

ENGINEERING CYANOVIRIN-N FOR ENHANCED VIRAL NEUTRALIZATION

Thesis by

Jennifer R. Keeffe

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2009

(Defended December 12, 2008)

© 2009

Jennifer R. Keefe

All Rights Reserved

Acknowledgements

I came to Caltech 5 ½ years ago as an enthusiastic first year graduate student ready to tackle the hard problems in protein design and learn computational methods. I quickly realized, however, that there was a reason these problems hadn't been solved before and that I really wasn't much of a computationalist. Luckily, through the good times and the bad, I had an amazing team of supporters who would both listen to me complain endlessly and join in the celebration of successful experiments.

I would especially like to thank my advisor, Steve Mayo. I came to Caltech hoping that I would learn some amazing new techniques and ideas, and with Steve's help I was able to do that. Although the computational component didn't exactly work out for me, Steve and I worked together to create and work on a project that was right up my alley—experimental work!

I would also like to thank my thesis committee: Professors Pamela Bjorkman, David Chan, and Peter Dervan. Whether they knew it or not, they pushed me to think and understand more than even I thought was possible. It was an incredible opportunity to work with all of these fabulous scientists.

I couldn't have done 90% of the work presented here without amazing collaborators. The most important collaborator to this work has been (and hopefully will continue to be) Priyanthi Peiris. Pri ran almost all of the HIV neutralization assays on my variants and did it with a smile! Even when I knew she was dreading that email from me telling her I had a new set of samples for her, she was always there for me! I also appreciate her showing me the assay so I could get a better understanding of what was going on and answering my endless questions!

I also had the opportunity to pick the brains of many other Bjorkman lab members, including Josh Klein, Anthony West, and Noreen Tiangco. These discussions would sometimes open up a whole new world of understanding and sometimes would just make me happy! I would especially like to thank Anthony West for setting up the collaboration with the CAVD facility and Dr. Michael Seaman. I would also like to thank the Protein Expression Center, specifically Jost Vielmetter and Michael Anaya, for their help in various aspects of my work here at Caltech.

I would like to thank my “team” of crystallographers who helped me at every step of solving my crystal structures. Leonard Thomas was my “go-to-guy” and was always available to help me when I ran into trouble refining and Pavle Nikolovski in the Molecular Observatory set up many crystal trays and always managed to be excited when I saw him!

My time here at Caltech would not have even been a reality had it not been for the support and guidance of my previous advisors, Dr. Spencer Anthony-Cahill at Western Washington University, and Dr. Rachel Klevit at the University of Washington. Both my time as an undergraduate researcher and as a lab scientist helped prepare me for my Caltech experience and made me realize how much I love science and experimental work.

The science has been amazing, but the people that I have met here at Caltech have made these 5 ½ years go by so much smoother. The Mayo lab is an extremely supportive place to be and I’d like to thank my two best friends, Heidi Privett and Ben Allen, for being there when I needed them. Ben, my “lab husband,” is a friend like no other, and I

will surely miss our conversations and such. Heidi is an awesome friend—the one that you take for granted until she's gone for the summer and you miss her SO MUCH!

I would also like to thank all of the other members of the Mayo lab, past and present, for making this a memorable experience. I'd specifically like to thank Jonathan Kyle Lassila for allowing me to work with him on the chorismate mutase project and teaching me a lot about enzymes, computational design, and graduate school in general. I'd also like to thank John Yong, a former rotation student in the lab, for putting up with me for 3 months and still wanting to talk to me! John made a number of CVN₂ variants for me and assayed them and was really instrumental in getting this project going. He did all of that while having to deal with me...a saint, I think!

I'd also like to thank the staff in the Mayo lab. While many times their jobs are unglamorous or unappreciated, my time here would not have been the same without them and they are all indispensable. Rhonda DiGiusto is not only a lab manager and technician, but she is a great friend. All of our administrative assistants have been helpful and great fun to be around: Erin Drez, Cathy Miles, and Cynthia Carlson. And of course, this thesis would be full of missing commas and other mistakes if it weren't for the fabulous services of Marie Ary, our technical writer and awesome “I-want-to-procrastinate-while-picking-something-up-off-the-printer-and-talk-to-Marie” conversation partner!

I'd also like to thank all of the friends that I've met here in Southern California. Michelle Fontes and Tara Gomez were always willing and able to distract me from the failures in the lab with scrapbooking, while “the zoo crew” would gladly meet us at Disneyland to take a load off.

Finally, I'd like to thank my ever-supportive family. My mom and step-dad, Roxanne and Gary Lowe, and my dad and step-mom, Jim and Anne White, instilled in me from the beginning good work ethic and excitement about learning. I'd also like to thank my brother and stepsisters for resisting my pleas to come down and visit with their families—I know you just wanted me to work harder!

Most of all, I'd like to thank my husband, Dan Keeffe. Without him, life is uninteresting and not nearly as much fun! His support got me through the toughest times—including Rob Phillips' Applied Physics class when Dan made me eat ice cream because I was too stressed to eat anything else. After all the hard times, we've made it together!

Abstract

Cyanovirin-N (CVN) is an 11-kDa lectin originally isolated from the cyanobacterium *Nostoc ellipsosporum* during a high-throughput screen for novel anti-HIV activities. In addition to having anti-HIV activity, CVN has since been shown to neutralize a number of other enveloped viruses including influenza and Ebola. This antiviral activity is attributed to two homologous carbohydrate binding sites that specifically bind $\alpha(1-2)$ -linked oligomannose glycosylation sites present on many envelope glycoproteins. Because of its broad ability to neutralize enveloped viruses, CVN is a promising target as a potential therapeutic or prophylactic.

In this work, we oligomerized CVN to determine whether an increase in the number of carbohydrate binding sites has an effect on its viral neutralization activity. To create obligate dimers, we covalently linked multiple copies of CVN through flexible polypeptide linkers. Using HIV-1 as our viral system, we found that a tandem repeat of two CVN molecules (CVN₂) increased the efficacy of HIV-1 neutralization by up to 10-fold. An additional benefit was not seen when CVN was trimerized. We also show here that CVN and the CVN₂ variants show extensive cross-clade reactivity and higher neutralization efficacy than the most broadly reactive neutralizing antibodies. To determine whether any major structural changes or differences in domain swapping occurred because of the linkage, we solved the crystal structures of three dimeric variants and showed that all variants are intramolecularly domain-swapped.

Additionally, we present in this thesis a novel CVN-Fc chimera, a “lectibody,” which shows antiviral activity similar to wild-type CVN. This variant is dimerized through the Fc region of an antibody and has the additional benefit of incorporating Fc-

mediated effector functions, which may be therapeutically advantageous. Initial results on the lectin body indicate that domain swapping of CVN has an integral role in the antiviral function as well as in the overall folding and stability of the molecule. Future work on this variant to assay the effector functions as well as create a monodisperse, stable variant are underway.

Although CVN is already a promising candidate for antiviral therapeutics, we show here that engineering CVN to add additional functionalities or creating variants with an increased number of active sites can significantly enhance the potential benefit of these molecules.

TABLE OF CONTENTS

Acknowledgements		iii
Abstract		vii
Table of Contents		ix
Figures and Tables		x
Abbreviations		xii
Chapters		
Chapter 1	<i>Introduction</i>	1
Chapter 2	<i>Engineered cyanovirin-N oligomers show enhanced HIV neutralization</i>	20
Chapter 3	<i>Structural characterization of engineered cyanovirin-N dimers</i>	53
Chapter 4	<i>Lectibody: design and characterization of a cyanovirin-N – Fc chimera</i>	83
Appendix		
Appendix A	<i>Toward computationally designed calmodulin variants with enhanced peptide binding specificity</i>	109
Appendix B	<i>Computationally designed variants of Escherichia coli chorismate mutase show altered catalytic activity</i>	154
Appendix C	<i>Exhaustive mutagenesis of six secondary active site residues in E. coli chorismate mutase shows the importance of hydrophobic side chains and a helix N-capping position for stability and catalysis</i>	169

FIGURES AND TABLES

Figure 1-1.	Wild type CVN structures	19
Table 2-1.	CVN ₂ and CVN ₃ linker sequences	43
Table 2-2.	IC ₅₀ s (nM) of CVN and HIV neutralizing antibodies against various envelopes in HIV clades A, B, and C.....	44
Figure 2-1.	Model of generic CVN ₂ protein.....	45
Figure 2-2.	HIV neutralization assay results	46
Figure 2-3.	CVN ₃ HIV neutralization data	47
Figure 2-4.	WT CVN cross-clade reactivity compared to broadly HIV neutralizing antibodies.....	48
Figure 2-5.	Engineered CVN ₂ variants neutralize most HIV pseudoviruses with a lower IC ₅₀ compared to the most effective broadly neutralizing antibody (NAb).....	49
Figure 2-6.	Engineered CVN variants are more effective at neutralizing various HIV pseudoviruses than WT CVN	50
Figure 2-7.	Cellular toxicity assay of CVN and CVN ₂ s.....	51
Figure 2-8.	Carbohydrate binding site spacing in CVN and 2G12 anti-HIV Fab	52
Table 3-1.	Crystallographic statistics	73
Figure 3-1.	Wild type CVN structures	74
Figure 3-2.	CVN ₂ L0 crystal structure compared to WT CVN	75
Figure 3-3.	CVN ₂ L0 structure.....	76
Figure 3-4.	CVN ₂ L1 crystal structures compared to WT CVN.....	77
Figure 3-5.	CVN ₂ L1 P3 ₂ 21 structure	78
Figure 3-6.	CVN ₂ L1 P4 ₁ 2 ₁ 2 structure	79
Figure 3-7.	CVN ₂ L10 structure comparison to WT CVN.....	80
Figure 3-8.	CVN ₂ L10 structure.....	81
Figure 3-9.	Carbohydrate binding site spacing in CVN and 2G12 anti-HIV Fab	82

