	1.35
G74D	V80A	A55V	F103H	K72R	7	1.35
G74D	V80A	A55V	F103H	R47G	7	1.35
G74D	V80A	A55V	F103H	R14T	7	1.35
G74D	V80A	R14T	F103H	M39I	7	1.35
G74D	V80A	R14T	F103H	K72R	7	1.35
G74D	V80A	R14T	F103H	R47G	7	1.35
G74D	V80A	R14T	F103H	A55V	7	1.35
G74D	R47G	A55V	R14T	M39I	7	1.35
G74D	R47G	A55V	R14T	K72R	7	1.35
G74D	R47G	A55V	R14T	V80A	7	1.35
G74D	R47G	A55V	R14T	F103H	7	1.35
G74D	R47G	A55V	F103H	M39I	7	1.35
G74D	R47G	A55V	F103H	K72R	7	1.35
G74D	R47G	A55V	F103H	V80A	7	1.35
G74D	R47G	A55V	F103H	R14T	7	1.35
G74D	R47G	R14T	F103H	M39I	7	1.35
G74D	R47G	R14T	F103H	K72R	7	1.35
G74D	R47G	R14T	F103H	V80A	7	1.35
G74D	R47G	R14T	F103H	A55V	7	1.35
G74D	A55V	R14T	F103H	M39I	7	1.35
G74D	A55V	R14T	F103H	K72R	7	1.35
G74D	A55V	R14T	F103H	V80A	7	1.35
G74D	A55V	R14T	F103H	R47G	7	1.35
K72R	V80A	R47G	A55V	M39I	7	1.35
K72R	V80A	R47G	A55V	G74D	7	1.35
K72R	V80A	R47G	A55V	R14T	7	1.35
K72R	V80A	R47G	A55V	F103H	7	1.35
K72R	V80A	R47G	R14T	M39I	7	1.35
K72R	V80A	R47G	R14T	G74D	7	1.35
K72R	V80A	R47G	R14T	A55V	7	1.35
K72R	V80A	R47G	R14T	F103H	7	1.35
K72R	V80A	R47G	F103H	M39I	7	1.35
K72R	V80A	R47G	F103H	G74D	7	1.35
K72R	V80A	R47G	F103H	A55V	7	1.35
K72R	V80A	R47G	F103H	R14T	7	1.35
K72R	V80A	A55V	R14T	M39I	7	1.35
K72R	V80A	A55V	R14T	G74D	7	1.35
K72R	V80A	A55V	R14T	R47G	7	1.35
K72R	V80A	A55V	R14T	F103H	7	1.35
K72R	V80A	A55V	F103H	M39I	7	1.35

K72R	V80A	A55V	F103H	G74D	7	1.35
K72R	V80A	A55V	F103H	R47G	7	1.35
K72R	V80A	A55V	F103H	R14T	7	1.35
K72R	V80A	R14T	F103H	M39I	7	1.35
K72R	V80A	R14T	F103H	G74D	7	1.35
K72R	V80A	R14T	F103H	R47G	7	1.35
K72R	V80A	R14T	F103H	A55V	7	1.35
K72R	R47G	A55V	R14T	M39I	7	1.35
K72R	R47G	A55V	R14T	G74D	7	1.35
K72R	R47G	A55V	R14T	V80A	7	1.35
K72R	R47G	A55V	R14T	F103H	7	1.35
K72R	R47G	A55V	F103H	M39I	7	1.35
K72R	R47G	A55V	F103H	G74D	7	1.35
K72R	R47G	A55V	F103H	V80A	7	1.35
K72R	R47G	A55V	F103H	R14T	7	1.35
K72R	R47G	R14T	F103H	M39I	7	1.35
K72R	R47G	R14T	F103H	G74D	7	1.35
K72R	R47G	R14T	F103H	V80A	7	1.35
K72R	R47G	R14T	F103H	A55V	7	1.35
K72R	A55V	R14T	F103H	M39I	7	1.35
K72R	A55V	R14T	F103H	G74D	7	1.35
K72R	A55V	R14T	F103H	V80A	7	1.35
K72R	A55V	R14T	F103H	R47G	7	1.35
V80A	R47G	A55V	R14T	M39I	7	1.35
V80A	R47G	A55V	R14T	G74D	7	1.35
V80A	R47G	A55V	R14T	K72R	7	1.35
V80A	R47G	A55V	R14T	F103H	7	1.35
V80A	R47G	A55V	F103H	M39I	7	1.35
V80A	R47G	A55V	F103H	G74D	7	1.35
V80A	R47G	A55V	F103H	K72R	7	1.35
V80A	R47G	A55V	F103H	R14T	7	1.35
V80A	R47G	R14T	F103H	M39I	7	1.35
V80A	R47G	R14T	F103H	G74D	7	1.35
V80A	R47G	R14T	F103H	K72R	7	1.35
V80A	R47G	R14T	F103H	A55V	7	1.35
V80A	A55V	R14T	F103H	M39I	7	1.35
V80A	A55V	R14T	F103H	G74D	7	1.35
V80A	A55V	R14T	F103H	K72R	7	1.35
V80A	A55V	R14T	F103H	R47G	7	1.35
D36S	T100A	T68N	L80S	D90E	7	1.35
D36S	T100A	T68N	D90E	L80S	7	1.35
D36S	T100A	L80S	D90E	T68N	7	1.35
D36S	T68N	L80S	D90E	T100A	7	1.35
R47G	A55V	R14T	F103H	M39I	7	1.35
R47G	A55V	R14T	F103H	G74D	7	1.35
R47G	A55V	R14T	F103H	K72R	7	1.35
R47G	A55V	R14T	F103H	V80A	7	1.35
T100A	T68N	L80S	D90E	D36S	7	1.35

REFERENCE LIST

Adolf-Bryfogle, J., Kalyuzhniy, O., Kubitz, M., Weitzner, B. D., Hu, X., Adachi, Y., ... & Dunbrack Jr, R. L. (2018). RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. *PLoS computational biology*, *14*(4), e1006112.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular biology of the cell*. New York: Garland Science.

Allen, D., Cumano, A., Dildrop, R., Kocks, C., Rajewsky, K., Rajewsky, N., ... & Siekevitz, M. (1987). Timing, genetic requirements and functional consequences of somatic hypermutation during B-cell development. *Immunological reviews*, *96*, 5-22.

Baran, D., Pszolla, M. G., Lapidoth, G. D., Norn, C., Dym, O., Unger, T., ... & Fleishman, S. J. (2017). Principles for computational design of binding antibodies. *Proceedings of the National Academy of Sciences*, *114*(41), 10900-10905.

Benkovic, S. J. (1992). Catalytic antibodies. *Annual review of biochemistry*, *61*(1), 29-54.

Bhayani, J. A., & Ballicora, M. A. (2022). Determination of dissociation constants of protein ligands by thermal shift assay. Biochemical and biophysical research communications, 590, 1-6.

Boehr, D. D., McElheny, D., Dyson, H. J., & Wright, P. E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*, *313*(5793), 1638-1642.

Braden, B. C., Dall'Acqua, W., Eisenstein, E., Fields, B. A., Goldbaum, F. A., Malchiodi, E. L., ... & Poljak, R. J. (1995). Protein motion and lock and key complementarity in antigen-antibody reactions. *Pharmaceutica Acta Helvetiae*, *69*(4), 225-230.

Bugg, T. D. (2012). *Introduction to enzyme and coenzyme chemistry*. John Wiley & Sons. pg. 232

Burnett, D. L., Schofield, P., Langley, D. B., Jackson, J., Bourne, K., Wilson, E., ... & Christ, D. (2020). Conformational diversity facilitates antibody mutation trajectories and discrimination between foreign and self-antigens. *Proceedings of the National Academy of Sciences*, *117*(36), 22341-22350.