Figure 4-1.	Model of the CVN-Fc lectibody	104
Figure 4-2.	Assessment of glycosylation site deletion variants.....	105
Figure 4-3.	CVN-Fc N30S purification and activity.....	106
Figure 4-4.	Surface plasmon resonance assays of lectibodies and Fc.....	107
Figure 4-5.	N30S/P51G-Fc purification and activity	108
Table A-1.	CaM-binding peptides	143
Table A-2.	Sequence alignment of calmodulin designs.....	144
Table A-3.	K_{DS} for CaM variants as determined by fluorescence assays.....	145
Table A-4.	K_{DS} for CaMKIp positive design-derived single mutant variants as determined by fluorescence assays	146
Figure A-1.	CaM-peptide structures	147
Figure A-2.	Negative design result structures	148
Figure A-3.	Tryptophan fluorescence assay data analysis	149
Figure A-4.	Biacore kinetics of wild type CaM and smMLCK _p	150
Figure A-5.	Biacore equilibrium data	151
Figure A-6.	CaM variants – peptide equilibrium binding data.....	152
Figure A-7.	Competition Biacore assay	153
Table B-1.	Kinetic parameters of wild type and mutant EcCM.....	166
Figure B-1.	Chorismate to prephenate rearrangement	167
Figure B-2.	Predicted hydrogen bonding in the Ala32Ser variant	168
Table C-1.	Kinetics and stabilities of EcCM variants	194
Table C-2.	Active site alignment of solved AroQ mutase structures	197
Scheme C-1.	Chorismate to prephenate rearrangement	198
Figure C-1.	EcCM structure and active site	199
Figure C-2.	Asp48 interactions.....	200

ABBREVIATIONS

BsCM	chorismate mutase from <i>Bacillus subtilis</i>
CaM	calmodulin
CaMKI _p	calmodulin kinase I peptide
CD	circular dichroism
CV	column volume
CVN/CV-N	cyanovirin-N
CVN ₂ LX	cyanovirin-N dimer with a linker length of X amino acids
CVN ₃ LX	cyanovirin-N trimer with a linker length of X amino acids
Da	daltons
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EcCM	chorismate mutase domain of <i>E. coli</i> chorismate mutase-prephenate dehydratase
EDTA	ethylene diamine tetraacetic acid
EGTA	ethylene glycol tetraacetic acid
Fab	antigen binding fragment of an antibody
Fc	constant region of an antibody
FMEC	FASTER-derived minimum energy conformation
GnHCl	guanidine HCl
HERO	hybrid exact rotamer optimization
HIV	human immunodeficiency virus
IC ₅₀	concentration at which there is 50% inhibition
IPTG	isopropyl β-D-1-thiogalactopyranoside
k _{cat}	catalytic rate constant
K _D	dissociation binding constant
K _M	Michaelis constant
LB	Luria-Bertani broth
MWCO	molecular weight cut off
NAb	neutralizing antibody
Ni-NTA	nickel-nitrilotriacetic
NMR	nuclear magnetic resonance spectroscopy
OD	optical density at a specific wavelength
ORBIT	optimization of rotamers by iterative techniques
PCR	polymerase chain reaction
PDB	protein data bank
PMSF	phenylmethanesulphonyl fluoride (a serine protease inhibitor)
RU	response unit (units for binding on Biacore)
scFv	single chain variable region fragment
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
smMLCK _p	smooth muscle light chain kinase peptide
TCID ₅₀	tissue culture infectious dose 50%
T _m	midpoint of temperature denaturation curve
WT	wild type

Chapter 1

Introduction

Viruses are subcellular agents that rely on the molecular machinery of a host cell to replicate. They have evolved to infect almost every organism and are increasingly being employed in scientific research. Although viruses contain very small genomes encoding only a few necessary proteins, they have adapted to evade immune systems and transmit efficiently from host to host.^{1,2} While viruses come in different shapes, sizes, and types of genomic information, here, we are specifically interested in enveloped viruses. Enveloped viruses are a class of viruses that bud from the plasma or internal membrane of plants or animals during their replication.² The newly budded viral particle contains the genomic material inside a protein capsid, which in turn is surrounded by membrane from the host and envelope proteins. Envelope proteins are often heavily glycosylated by the host machinery, and therefore are often not immediately recognized by the immune system. The envelope glycoproteins are usually responsible for interactions with cellular receptors on target host cells, triggering membrane fusion and infection.

Enveloped viruses cause many well known diseases, including influenza, Ebola, chicken pox, SARS (severe acute respiratory syndrome), small pox, and AIDS.^{1,2} Human immunodeficiency virus (HIV), the virus that causes AIDS, affects approximately 33 million people throughout the world and causes approximately 2 million HIV-related deaths per year.³ While current retroviral therapies have extended the length and quality of life of those infected with HIV, resistant strains are becoming increasingly common, and additional treatments and a broad vaccine are necessary to prevent additional infections. Although influenza does not typically cause the mortality of HIV, it is a highly contagious virus that can be lethal, usually in the very young and very old and in those with immune deficiencies. Moreover, influenza pandemics, such as the one in 1918 when

an estimated 40 million people worldwide were killed, are capable of causing a significant number of deaths, including in healthy young adults.^{2,4}

Current treatment for viral infection varies widely for different viruses. Effective vaccines are available for smallpox, measles, hepatitis, and varicella-zoster (chicken pox) viruses, among others.² Influenza vaccines are typically effective against the strains included in the vaccine, but must be readministered every year due to the rapid mutation rate of the virus.⁵ For some enveloped viruses, however, there is no vaccine and therefore treatment of the infection is the primary clinical goal. There are currently no vaccines available for Ebola virus, herpesviruses, hanta viruses, HIV, and many other potentially deadly viruses. For many of these diseases, treatment is administered to make the patient more comfortable, provide symptom relief, or decrease the viral load to allow the immune system to more easily fight off the infection.^{2,6,7}

Efforts to develop a vaccine for HIV have been met with limited success, with promising laboratory results thus far leading only to failures in clinical trials.⁸⁻¹⁰ Although an effective vaccine has not yet been developed, researchers have isolated some broadly neutralizing anti-HIV antibodies from non-progressing patients, leading to hope that a vaccine may be possible.¹¹⁻¹⁴ While research continues on developing an effective and cross-reactive vaccine, patients currently rely on antiviral drugs to decrease their viral load and prolong their lives. HIV antiviral therapy usually consists of three or more antiretroviral drugs from at least two inhibitory classes in a therapeutic regimen known as highly active antiretroviral therapy (HAART).^{1,2,10} As of 2008, 32 antivirals have been approved by the FDA for treatment of HIV-1 infections. These mostly small molecule drugs can be divided into six categories: nucleoside reverse transcriptase inhibitors

(NRTIs), nonnucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors, fusion inhibitors, entry inhibitors, and integrase inhibitors. Although HAART has been quite successful at reducing the viral load of patients, the rapid mutation rate of HIV often eventually leads to drug resistant strains, rendering antiviral treatment ineffective.¹⁵ New and improved small molecule and protein-based therapeutics that limit the development of drug resistant strains, along with the creation of effective vaccines, will greatly improve the outlook for currently infected patients as well as those at high risk for becoming infected.

Unlike for HIV, there is a very effective vaccine for influenza.^{2,5} However, the inconvenience and cost of yearly immunization, as well as the unpredictable mutation of influenza, means that millions of people are susceptible to infection every year.^{16,17} In addition, the recent emergence of a highly lethal H5N1 strain (“bird flu”) has led to concerns that this strain could become easily transmittable from human to human or weaponized, creating a massive influenza pandemic.¹⁸⁻²¹ Influenza A, including H5N1 strains, can be treated with antiviral medications, including oseltamivir (Tamiflu) and zanamivir (Relenza),^{1,22,23} although resistance to oseltamivir and other common influenza antivirals have already been reported in H5N1 cases.²⁴⁻²⁷ Additional anti-influenza therapeutics and prophylactics would be beneficial in the case of a pandemic, especially for medical personnel and first-line defenders.

Cyanovirin-N (CVN), a potent antiviral lectin, is uniquely positioned to become a novel therapeutic and prophylactic for enveloped virus treatment and prevention. CVN is a small 11-kDa protein that was originally isolated from the cyanobacterium *Nostoc ellipsosporum* during a high-throughput screen intended to discover novel anti-HIV

activities. CVN was shown to be active against various strains of HIV, including primary isolates of HIV-1, T-lymphocyte-tropic strains, macrophage-tropic strains, and HIV-2.²⁸ This study also showed that CVN specifically interacts with the HIV envelope glycoprotein, gp120. Because of the great potential of CVN as an antiviral therapeutic, a number of additional studies quickly elucidated that CVN is a lectin with two carbohydrate binding sites that specifically bind to $\alpha(1-2)$ linked oligomannose moieties within Man-8 or Man-9 glycosylation sites.²⁹⁻³² Interestingly, high mannose glycosylation is very uncommon in mammalian oligosaccharides, but is often seen on the surface of viruses and microorganisms, creating an important distinction between the recognition of pathogens during potential treatment with CVN.³³ The two carbohydrate binding sites in CVN show distinct affinities for Man-9: the “high affinity” binding site has a K_a of 7.2×10^6 M, and the “low affinity” binding site has an approximately 10-fold lower affinity.³⁰ Later studies confirmed that both binding sites are important for HIV neutralization, and the destruction of either site renders the CVN variant inactive.³⁴⁻³⁶ These two binding sites provide a mechanism for high affinity and high avidity interactions with glycosylated envelope proteins on viruses.