Bustos-Jaimes, I., Mora-Lugo, R., Calcagno, M. L., & Farrés, A. (2010). Kinetic studies of Gly28: Ser mutant form of Bacillus pumilus lipase: Changes in kcat and thermal dependence. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1804(12), 2222-2227.

Dalkas, G. A., Teheux, F., Kwasigroch, J. M., & Rooman, M. (2014). Cation–π, amino–π, π–π, and H-bond interactions stabilize antigen–antibody interfaces. *Proteins: Structure, Function, and Bioinformatics*, 82(9), 1734-1746.

Debler, E. W., Müller, R., Hilvert, D., & Wilson, I. A. (2008). Conformational isomerism can limit antibody catalysis. *Journal of Biological Chemistry*, *283*(24), 16554-16560.

Debler, E. W., Müller, R., Hilvert, D., & Wilson, I. A. (2009). An aspartate and a water molecule mediate efficient acid-base catalysis in a tailored antibody pocket. *Proceedings of the National Academy of Sciences, 106*(44), 18539-18544.

Demchenko, A. P. (2002). The red-edge effects: 30 years of exploration. *Luminescence: the journal of biological and chemical luminescence*, *17*(1), 19-42.

Ehrenmann F., Kaas Q. and Lefranc M.-P. *Nucleic Acids Res*., 38:D301-D307 (2010).

Ferdous, S., & Martin, A. C. (2018). AbDb: antibody structure database—a database of PDB-derived antibody structures. *Database*, 2018.

Finkelstein, I. J., Ishikawa, H., Kim, S., Massari, A. M., & Fayer, M. D. (2007). Substrate binding and protein conformational dynamics measured by 2D-IR vibrational echo spectroscopy. *Proceedings of the National Academy of Sciences*, *104*(8), 2637-2642.

Foote, J., & Winter, G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *Journal of molecular biology*, *224*(2), 487-499.

Gelin, B. R., & Karplus, M. (1979). Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry*, *18*(7), 1256-1268.

Goethe, M., Fita, I., & Rubi, J. M. (2015). Vibrational entropy of a protein: large differences between distinct conformations. *Journal of Chemical Theory and Computation*, *11*(1), 351-359.

Goryanova, B., Goldman, L. M., Ming, S., Amyes, T. L., Gerlt, J. A., & Richard, J. P. (2015). Rate and Equilibrium Constants for an Enzyme Conformational Change during Catalysis by Orotidine 5′-Monophosphate Decarboxylase. Biochemistry, 54(29), 4555-4564.

Guddat, L. W., Shan, L., Anchin, J. M., Linthicum, D. S., & Edmundson, A. B. (1994). Local and transmitted conformational changes on complexation of an anti-sweetener Fab. *Journal of molecular biology*, *236*(1), 247-274.

Haji-Ghassemi, O., Müller-Loennies, S., Brooks, C. L., MacKenzie, C. R., Caveney, N., Van Petegem, F., ... & Evans, S. V. (2018). Subtle Changes in the Combining Site of the Chlamydiaceae-Specific mAb S25-23 Increase the Antibody–Carbohydrate Binding Affinity by an Order of Magnitude. *Biochemistry*, *58*(6), 714-726.

Haynes, M. R., Stura, E. A., Hilvert, D., & Wilson, I. A. (1994). Routes to catalysis: structure of a catalytic antibody and comparison with its natural counterpart. *Science*, *263*(5147), 646-652.

Ishikawa, F., Uno, K., Nishikawa, M., Tsumuraya, T., & Fujii, I. (2013). Antibody-catalyzed decarboxylation and aldol reactions using a primary amine molecule as a functionalized small nonprotein component. *Bioorganic & medicinal chemistry*, *21*(22), 7011-7017.

Ishikawa, H., Kwak, K., Chung, J. K., Kim, S., & Fayer, M. D. (2008). Direct observation of fast protein conformational switching. *Proceedings of the National Academy of Sciences*, *105*(25), 8619-8624.

Johnson, K. A. (2019). Kinetic Analysis for the New Enzymology: Using computer simulation to learn kinetics and solve mechanisms. KinTek Corporation. pp.149-191.