In addition to its potent activity against HIV,^{28,37-39} CVN has also been shown to effectively neutralize influenza,^{40,41} Ebola,^{42,43} hepatitis C,⁴⁴ herpesvirus 6, and measles virus.⁴⁵ In each case, CVN binds specifically to high mannose glycosylation sites on envelope glycoproteins and inhibits vital interactions between the virus and the host cell. To date, CVN has shown no antiviral activity against any non-enveloped viruses, including rhinoviruses and enteric viruses,⁴⁰ and also appears to be inactive against some enveloped viruses, including vaccinia.⁴⁵

In the case of influenza, CVN interacts with glycosylation sites on hemagglutinin, one of the two surface glycoproteins expressed on influenza particles. CVN showed highly potent antiviral activity against strains of influenza A, including H1N1 and N3N2, exhibited moderate neutralization against influenza B strains,^{40,41} and was able to protect mice from a highly fatal strain of influenza when administered before infection.⁴⁶ Unfortunately, there was no apparent activity against H5N1 strains (“bird flu”).⁴¹ However, with increased understanding of the specific interactions between CVN and hemagglutinin, engineered variants may provide increased neutralization of H5 and other strains, allowing a broad and potentially successful method for preventing infection in the case of an influenza outbreak in the absence of an effective vaccine.

Similarly to the mechanism for influenza neutralization, CVN inhibits HIV by binding to glycosylated surface proteins. In this case, CVN binds specifically and with high affinity to glycosylated gp120²⁸ and with significantly lower affinity to gp41.⁴⁷ CVN binds with approximately 5:1 stoichiometry to soluble gp120, indicating that there are not only multiple sites of glycosylation to which CVN can bind, but that avidity may also play a significant role in the neutralization of HIV.⁴⁷ Additionally, studies have shown that CVN does not bind to a single glycan on gp120, but instead three to five separate N-linked glycosylation sites must be mutated before CVN resistance is incurred.^{39,48,49} Although CVN-treated gp120 can still bind to soluble CD4,⁵⁰ membrane-bound CD4 binding is inhibited, probably due to steric constraints.^{45,50} CVN also blocks the interaction between gp120 and the CCR5 coreceptor, adding a secondary inhibitory effect.⁴⁵ These two mechanisms together make CVN an efficient inhibitor of the pre-membrane fusion event of HIV infection.

The structure of CVN has been extensively studied to attempt to elucidate a mechanism for its broad antiviral activity. CVN exists in solution mainly as a monomer, but a trapped, metastable domain-swapped dimer can be formed.⁵¹ In crystal form, however, wild-type (WT) CVN is only seen as a domain-swapped dimer (Figure 1-1). The monomer contains two pseudo domains that display high sequence homology. Interestingly, however, the gene does not appear to result from a simple gene duplication since the first domain contains residues 1-39 and 90-101, and the second domain contains residues 39-89.⁵² Instead, there was probably a gene duplication then rearrangement or an uneven recombination event that resulted in the CVN gene. Monomeric CVN also contains two native disulfide bridges: between residues 8 and 22, and between residues 58 and 73. These two disulfide bridges are located near each end of the molecule and anchor the secondary structure. The dimer contains the same topology, but is domain-swapped at residues 51-53.⁵³ In the dimeric structure, the first domain of one chain (A) forms a “monomer-like” structure with the second domain of the other chain (B’) in an almost symmetric domain swapping (Figure 1-1B). The two quasi-monomers can sample different orientations relative to each other due to the flexibility of the domain-swapped region, and the orientation appears to be pH dependent in crystal structures.^{51,53,54}

A number of groups have attempted to modulate the domain swapping of CVN to determine whether this property is a crystallographic artifact or a biologically relevant state. Because the domain-swapped dimer of WT CVN is metastable at physiological temperatures, purified dimer quickly converts to monomer during the course of a neutralization assay.⁵¹ Variants have been generated that stabilize both the monomeric state⁵¹ and the dimeric state.^{55,56} However, the effect of dimerization remains unclear,

since some groups have concluded that the dimeric state is more active than monomeric WT CVN⁵⁵ and others have shown that monomeric and dimeric variants have the same antiviral activity.⁵⁶

CVN has great potential therapeutic value both as a prophylactic as well as a treatment for viral infection. In fact, CVN is currently in clinical trials as a prophylactic gel (Cellegy Pharmaceuticals, Inc.) and has been shown to be effective against both rectal and vaginal SIV/HIV-1 transmission in non-human primate studies when used as a topical microbicide.^{37,38} Additionally, it has been shown that CVN has limited toxicity in tissue culture,^{28,45,57} in mice,^{43,46,58,59} and in non-human primate models,^{37,38} although a recent study indicates that CVN can increase the levels of chemokines in treated cells and potentially allow much higher susceptibility for viral replication after CVN is removed.⁶⁰ Additionally, CVN can be prepared in large quantities, is stable for long periods of time, and is extremely resistant to degradation.^{28,61-63} It was also demonstrated that WT CVN can be specifically PEGylated to increase the serum half-life while retaining most of the anti-HIV activity.⁵⁹

Although the viral neutralization activity of CVN is important in the prevention of infection, this function may prove even more beneficial as a potential therapeutic. Because CVN specifically targets glycosylation on viral envelopes, escape variants will likely appear rapidly upon treatment with this lectin. Under evolutionary pressure by CVN^{39,41,48,49} and other carbohydrate-binding proteins,⁶⁴⁻⁶⁶ HIV and influenza have both been shown to eliminate N-linked glycosylation sites on their envelope proteins to escape neutralization. However, HIV-1 and other viruses use glycosylation to prevent recognition by the innate and adaptive immune systems.⁶⁷ With the removal of

glycosylation and the exposure of antigen, these viruses may become more sensitive to neutralization and clearance by the immune system.⁶⁸⁻⁷⁰ In fact, Reitter *et al.* found this to be true when rhesus monkeys were infected with SIV (simian immunodeficiency virus, an HIV homolog) lacking various glycosylation sites. In this case, the viruses were significantly more susceptible to antibody neutralization.⁷¹ Additionally, glycosylation of these viral proteins is often necessary for their proper folding and function, and therefore treatment with CVN or other lectins may decrease their viability.^{60,72,73}

Various attempts to increase the HIV neutralization of CVN have met with some success. Mori *et al.* showed that a chimera of CVN and an exotoxin from *Pseudomonas* had enhanced cytotoxicity to HIV-infected cells.⁷⁴ Another chimera between CVN and an allosteric peptide inhibitor of HIV-1 fusion also showed synergy between the two components, creating a more effective compound against HIV.⁷⁵ Attempts to engineer CVN itself, however, have not resulted in variants with increased potency.

Although the Mayo lab has typically used computational methods to study and engineer proteins, for the last few years I have worked on a non-computational project involving engineering CVN to create variants with increased antiviral potency. As described above, the information we gathered through characterizing these variants has led us to believe that they have potential therapeutic as well as scientific value. In Chapter 2 of this thesis, I describe the generation and characterization of dimeric and trimeric CVN variants that display increased HIV neutralization activity. Although we showed that these variants exhibit increased potency, a mechanism was elusive. We therefore solved four crystal structures of three CVN₂ variants to determine whether

structural changes or differences in domain swapping could account for the increase in activity. These data and results are presented in Chapter 3.

In Chapter 4 I discuss a chimeric CVN-Fc construct, which we have termed a “lectibody” because it is a fusion of a lectin (CVN) and the constant region of an antibody (Fc). This construct may have therapeutic value as a molecule that not only neutralizes free viral particles through the CVN domain, but also functions as part of the human immune system through the Fc, inducing immune response against infected cells which are budding new virus.

While the vast amount of research has been focused on CVN projects, I also had the opportunity to work on some computational protein design projects that are presented as chapters in the Appendix of this thesis. One project involved designing calmodulin (CaM) to optimize the protein-peptide surface and provide specificity between two high-affinity native CaM peptide partners (Appendix A). While this research did not ultimately result in a variant that displayed increased specificity, the groundwork was set for future CaM designs and experiments that may prove more successful.

I also had the opportunity to collaborate with Jonathan Kyle Lassila on the design and characterization of various chorismate mutase variants (Appendices B, C). The goal of this project was to investigate the role of secondary active site residues in an enzyme (these residues do not directly contact substrate but instead interact with active site residues). We used our protein design software (ORBIT) to generate a variant that showed a modest increase in catalytic efficiency and to identify other mutations that were consistent with activity. We also performed site-saturation mutagenesis on six secondary active site residues and characterized each active variant. This data allowed us to

determine the tolerance for mutation in a natural enzyme system and use that information for future computational studies involving functional proteins.

References

1. Fields, B. N., Knipe, D. M. & Howley, P. M. (2007). *Fields virology*. 5th edit, Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia.
2. Strauss, J. H. & Strauss, E. G. (2008). *Viruses and human disease*. 2nd edit, Elsevier / Academic Press, Amsterdam; Boston.
3. UNAIDS. (2008). Report on the global AIDS epidemic.
4. Reid, A. H. & Taubenberger, J. K. (2003). The origin of the 1918 pandemic influenza virus: a continuing enigma. *J Gen Virol* **84**, 2285-92.
5. Mossad, S. B. (2007). Influenza update 2007-2008: vaccine advances, pandemic preparation. *Cleve Clin J Med* **74**, 889-94.
6. Enserink, M. (2003). Virology. New vaccine and treatment excite Ebola researchers. *Science* **302**, 1141-2.
7. Emmert, D. H. (2000). Treatment of common cutaneous herpes simplex virus infections. *Am Fam Physician* **61**, 1697-706, 1708.
8. Burton, D. R., Desrosiers, R. C., Doms, R. W., Koff, W. C., Kwong, P. D., Moore, J. P., Nabel, G. J., Sodroski, J., Wilson, I. A. & Wyatt, R. T. (2004). HIV vaccine design and the neutralizing antibody problem. *Nat Immunol* **5**, 233-6.
9. Karlsson Hedestam, G. B., Fouchier, R. A., Phogat, S., Burton, D. R., Sodroski, J. & Wyatt, R. T. (2008). The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. *Nat Rev Microbiol* **6**, 143-55.
10. Endsley, A. N., Salama, N. N. & Ho, R. J. (2008). Combining drug and immune therapy: a potential solution to drug resistance and challenges of HIV vaccines? *Curr HIV Res* **6**, 401-10.
11. Burton, D. R., Stanfield, R. L. & Wilson, I. A. (2005). Antibody vs. HIV in a clash of evolutionary titans. *Proc Natl Acad Sci U S A* **102**, 14943-8.
12. Cardoso, R. M., Zwick, M. B., Stanfield, R. L., Kunert, R., Binley, J. M., Katinger, H., Burton, D. R. & Wilson, I. A. (2005). Broadly neutralizing anti-HIV antibody 4E10 recognizes a helical conformation of a highly conserved fusion-associated motif in gp41. *Immunity* **22**, 163-73.
13. Trkola, A., Purtscher, M., Muster, T., Ballaun, C., Buchacher, A., Sullivan, N., Srinivasan, K., Sodroski, J., Moore, J. P. & Katinger, H. (1996). Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol* **70**, 1100-8.

14. Burton, D. R., Pyati, J., Koduri, R., Sharp, S. J., Thornton, G. B., Parren, P. W., Sawyer, L. S., Hendry, R. M., Dunlop, N., Nara, P. L. & et al. (1994). Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* **266**, 1024-7.
15. Shafer, R. W. & Schapiro, J. M. (2008). HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev* **10**, 67-84.
16. Tosh, P. K., Boyce, T. G. & Poland, G. A. (2008). Flu myths: dispelling the myths associated with live attenuated influenza vaccine. *Mayo Clin Proc* **83**, 77-84.
17. Weingarten, S., Riedinger, M., Bolton, L. B., Miles, P. & Ault, M. (1989). Barriers to influenza vaccine acceptance. A survey of physicians and nurses. *Am J Infect Control* **17**, 202-7.
18. Webby, R. J. & Webster, R. G. (2003). Are we ready for pandemic influenza? *Science* **302**, 1519-22.
19. von Itzstein, M. (2008). Avian influenza virus, a very sticky situation. *Curr Opin Chem Biol* **12**, 102-8.
20. Ong, A., Kindhauser, M., Smith, I. & Chan, M. (2008). A global perspective on avian influenza. *Ann Acad Med Singapore* **37**, 477-81.
21. Krug, R. M. (2003). The potential use of influenza virus as an agent for bioterrorism. *Antiviral Res* **57**, 147-50.
22. Sugrue, R. J., Tan, B. H., Yeo, D. S. & Sutejo, R. (2008). Antiviral drugs for the control of pandemic influenza virus. *Ann Acad Med Singapore* **37**, 518-24.
23. Proenca-Modena, J. L., Macedo, I. S. & Arruda, E. (2007). H5N1 avian influenza virus: an overview. *Braz J Infect Dis* **11**, 125-33.
24. de Jong, M. D., Tran, T. T., Truong, H. K., Vo, M. H., Smith, G. J., Nguyen, V. C., Bach, V. C., Phan, T. Q., Do, Q. H., Guan, Y., Peiris, J. S., Tran, T. H. & Farrar, J. (2005). Oseltamivir resistance during treatment of influenza A (H5N1) infection. *N Engl J Med* **353**, 2667-72.
25. Le, Q. M., Kiso, M., Someya, K., Sakai, Y. T., Nguyen, T. H., Nguyen, K. H., Pham, N. D., Ngyen, H. H., Yamada, S., Muramoto, Y., Horimoto, T., Takada, A., Goto, H., Suzuki, T., Suzuki, Y. & Kawaoka, Y. (2005). Avian flu: isolation of drug-resistant H5N1 virus. *Nature* **437**, 1108.
26. He, G., Qiao, J., Dong, C., He, C., Zhao, L. & Tian, Y. (2008). Amantadine-resistance among H5N1 avian influenza viruses isolated in Northern China. *Antiviral Res* **77**, 72-6.

27. Cheung, C. L., Rayner, J. M., Smith, G. J., Wang, P., Naipospos, T. S., Zhang, J., Yuen, K. Y., Webster, R. G., Peiris, J. S., Guan, Y. & Chen, H. (2006). Distribution of amantadine-resistant H5N1 avian influenza variants in Asia. *J Infect Dis* **193**, 1626-9.
28. Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T., Gulakowski, R. J., Wu, L., Rivera, M. I., Laurencot, C. M., Currens, M. J., Cardellina, J. H., 2nd, Buckheit, R. W., Jr., Nara, P. L., Pannell, L. K., Sowder, R. C., 2nd & Henderson, L. E. (1997). Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother* **41**, 1521-30.
29. Bolmstedt, A. J., O'Keefe, B. R., Shenoy, S. R., McMahon, J. B. & Boyd, M. R. (2001). Cyanovirin-N defines a new class of antiviral agent targeting N-linked, high-mannose glycans in an oligosaccharide-specific manner. *Mol Pharmacol* **59**, 949-54.
30. Bewley, C. A. & Otero-Quintero, S. (2001). The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120. *J Am Chem Soc* **123**, 3892-902.
31. Shenoy, S. R., Barrientos, L. G., Ratner, D. M., O'Keefe, B. R., Seeberger, P. H., Gronenborn, A. M. & Boyd, M. R. (2002). Multisite and multivalent binding between cyanovirin-N and branched oligomannosides: calorimetric and NMR characterization. *Chem Biol* **9**, 1109-18.
32. Shenoy, S. R., O'Keefe, B. R., Bolmstedt, A. J., Cartner, L. K. & Boyd, M. R. (2001). Selective interactions of the human immunodeficiency virus-inactivating protein cyanovirin-N with high-mannose oligosaccharides on gp120 and other glycoproteins. *J Pharmacol Exp Ther* **297**, 704-10.
33. Weis, W. I., Taylor, M. E. & Drickamer, K. (1998). The C-type lectin superfamily in the immune system. *Immunol Rev* **163**, 19-34.
34. Fromme, R., Katiliene, Z., Fromme, P. & Ghirlanda, G. (2008). Conformational gating of dimannose binding to the antiviral protein cyanovirin revealed from the crystal structure at 1.35 Å resolution. *Protein Sci* **17**, 939-44.
35. Fromme, R., Katiliene, Z., Giomarelli, B., Bogani, F., Mc Mahon, J., Mori, T., Fromme, P. & Ghirlanda, G. (2007). A monovalent mutant of cyanovirin-N provides insight into the role of multiple interactions with gp120 for antiviral activity. *Biochemistry* **46**, 9199-207.
36. Matei, E., Furey, W. & Gronenborn, A. M. (2008). Solution and crystal structures of a sugar binding site mutant of cyanovirin-N: no evidence of domain swapping. *Structure* **16**, 1183-94.