Johnson, K. A., Simpson, Z. B., & Blom, T. (2009). FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. Analytical biochemistry, 387(1), 30-41.

Kabir, K. L., Nussinov, R., Ma, B., & Shehu, A. (2021). Antigen Binding Reshapes Antibody Energy Landscape and Conformation Dynamics. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2519-2526). IEEE.

Karshikoff, A., Nilsson, L., & Ladenstein, R. (2015). Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *The FEBS journal*, *282*(20), 3899-3917.

Kodandapani, R., Veerapandian, L., Ni, C. Z., Chiou, C. K., Whittal, R. M., Kunicki, T. J., & Ely, K. R. (1998). Conformational change in an anti-integrin antibody: Structure of OPG2 Fab bound to a β3 peptide. *Biochemical and biophysical research communications*, *251*(1), 61-66.

Koenig, P., Lee, C. V., Walters, B. T., Janakiraman, V., Stinson, J., Patapoff, T. W., & Fuh, G. (2017). Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding. *Proceedings of the National Academy of Sciences*, *114*(4), E486-E495.

Kokkonen, P., Beier, A., Mazurenko, S., Damborsky, J., Bednar, D., & Prokop, Z. (2021). Substrate inhibition by the blockage of product release and its control by tunnel engineering. *RSC chemical biology*, *2*(2), 645-655.

Kuah, E., Toh, S., Yee, J., Ma, Q., & Gao, Z. (2016). Enzyme mimics: advances and applications. *Chemistry–A European Journal*, *22*(25), 8404-8430.

Kwok, A., Camacho, I. S., Winter, S., Knight, M., Meade, R. M., Van der Kamp, M. W., ... & Pudney, C. R. (2021). A thermodynamic model for interpreting tryptophan excitation-

energy-dependent fluorescence spectra provides insight into protein conformational sampling and stability. *Frontiers in molecular biosciences*, *8*.

Lerner, R. A., Benkovic, S. J., & Schultz, P. G. (1991). At the crossroads of chemistry and immunology: catalytic antibodies. *Science*, *252*(5006), 659-6

Li, D., Liu, M. S., & Ji, B. (2015). Mapping the dynamics landscape of conformational transitions in enzyme: the adenylate kinase case. *Biophysical journal*, *109*(3), 647-660.

Li, T., Tracka, M. B., Uddin, S., Casas-Finet, J., Jacobs, D. J., & Livesay, D. R. (2015). Rigidity emerges during antibody evolution in three distinct antibody systems: evidence from QSFR analysis of Fab fragments. *PLoS Comput Biol*, *11*(7), e1004327.

Lin, J., Cheng, N., Hogle, J. M., Steven, A. C., & Belnap, D. M. (2013). Conformational shift of a major poliovirus antigen confirmed by immuno-cryogenic electron microscopy. *The Journal of Immunology*, *191*(2), 884-891.

Lindner, A. B., Eshhar, Z., & Tawfik, D. S. (1999). Conformational changes affect binding and catalysis by ester-hydrolysing antibodies. *Journal of molecular biology*, *285*(1), 421-430.

MacCallum RM, Martin AC, Thornton JM. Antibody-antigen interactions: contact analysis and binding site topography. *J Mol Biol* (1996) **262**:732–45.doi:10.1006/ jmbi.1996.0548

Moritsugu, K., Njunda, B. M., & Smith, J. C. (2010). Theory and normal-mode analysis of change in protein vibrational dynamics on ligand binding. *The Journal of Physical Chemistry B*, *114*(3), 1479-1485.

Mukherjee, S., & Chattopadhyay, A. (1995). Wavelength-selective fluorescence as a novel tool to study organization and dynamics in complex biological systems. *Journal of Fluorescence*, *5*(3), 237-246.

Nicholas, K. M., Wentworth, P., Harwig, C. W., Wentworth, A. D., Shafton, A., & Janda, K. D. (2002). A cofactor approach to copper-dependent catalytic antibodies. *Proceedings of the National Academy of Sciences*, *99*(5), 2648-2653.