37. Tsai, C. C., Emau, P., Jiang, Y., Agy, M. B., Shattock, R. J., Schmidt, A., Morton, W. R., Gustafson, K. R. & Boyd, M. R. (2004). Cyanovirin-N inhibits AIDS virus infections in vaginal transmission models. *AIDS Res Hum Retroviruses* **20**, 11-8.
38. Tsai, C. C., Emau, P., Jiang, Y., Tian, B., Morton, W. R., Gustafson, K. R. & Boyd, M. R. (2003). Cyanovirin-N gel as a topical microbicide prevents rectal transmission of SHIV89.6P in macaques. *AIDS Res Hum Retroviruses* **19**, 535-41.
39. Balzarini, J., Van Laethem, K., Peumans, W. J., Van Damme, E. J., Bolmstedt, A., Gago, F. & Schols, D. (2006). Mutational pathways, resistance profile, and side effects of cyanovirin relative to human immunodeficiency virus type 1 strains with N-glycan deletions in their gp120 envelopes. *J Virol* **80**, 8411-21.
40. O'Keefe, B. R., Smee, D. F., Turpin, J. A., Saucedo, C. J., Gustafson, K. R., Mori, T., Blakeslee, D., Buckheit, R. & Boyd, M. R. (2003). Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. *Antimicrob Agents Chemother* **47**, 2518-25.
41. Smee, D. F., Wandersee, M. K., Checketts, M. B., O'Keefe, B. R., Saucedo, C., Boyd, M. R., Mishin, V. P. & Gubareva, L. V. (2007). Influenza A (H1N1) virus resistance to cyanovirin-N arises naturally during adaptation to mice and by passage in cell culture in the presence of the inhibitor. *Antivir Chem Chemother* **18**, 317-27.
42. Barrientos, L. G. & Gronenborn, A. M. (2005). The highly specific carbohydrate-binding protein cyanovirin-N: structure, anti-HIV/Ebola activity and possibilities for therapy. *Mini Rev Med Chem* **5**, 21-31.
43. Barrientos, L. G., O'Keefe, B. R., Bray, M., Sanchez, A., Gronenborn, A. M. & Boyd, M. R. (2003). Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral Res* **58**, 47-56.
44. Helle, F., Wychowski, C., Vu-Dac, N., Gustafson, K. R., Voisset, C. & Dubuisson, J. (2006). Cyanovirin-N inhibits hepatitis C virus entry by binding to envelope protein glycans. *J Biol Chem* **281**, 25177-83.
45. Dey, B., Lerner, D. L., Lusso, P., Boyd, M. R., Elder, J. H. & Berger, E. A. (2000). Multiple antiviral activities of cyanovirin-N: blocking of human immunodeficiency virus type 1 gp120 interaction with CD4 and coreceptor and inhibition of diverse enveloped viruses. *J Virol* **74**, 4562-9.
46. Smee, D. F., Bailey, K. W., Wong, M. H., O'Keefe, B. R., Gustafson, K. R., Mishin, V. P. & Gubareva, L. V. (2008). Treatment of influenza A (H1N1) virus infections in mice and ferrets with cyanovirin-N. *Antiviral Res* **80**, 266-71.
47. O'Keefe, B. R., Shenoy, S. R., Xie, D., Zhang, W., Muschik, J. M., Currens, M. J., Chaiken, I. & Boyd, M. R. (2000). Analysis of the interaction between the HIV-

inactivating protein cyanovirin-N and soluble forms of the envelope glycoproteins gp120 and gp41. *Mol Pharmacol* **58**, 982-92.

48. Hu, Q., Mahmood, N. & Shattock, R. J. (2007). High-mannose-specific deglycosylation of HIV-1 gp120 induced by resistance to cyanovirin-N and the impact on antibody neutralization. *Virology* **368**, 145-54.
49. Witvrouw, M., Fikkert, V., Hantson, A., Pannecouque, C., O'Keefe B, R., McMahon, J., Stamatatos, L., de Clercq, E. & Bolmstedt, A. (2005). Resistance of human immunodeficiency virus type 1 to the high-mannose binding agents cyanovirin N and concanavalin A. *J Virol* **79**, 7777-84.
50. Mariner, J. M., McMahon, J. B., O'Keefe, B. R., Nagashima, K. & Boyd, M. R. (1998). The HIV-inactivating protein, cyanovirin-N, does not block gp120-mediated virus-to-cell binding. *Biochem Biophys Res Commun* **248**, 841-5.
51. Barrientos, L. G., Louis, J. M., Botos, I., Mori, T., Han, Z., O'Keefe, B. R., Boyd, M. R., Wlodawer, A. & Gronenborn, A. M. (2002). The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure* **10**, 673-86.
52. Bewley, C. A., Gustafson, K. R., Boyd, M. R., Covell, D. G., Bax, A., Clore, G. M. & Gronenborn, A. M. (1998). Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nat Struct Biol* **5**, 571-8.
53. Yang, F., Bewley, C. A., Louis, J. M., Gustafson, K. R., Boyd, M. R., Gronenborn, A. M., Clore, G. M. & Wlodawer, A. (1999). Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J Mol Biol* **288**, 403-12.
54. Botos, I. & Wlodawer, A. (2003). Cyanovirin-N: a sugar-binding antiviral protein with a new twist. *Cell Mol Life Sci* **60**, 277-87.
55. Kelley, B. S., Chang, L. C. & Bewley, C. A. (2002). Engineering an obligate domain-swapped dimer of cyanovirin-N with enhanced anti-HIV activity. *J Am Chem Soc* **124**, 3210-1.
56. Barrientos, L. G., Lasala, F., Delgado, R., Sanchez, A. & Gronenborn, A. M. (2004). Flipping the switch from monomeric to dimeric CV-N has little effect on antiviral activity. *Structure* **12**, 1799-807.
57. Esser, M. T., Mori, T., Mondor, I., Sattentau, Q. J., Dey, B., Berger, E. A., Boyd, M. R. & Lifson, J. D. (1999). Cyanovirin-N binds to gp120 to interfere with CD4-dependent human immunodeficiency virus type 1 virion binding, fusion, and infectivity but does not affect the CD4 binding site on gp120 or soluble CD4-induced conformational changes in gp120. *J Virol* **73**, 4360-71.

58. Bringans, S. D., O'Keefe, B. R., Bray, M., Whitehouse, C. A. & Boyd, M. R. (2004). Development of a fluorescent microplate assay for determining cyanovirin-N levels in plasma. *Anal Bioanal Chem* **380**, 269-74.
59. Zappe, H., Snell, M. E. & Bossard, M. J. (2008). PEGylation of cyanovirin-N, an entry inhibitor of HIV. *Adv Drug Deliv Rev* **60**, 79-87.
60. Sugrue, R. J. (2007). Viruses and glycosylation: an overview. *Methods Mol Biol* **379**, 1-13.
61. Mori, T., Barrientos, L. G., Han, Z., Gronenborn, A. M., Turpin, J. A. & Boyd, M. R. (2002). Functional homologs of cyanovirin-N amenable to mass production in prokaryotic and eukaryotic hosts. *Protein Expr Purif* **26**, 42-9.
62. Mori, T., Gustafson, K. R., Pannell, L. K., Shoemaker, R. H., Wu, L., McMahon, J. B. & Boyd, M. R. (1998). Recombinant production of cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. *Protein Expr Purif* **12**, 151-8.
63. Colleluori, D. M., Tien, D., Kang, F., Pagliei, T., Kuss, R., McCormick, T., Watson, K., McFadden, K., Chaiken, I., Buckheit, R. W., Jr. & Romano, J. W. (2005). Expression, purification, and characterization of recombinant cyanovirin-N for vaginal anti-HIV microbicide development. *Protein Expr Purif* **39**, 229-36.
64. Wei, X., Decker, J. M., Wang, S., Hui, H., Kappes, J. C., Wu, X., Salazar-Gonzalez, J. F., Salazar, M. G., Kilby, J. M., Saag, M. S., Komarova, N. L., Nowak, M. A., Hahn, B. H., Kwong, P. D. & Shaw, G. M. (2003). Antibody neutralization and escape by HIV-1. *Nature* **422**, 307-12.
65. Balzarini, J., Van Laethem, K., Hatse, S., Froeyen, M., Peumans, W., Van Damme, E. & Schols, D. (2005). Carbohydrate-binding agents cause deletions of highly conserved glycosylation sites in HIV GP120: a new therapeutic concept to hit the achilles heel of HIV. *J Biol Chem* **280**, 41005-14.
66. Balzarini, J., Van Laethem, K., Hatse, S., Froeyen, M., Van Damme, E., Bolmstedt, A., Peumans, W., De Clercq, E. & Schols, D. (2005). Marked depletion of glycosylation sites in HIV-1 gp120 under selection pressure by the mannose-specific plant lectins of *Hippeastrum hybrid* and *Galanthus nivalis*. *Mol Pharmacol* **67**, 1556-65.
67. Balzarini, J. (2007). Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. *Nat Rev Microbiol* **5**, 583-97.
68. Kang, S. M., Quan, F. S., Huang, C., Guo, L., Ye, L., Yang, C. & Compans, R. W. (2005). Modified HIV envelope proteins with enhanced binding to neutralizing monoclonal antibodies. *Virology* **331**, 20-32.

69. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648-59.
70. Malenbaum, S. E., Yang, D., Cavacini, L., Posner, M., Robinson, J. & Cheng-Mayer, C. (2000). The N-terminal V3 loop glycan modulates the interaction of clade A and B human immunodeficiency virus type 1 envelopes with CD4 and chemokine receptors. *J Virol* **74**, 11008-16.
71. Reitter, J. N., Means, R. E. & Desrosiers, R. C. (1998). A role for carbohydrates in immune evasion in AIDS. *Nat Med* **4**, 679-84.
72. Li, Y., Luo, L., Rasool, N. & Kang, C. Y. (1993). Glycosylation is necessary for the correct folding of human immunodeficiency virus gp120 in CD4 binding. *J Virol* **67**, 584-8.
73. Sjolander, S., Bolmstedt, A., Akerblom, L., Horal, P., Olofsson, S., Morein, B. & Sjolander, A. (1996). N-linked glycans in the CD4-binding domain of human immunodeficiency virus type 1 envelope glycoprotein gp160 are essential for the in vivo priming of T cells recognizing an epitope located in their vicinity. *Virology* **215**, 124-33.
74. Mori, T., Shoemaker, R. H., McMahon, J. B., Gulakowski, R. J., Gustafson, K. R. & Boyd, M. R. (1997). Construction and enhanced cytotoxicity of a [cyanovirin-N]-[Pseudomonas exotoxin] conjugate against human immunodeficiency virus-infected cells. *Biochem Biophys Res Commun* **239**, 884-8.
75. McFadden, K., Cocklin, S., Gopi, H., Baxter, S., Ajith, S., Mahmood, N., Shattock, R. & Chaiken, I. (2007). A recombinant allosteric lectin antagonist of HIV-1 envelope gp120 interactions. *Proteins* **67**, 617-29.
76. Bewley, C. A. (2001). Solution structure of a cyanovirin-N:Man alpha 1-2Man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* **9**, 931-40.
77. Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H., Boyd, M. R. & Wlodawer, A. (2002). Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem* **277**, 34336-42.

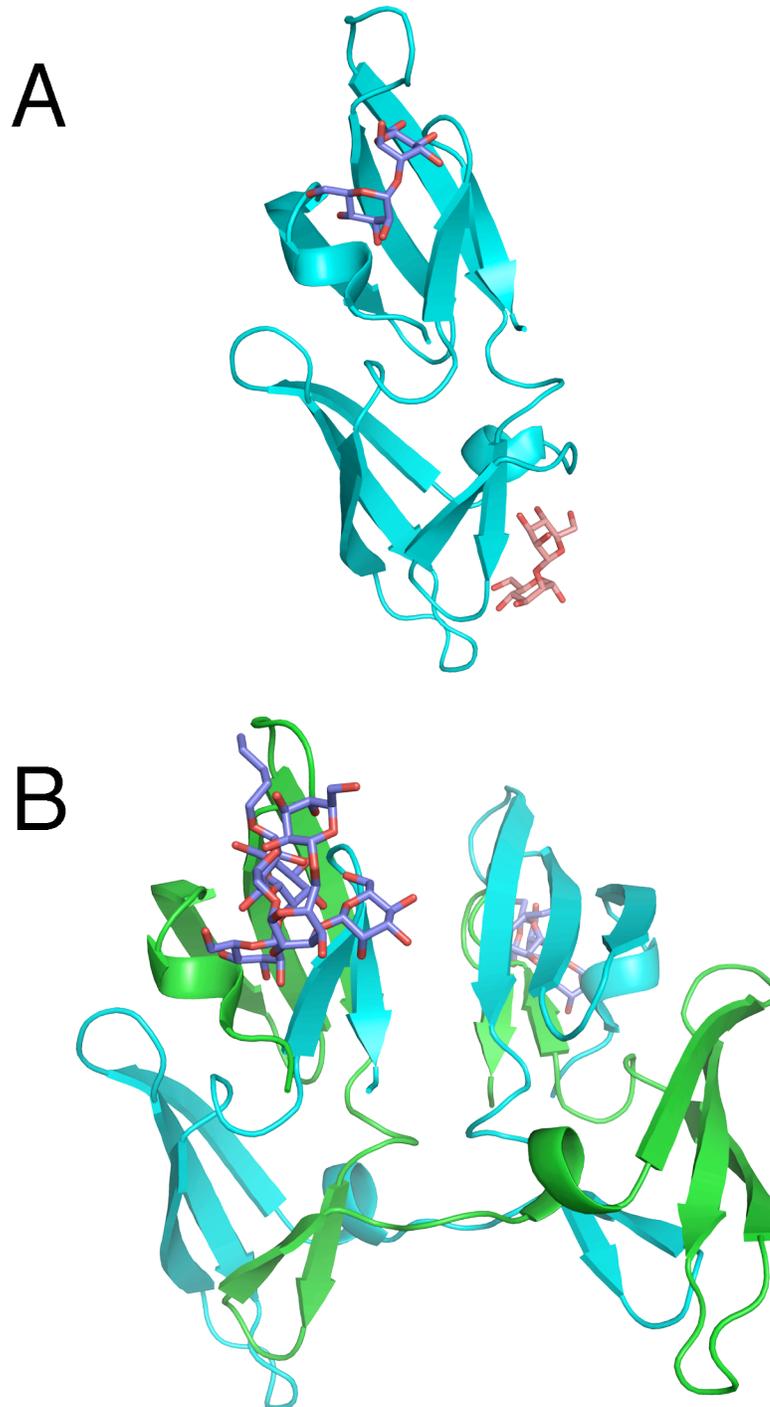


Figure 1-1. Wild-type (WT) CVN structures. In solution, WT CVN exists mainly as a monomer (A),⁷⁶ while it always crystallizes as a domain-swapped dimer (B).⁷⁷ CVN is shown in green and cyan ribbons to indicate protein chains. Carbohydrates bound in the high-affinity site are shown with orange carbons (present only in A), and carbohydrates bound in the low affinity site are shown with blue carbons (present in both A and B). The monomer and the left half of the dimer are in approximately the same orientation.

Chapter 2

Engineered cyanovirin-N oligomers show enhanced HIV neutralization

Abstract

Cyanovirin-N (CVN), a protein originally isolated from the cyanobacterium *Nostoc ellipsosporum*, has been shown to bind specifically to glycosylated gp120 on HIV particles, preventing viral fusion and neutralizing the virus. Here, we present dimeric and trimeric variants of CVN that display increased potency for neutralizing HIV-1. The molecules were expressed and purified to homogeneity and then assessed for their ability to prevent HIV-1 infection in a cell culture-based neutralization assay. We found that covalently linking two CVN monomers (CVN₂) through a flexible polypeptide linker decreased the concentration of protein at which 50% of the virus was neutralized (IC₅₀) up to ten-fold compared to wild-type CVN. The addition of a third CVN monomer (CVN₃), however, showed minimal further neutralization enhancement. Our data also suggest a linker-length dependence in the neutralization activity, although variability in the assays masks a definitive trend. In addition, CVN and dimeric variants displayed significant cross-clade and cross-strain reactivity against 33 strains of HIV-1 and neutralized most strains with decreased IC₅₀s compared to the most effective broadly neutralizing antibody tested.

The increased neutralization exhibited by these linked CVN variants provides a novel path to improve our understanding of how CVN prevents HIV infection, and the broad cross-strain reactivity holds promise for the future therapeutic utility of these and other engineered CVN variants.

Introduction

Treatment and prevention of HIV-1 have proven to be difficult and complex problems. Vaccines thus far have been unsuccessful in generating broadly reactive neutralizing antibodies that confer immunity to the virus, partly because of HIV-1's rapid mutation rate and partly because of the lack of epitopes on its envelope proteins. The HIV-1 envelope is composed mainly of two proteins: gp41 and gp120, which are products of a single precursor protein, gp160. gp41 contains a transmembrane region, which anchors the envelope protein to the membrane, as well as a region that interacts specifically with gp120; gp120 contains binding sites for the primary receptor, CD4, and coreceptors, CCR5 and CXCR4. These gp41-gp120 dimers form trimers on the surface of the virus, creating the envelope spikes.¹ While gp41 and gp120 by nature must contain some invariable regions, these conserved regions are typically masked or difficult to access. For example, the CD4 binding site is located in a cleft in the protein, allowing this region to evade the human immune system,^{2,3} and the binding sites for the coreceptors are revealed only after a conformational change induced by CD4 binding.⁴ Additionally, gp120 and gp41 are both heavily glycosylated, obscuring many potential epitopes. In fact, approximately 50% of the apparent molecular weight of gp120 is attributed to complex carbohydrates.⁵ This “silent face” of the HIV envelope is fairly resistant to the human immune system because it is generated by the host and often recognized as “self.”

In addition to the lack of functional epitopes on the envelope proteins of HIV-1, HIV is a retrovirus, and therefore has a very high mutation rate due to the error-prone reverse-transcriptase. Because of this rapid mutation rate, there is a great deal of variation in HIV viruses within an individual, between individuals, and geographically. A

successful treatment or vaccine must therefore induce a broadly neutralizing response such that it affects many strains and many clades. Thus far, however, only a few broadly neutralizing antibodies have been isolated that are capable of neutralizing primary isolates of HIV-1.⁶⁻¹⁰ To date, four broadly neutralizing antibodies have been discovered and extensively studied. The first, IgG1b12, recognizes a conserved, recessed area of gp120 that overlaps with the CD4 binding site.³ IgG1b12 is capable of neutralizing approximately half of HIV-1 strains tested, including some strains from each clade.^{6,7,9} The monoclonal antibodies 4E10 and 2F5 both recognize conserved areas of gp41, near the viral membrane,^{11,12} but differ in their abilities to neutralize HIV-1. 4E10 is the most broadly reactive of these neutralizing antibodies, neutralizing all 93 strains tested from 12 different clades. Unfortunately, although it is able to neutralize broadly, it does so with only modest potency.⁹ 2F5, on the other hand, is not capable of neutralizing most viruses from clade C and is effective against only approximately 50% of clade D viruses.^{7,9} The final anti-HIV antibody is 2G12. 2G12 differs from the other broadly neutralizing antibodies in that it recognizes the carbohydrates on gp120 and not the actual protein itself.¹³ In contrast to the standard “Y” structure of antibodies, the Fabs of 2G12 form a domain-swapped structure that allows it to bind two carbohydrate chains approximately 35 Å apart.¹⁴ 2G12 is mostly effective against viruses from clade B and exhibits limited or no neutralization of viruses from other clades.^{6,7,9}

Cyanovirin (CVN), a lectin from the cyanobacterium *Nostoc ellipsosporum*, is uniquely suited to play a role in HIV treatment and prevention.¹⁵ CVN, like the 2G12 antibody, binds specifically to α 1-2 oligomannose molecules,¹⁶ which are highly expressed on gp120, and neutralizes enveloped viruses including HIV,¹⁵ Ebola,¹⁷ and

influenza.¹⁸ Also like 2G12, CVN contains two carbohydrate binding sites per molecule, indicating there is a potential avidity effect upon binding.¹⁶ CVN is also distinctive in its small size. The 11-kDa protein is much smaller than even a single-chain Fv fragment (scFv) and therefore has the ability to bind to areas on gp120 that are sterically occluded from scFv, Fab, or IgG binding.¹⁹ Additionally, unlike 2G12, which is specific to carbohydrates on specific residues,²⁰ CVN is specific only to the type of linkage and therefore less sensitive to escape mutations that eliminate a single glycosylation site. In fact, glycosylation on gp120 has been shown to increase over the course of infection²⁰ and act as a mechanism for escape from neutralizing antibodies.²¹ CVN may be an optimal therapeutic in these cases due to its reaction to a broad range of high mannose carbohydrates.