Niessen, K. A., Xu, M., Paciaroni, A., Orecchini, A., Snell, E. H., & Markelz, A. G. (2017). Moving in the right direction: protein vibrations steering function. *Biophysical Journal*, *112*(5), 933-942.

Ofran, Y., Schlessinger, A., & Rost, B. (2008). Automated identification of complementarity determining regions (CDRs) reveals peculiar characteristics of CDRs and B cell epitopes. *The Journal of Immunology*, *181*(9), 6230-6235.

Ohkubo, Y. Z., & Thorpe, I. F. (2006). Evaluating the conformational entropy of macromolecules using an energy decomposition approach. *The Journal of chemical physics*, *124*(2), 024910.

Otten, R., Pádua, R. A., Bunzel, H. A., Nguyen, V., Pitsawong, W., Patterson, M., ... & Kern, D. (2020). How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science*, *370*(6523), 1442-1446.

Padlan, E. A. (1994). Anatomy of the antibody molecule. *Molecular immunology*, *31*(3), 169-217.

Padlan, E. A., Abergel, C., & Tipper, J. P. (1995). Identification of specificity-determining residues in antibodies. *The FASEB journal*, *9*(1), 133-139.

Pang, X., & Zhou, H. X. (2017). Rate constants and mechanisms of protein–ligand binding. Annual review of biophysics, 46, 105-130.

Pollack, S. J., Jacobs, J. W., & Schultz, P. G. (1986). Selective chemical catalysis by an antibody. *Science*, 234(4783), 1570-1573.

Reed, M. C., Lieb, A., & Nijhout, H. F. (2010). The biological significance of substrate inhibition: a mechanism with diverse functions. *Bioessays*, *32*(5), 422-429.

Rees, A. R. (2020, January). Understanding the human antibody repertoire. In *MAbs* (Vol. 12, No. 1, p. 1729683). Taylor & Francis.

Richard, J. P. (2019). Protein flexibility and stiffness enable efficient enzymatic catalysis. Journal of the American Chemical Society, 141(8), 3320-3331.

Rini, J. M., Schulze-Gahmen, U., & Wilson, I. A. (1992). Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science*, *255*(5047), 959-965.

Rodrigues, C. H., Pires, D. E., & Ascher, D. B. (2018). DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic acids research*, *46*(W1), W350-W355.

Rodrigues, C. H., Pires, D. E., & Ascher, D. B. (2021). DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Science*, *30*(1), 60-69.

Schleif R. AraC protein, regulation of the L-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. FEMS Microbiol Rev (2010) 1–18.

Sela-Culang, I., Kunik, V., & Ofran, Y. (2013). The structural basis of antibody-antigen recognition. *Frontiers in immunology*, *4*, 302.

Sheng, Z., Bimela, J. S., Katsamba, P. S., Patel, S. D., Guo, Y., Zhao, H., ... & Shapiro, L. (2022). Structural basis of antibody conformation and stability modulation by framework somatic hypermutation. *Frontiers in immunology*, 5573.

Sheng, Z., Schramm, C. A., Kong, R., Mullikin, J. C., Mascola, J. R., Kwong, P. D., ... & NISC Comparative Sequencing Program. (2017). Gene-specific substitution profiles describe the types and frequencies of amino acid changes during antibody somatic hypermutation. *Frontiers in immunology*, *8*, 537.

Stanfield, R. L., Takimoto-Kamimura, M., Rini, J. M., Profy, A. T., & Wilson, I. A. (1993). Major antigen-induced domain rearrangements in an antibody. *Structure*, *1*(2), 83-93

Stevenson, J. D., & Thomas, N. R. (2000). Catalytic antibodies and other biomimetic catalysts. *Natural product reports*, *17*(6), 535-577.

Tanaka, F., Lerner, R. A., & Barbas, C. F. (2000). Reconstructing aldolase antibodies to alter their substrate specificity and turnover. *Journal of the American Chemical Society*, *122*(19), 4835-4836.