We are interested in studying the effects of oligomerization on CVN to see whether increasing the number of binding sites and varying the distances between those sites increases the efficacy of neutralization for HIV-1 and other enveloped viruses. Previous studies of 2G12 indicated that higher order oligomers are more effective at neutralizing HIV.^{22,23} West *et al.* showed that natural dimers of 2G12 are up to 80-fold more potent than the monomer,²² and the oligomeric 2G12-IgM engineered antibody tested by Wolbank *et al.* exhibited up to 28-fold greater efficacy than 2G12-IgG.²³ Although CVN can also exist in a domain-swapped form,²⁴ it is unclear whether the domain-swapped dimer exhibits similar increases in neutralization to 2G12, since differing accounts have been published.^{25,26} By dimerizing or trimerizing CVN, we not only increase the number of binding sites and therefore potentially affect the avidity of

binding, we can also effectively increase the separation of binding sites and therefore possibly crosslink glycosylation sites that are separated by a greater distance.

In addition to determining the effect of oligomerization on CVN efficacy, we are interested in testing these variants against a large number of HIV-1 strains to ascertain their cross-clade neutralization. Toward these goals, we have successfully created linked dimers (CVN₂) and trimers (CVN₃) of CVN and assayed their HIV neutralization function against 33 viruses from three clades. These variants contain a polypeptide linker of varying length (L0 to L20) connecting the C-terminus of one CVN to the N-terminus of another (Table 2-1, Figure 2-1).

Methods

Construct generation. The gene for wild-type (WT) CVN was constructed using a recursive PCR method with 40-mer synthesized oligos,²⁷ and cloned into the NdeI and BamHI sites of pET11a. The protein contained an N-terminal 6-histidine purification tag followed by a Factor Xa protease cleavage site. CVN₂ L5 and CVN₂ L10 were constructed using PCR-based cloning to insert a tandem repeat of the WT CVN gene into the WT plasmid. The CVN₃ L5 gene was created by inserting an *Escherichia coli*-optimized WT CVN DNA sequence between the two existing copies of the WT gene in CVN₂ L5. Other dimeric and trimeric genes of varying linker lengths were constructed using the QuikChange Site-Directed Mutagenesis Kit to insert or delete codons corresponding to linker amino acids (Stratagene). All constructs were verified through DNA sequencing and restriction analysis to ensure the correct sequence and number of CVN repeats.

Expression and purification. WT CVN and all oligomeric variants were expressed in BL21(DE3) *E. coli* cells in LB including ampicillin. The cultures were induced with 1 mM IPTG when the cells reached mid-log and grown for an additional 3-5 hours at 37°C. The harvested cells were lysed using an EmulsiFlex-C5 (Avestin, Inc.) and the insoluble fraction was resuspended in buffer containing 6 M GnHCl and 10 mM imidazole and centrifuged to remove debris. The solubilized CVN was then purified using a denatured protocol on a Ni-NTA gravity column (Qiagen). The protein was eluted in buffer containing 6 M GnHCl and 250 mM imidazole and transferred to dialysis tubing with a MWCO of 5,000 Da. WT CVN and all variants were refolded by dialyzing the Ni-NTA eluate against native buffer overnight at room temperature.²⁸ Following refolding, the solution was filtered to remove any precipitant and concentrated using Amicon 5000 MWCO centrifugal concentrators to approximately 0.5 mL (Millipore). The proteins were then additionally purified on a Superdex-75 column and eluted in 25 mM sodium phosphate pH 7.4, 150 mM NaCl. Pure protein was concentrated or stored as eluted at 4°C. Attempts to cleave the His-tag using Factor Xa were unsuccessful under several conditions, so the tag was left intact for the studies described here.

Amino acid analysis was performed on WT CVN, CVN₂ L5, CVN₂ L10, CVN₃ L5 and CVN₃ L10 to determine extinction coefficients at 280 nm (Jinny Johnson, Texas A&M University). These experimentally determined extinction coefficients (WT: 10471 M⁻¹cm⁻¹, CVN₂S: 20800 M⁻¹cm⁻¹, CVN₃S: 32000 M⁻¹cm⁻¹) were used to calculate the protein concentration.

Surface plasmon resonance (SPR) assays. SPR (Biacore) experiments were performed on a T100 instrument (Biacore). Approximately 30 response units (RUs) of WT CVN were immobilized on flow cells 2 through 4 of a CM5 chip through standard amine coupling. Flow cell 1 was reserved as a control. All assays were conducted in HBS-EP buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 0.0005% v/v Surfactant P20, 1 mM EDTA; Biacore). Various analytes were injected over the surface for 60 seconds at a flow rate of 30 μ L/min. The chip was regenerated with two pulses of 50 mM NaOH. In some cases, complete regeneration was not achieved and a new chip was created. The data were analyzed for binding or lack of binding based on the sensorgram.

Cell viability assays. Cellular toxicity was assayed using an XTT cell proliferation kit (MD Biosciences). This assay is based on the observation that live cells can reduce XTT and other tetrazolium salts, resulting in a colored solution. Cells that are not metabolically active, however, are not able to reduce XTT and therefore there is no color change after the addition of the reagent. WT CVN or a variant protein was incubated with approximately 5000 Tzm-B1 cells in D-MEM high glucose medium with L-glutamine, sodium pyruvate, 50 μ g/mL gentamicin, and 10% heat inactivated BSA at 37°C in a CO₂ incubator for 48 hours. The prepared XTT reagent was added to each well and incubated for an additional two hours. The plates were shaken gently and the absorbance at 450 nm (the absorbance of the reduced XTT product) was measured on a Safire² plate reader (Tecan). Each protein was assayed in triplicate, and the average absorbance values were used to determine the percent of cells that were viable (*%Viable*) using Equation 2-1, where A_{CVN} is the absorbance of the well containing cells and CVN, $A_{negative}$ is the average

absorbance of wells containing no cells or protein, and A_{cell} is the average absorbance of wells containing cells but no CVN. The percent of viable cells was plotted as a function of the CVN concentration.

$$\%Viable = \frac{A_{CVN} - A_{negative}}{A_{cell} - A_{negative}} * 100 \quad (2-1)$$

HIV neutralization assays. HIV neutralization assays were performed by Priyanthi Peiris, a technician in Prof. Pamela Bjorkman's lab, according to the methods by Li *et al.*⁶ The assays used Tzm-B1 cells, a HeLa cell line that expresses CD4 and the HIV-coreceptors CCR5 and CXCR4 and contains a viral Tat-induced luciferase reporter gene. Only when infected by the HIV-1 pseudovirus will these cells express luciferase, allowing a high-throughput measurement of neutralization. HIV-1 pseudovirus particles from pseudotyped primary virus strains were prepared as described.^{6,21} The SC422661.8 strain (clade B) was used for all assays unless otherwise noted. One column of 8 wells in a 96-well plate contained cells but no virus and was used to determine the background level of luminescence (cell control). Additionally, one column contained cells and virus but no inhibitory compound, acting as both a positive control and a maximal signal of infection (viral control). In the remaining wells, approximately 250 TCID₅₀ of virus was incubated with varying amounts of CVN or CVN variant in triplicate for one hour at 37°C. Each plate contained WT in triplicate as an internal control. Typically eight three-fold dilutions starting with 200 nM protein were tested to create a neutralization curve. Approximately 10,000 freshly trypsinized cells were added to each well and the plate was incubated for 48 hours. The cells were then lysed using Bright Glo Luciferase Assay

Buffer (Promega), which was diluted 4X. The lysate was then transferred to a new plate and the luminescence was measured on a Victor³ Multilabel Counter (PerkinElmer, Inc.).

To determine the IC_{50} of neutralization, the luminescence corresponding to a given protein concentration was first averaged across the three replicates, then the percent neutralization (*%Neutralization*) was calculated based on Equation 2-2, where RLU is the average relative luminescence for a given concentration, CC is the average luminescence from the cell control wells, and VC is the average luminescence from the viral control wells. The percent of virus neutralized was then plotted as a function of neutralizing protein in Kaleidograph (Synergy Software) and fitted to Equation 2-3, where IC_{50} is the concentration of CVN at which 50% of the virus is neutralized and C is the concentration of CVN. The reported error is the error associated with the curve fit to the experimental data.

$$\%Neutralization = \left(1 - \frac{RLU - CC}{VC - CC}\right) * 100 \quad (2-2)$$

$$\%Neutralization = \frac{100}{1 + \frac{IC_{50}}{C}} \quad (2-3)$$

To minimize the plate-to-plate deviations between assays, we normalized each variant's IC_{50} compared to WT IC_{50} on the same plate. These data are presented as “ IC_{50} : fold lower than WT” and were calculated by dividing the IC_{50} for WT by the IC_{50} for the variant. Each variant was independently tested between one and five times. Error bars were calculated by propagating the error from the WT and variant curve fits as well as multiple trials if applicable, according to standard methods.