Tramontano, A., Chothia, C., & Lesk, A. M. (1990). Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *Journal of molecular biology*, *215*(1), 175-182.

Turton, D. A., Senn, H. M., Harwood, T., Lapthorn, A. J., Ellis, E. M., & Wynne, K. (2014). Terahertz underdamped vibrational motion governs protein-ligand binding in solution. *Nature communications*, *5*(1), 1-6.

van den Elsen, J., Vandeputte-Rutten, L., Kroon, J., & Gros, P. (1999). Bactericidal antibody recognition of meningococcal PorA by induced fit: comparison of liganded and unliganded Fab structures. *Journal of Biological Chemistry*, *274*(3), 1495-1501.

Vihinen, M. (1987). Relationship of protein flexibility to thermostability. *Protein Engineering, Design and Selection*, *1*(6), 477-480.

Vogt, A. D., Pozzi, N., Chen, Z., & Di Cera, E. (2014). Essential role of conformational selection in ligand binding. *Biophysical chemistry*, *186*, 13-21.

Wagner, J., Lerner, R. A., & Barbas, C. F. (1995). Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. *Science*, *270*(5243), 1797-1800.

Wagner, J., Lerner, R. A., & Barbas, C. F. (1995). Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. *Science*, *270*(5243), 1797-1800.

Wang, F., Sen, S., Zhang, Y., Ahmad, I., Zhu, X., Wilson, I. A., ... & Schultz, P. G. (2013). Somatic hypermutation maintains antibody thermodynamic stability during affinity maturation. *Proceedings of the National Academy of Sciences*, *110*(11), 4261-4266.

Wedemayer, G. J., Patten, P. A., Wang, L. H., Schultz, P. G., & Stevens, R. C. (1997). Structural insights into the evolution of an antibody combining site. *Science*, *276*(5319), 1665-1669.

Wu, T. T., & Kabat, E. A. (1970). An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *Journal of Experimental Medicine, 132*(2), 211-250.

Yin, J., Andryski, S. E., Beuscher, A. E., Stevens, R. C., & Schultz, P. G. (2003). Structural evidence for substrate strain in antibody catalysis. *Proceedings of the National Academy of Sciences, 100*(3), 856-861.

Zeymer, C., Zschoche, R., & Hilvert, D. (2017). Optimization of enzyme mechanism along the evolutionary trajectory of a computationally designed (retro-) aldolase. *Journal of the American Chemical Society, 139*(36), 12541-12549.

Zhong, G., Lerner, R. A., & Barbas III, C. F. (1999). Broadening the aldolase catalytic antibody repertoire by combining reactive immunization and transition state theory: new enantio- and diastereoselectivities. *Angewandte Chemie International Edition, 38*(24), 3738-3741.

Zhou, H. X. (2010). From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. Biophysical journal, 98(6), L15-L17.

Zhou, H. X., Wlodek, S. T., & McCammon, J. A. (1998). Conformation gating as a mechanism for enzyme specificity. *Proceedings of the National Academy of Sciences, 95*(16), 9280-9283.

Zhu, X., Tanaka, F., Hu, Y., Heine, A., Fuller, R., Zhong, G., ... & Wilson, I. A. (2004). The origin of enantioselectivity in aldolase antibodies: crystal structure, site-directed mutagenesis, and computational analysis. *Journal of molecular biology, 343*(5), 1269-1280.

Zimmermann, J., Oakman, E. L., Thorpe, I. F., Shi, X., Abbyad, P., Brooks, C. L., ... & Romesberg, F. E. (2006). Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics. *Proceedings of the National Academy of Sciences, 103*(37), 13722-13727.

VITA

Charles Mettler was born and raised in Huntington, IN. Before attending Loyola University Chicago, he attended Wabash College where he earned a Bachelor of Arts in Biology, with Highest Distinction, in 2018. As an undergraduate, Mettler conducted research through internships, advanced laboratory courses, and independent study courses on carbon and energy dynamics in natural ecosystems.