In addition to the assays done here at Caltech, WT and two dimer variants were tested against multiple envelopes from various HIV-1 clades by Professor Michael Seaman's lab at Harvard Medical School through the Collaboration for AIDS Vaccine Discovery (CAVD) Neutralizing Antibody Laboratory. These assays were performed according to the same protocol described above, but 33 pseudoviruses from clades A, B, and C were tested to determine the cross-clade reactivity. The data were analyzed as described above.

Results

Dimer and trimer variants. To directly assay the effects of dimerization and trimerization on the activity of CVN, we generated proteins consisting of two or three tandem repeats of CVN (Figure 2-1). The resulting proteins had one copy of the protein linked through its C-terminus to the N-terminus of the next copy through a flexible polypeptide linker encoded in the gene. In this study, we tested dimers (CVN₂S) with 14 different linkers ranging from 0 to 20 amino acids (all Gly or Ser) for their ability to neutralize HIV in a cell-based assay. In addition, we assayed trimers (CVN₃S) with three linkers comprised of 0, 5, or 10 amino acids (Table 2-1).

Expression and purification. WT and all CVN variants were expressed into inclusion bodies at 37°C using standard *E. coli* expression protocols. After solubilizing the proteins in 6 M GnHCl and running a Ni-NTA purification step, the proteins were refolded by dialyzing against native buffer. Most of the refolded protein solutions, including those for WT, CVN₂ L0, CVN₂ L10, and CVN₃ variants, had little or no

precipitation after dialysis, indicating the conditions were sufficient for refolding without favoring aggregation. Some variants, however, including CVN₂ L1 and CVN₂ L3, experienced significant precipitation during the refolding step. These solutions were filtered before continuing with the purification protocol. Gel filtration was performed on the refolded proteins to separate the desired species (WT: monomer; CVN₂S: dimer; CVN₃S: trimer) from higher molecular weight species including domain-swapped dimers or tetramers and aggregates. WT CVN eluted at approximately 0.59 CV, CVN₂S eluted at approximately 0.54 CV, and CVN₃S eluted at approximately 0.50 on the gel filtration column. Reinjection of purified sample indicated the protein was stable in its purified oligomeric state for weeks to months when stored at 4°C.

Surface plasmon resonance (Biacore) assays. To assess the efficacy of the refolding protocol, WT CVN was assayed for its ability to bind gp120. WT CVN was immobilized to a Biacore chip and 100 nM gp120HxBc2 was flowed over. Significant binding was observed at all surface densities tested and the binding was virtually irreversible (data not shown). Various regeneration conditions including low pH, high pH, and high concentrations of NaCl were used, but the chip was never fully regenerated. We therefore concluded that WT CVN was properly folded and able to bind specifically to gp120, but kinetic and binding constants could not be obtained using this method.

We also tested for domain-swapping and aggregation on the surface. WT CVN was immobilized and WT and various CVN₂ proteins were analyzed for binding. No binding was observed for WT, CVN₂ L0, CVN₂ L1 or CVN₂ L10 (data not shown).

These proteins did not aggregate on the surface or bind to WT CVN. We were not able to observe any intermolecular domain-swapping under these conditions.

HIV neutralization assays. WT CVN was tested for its ability to neutralize HIV pseudovirus in cell culture. The IC₅₀ of CVN is reported to be in the low nanomolar range.^{25,29-31} In our assays, WT CVN neutralized HIV envelope SC422661.8 with IC₅₀s between 2 and 14 nM over 16 independent trials, consistent with published values. Because of the variation we saw in these experiments and to minimize any plate-to-plate deviations due to incubation conditions, viral particle preparation, or other systematic differences, we report all variant data relative to the WT IC₅₀ from the same 96-well plate.

Dimeric CVN₂ proteins were also tested for their ability to neutralize HIV strain SC422661.8 (Figure 2-2). All variants displayed IC₅₀s lower than WT CVN, showing enhanced neutralizing compared to WT. For CVN₂ L0 and CVN₂ L10, the increase in efficacy is nearly ten-fold. While there appears to be a linker-length dependence to the data, this may be an artifact of the complicated biological assay and not a reflection of actual differences in activities.

To test the hypothesis that more binding sites make better neutralizing variants, we created and assayed trimeric CVN₃ molecules. Similar to the CVN₂ results, these variants were significantly more effective at HIV neutralization than WT. However, compared to the CVN₂ variant, adding an additional CVN repeat did not increase the efficacy of HIV neutralization, and the three variants tested (CVN₃ L0, CVN₃ L5, and

CVN₃ L10) did not have significantly different activities from their CVN₂ counterparts (Figure 2-3).

After identifying CVN₂ L0 and CVN₂ L10 as the best performing oligomeric variants against the SC422661.8 strain, we assayed these proteins as well as WT for their cross-clade reactivity. The proteins were tested against a total of 33 viruses from three clades of HIV (Table 2-2). One of the most difficult obstacles in developing HIV neutralizing monoclonal antibodies (NAbs) is their lack of cross-clade reactivity. Most NAbs effectively neutralize viruses from one or two clades, but often are not effective against other clades. It is important for a potential therapeutic to be effective against as broad a range of viruses as possible. In the case of WT CVN and the dimeric mutants, all 33 of the HIV pseudoviruses were neutralized with IC₅₀s less than 300 nM (Figure 2-4, Table 2-2). Only the 4E10 NAb was as cross-clade reactive, while 2G12 and 2F5 were not effective at neutralizing clade C viruses, and IgG1b12 was not effective against clade A viruses.

In addition to the broad cross-clade reactivity of WT CVN and the CVN₂ variants, we were interested in the overall efficacy of the CVN proteins as compared to the NAbs. To simplify the analysis, we chose the NAb with the lowest IC₅₀ for each individual envelope and compared WT CVN, CVN₂ L0, and CVN₂ L10 to that variant (Figure 2-5). This comparison therefore indicates how effective the CVNs are against the best broadly neutralizing antibody for each strain. For many strains, WT CVN is less effective than the best NAb, as indicated by bars with negative values. However, by dimerizing the protein, we increased the efficacy of neutralization and generated variants that exhibit better neutralization than the best NAb against a given HIV strain. In fact, CVN₂ L0 is better

than the best NAb for every virus tested except for 4 out of 31 cases (2 of the 33 viruses did not have corresponding NAb data). Although it is only moderately more effective against some strains, CVN₂ L0 is at least 5-fold better than the best NAb against 19 out of 31 envelopes and at least 10-fold better against 11 envelopes. Additionally, CVN₂ L0 has an IC₅₀ 215-fold lower than the NAbs in one case (clade B, TRJO4551.58). Through dimerization, we have created a variant that is not only broadly cross-clade reactive, but is also more effective at neutralizing HIV-1 than the commonly studied NAbs.

Similarly to the results from neutralization assays on SC422661.8, the cross-clade data indicate that the dimerized variants are significantly more effective at neutralizing various strains of HIV-1 than WT CVN (Figure 2-6). Specifically, CVN₂ L0 neutralized with a lower IC₅₀ than CVN₂ L10 in 32 out of 33 cases, and CVN₂ L10 neutralized with a lower IC₅₀ than WT in all 33 cases.

Cell viability assays. An alternative explanation to our HIV neutralization assay data is that CVN is not in fact neutralizing HIV, but instead killing the host cell. In this case, the cell would not express luciferase upon infection because its cellular machinery would be nonfunctional. To test this hypothesis, we checked WT CVN, CVN₂ L5, and CVN₂ L10 for toxicity against Tzm-B1 cells using an XTT cell proliferation assay. We tested concentrations up to 25-fold higher than the highest concentration used in the HIV neutralization assays. Our data indicate that at the concentrations relevant for the neutralization assays, CVN and the CVN₂ variants are not toxic (Figure 2-7). We did, however, see limited toxicity both by the XTT assay as well as visual inspection of the cells at protein concentrations above 1 μ M. This result is consistent with published

reports^{15,32} and indicates the activity we see against HIV is in fact neutralization of the virus and not toxicity to the host cell.

Discussion

We successfully created dimeric and trimeric CVN variants that have enhanced anti-HIV activity compared to WT CVN. These variants show excellent cross-clade and cross-strain reactivity and are more effective at neutralizing HIV than the most broadly neutralizing HIV antibodies.

Although the dimeric and trimeric variants show significant improvement in HIV neutralization, the mechanism responsible for the enhancement has thus far proved elusive. We hypothesize that differences in domain swapping may lead to an increase in efficacy. Previous reports are divided about whether the domain-swapped form of WT CVN is more active than the monomeric form.^{25,26} However, because of the meta-stable state of domain-swapped WT, it is difficult to assay the dimerized form in current biological assays. Our variants, however, because they are covalently linked at their termini, have a much higher local concentration of CVN and therefore may be more stable as a domain-swapped dimer, even at physiological temperatures. Additionally, for variants with short linker lengths, the link may force a domain-swapped structure and may sterically hinder a monomeric-like form. Because the tandem CVN repeats are covalently attached, however, it is difficult to determine whether the molecule is in a monomeric-like dimer form or a domain-swapped dimer (Figure 2-1). We can effectively remove any inter-molecularly domain-swapped protein using the gel filtration purification step, but intra-molecularly domain-swapped variants are difficult to

distinguish using standard purification protocols. We plan to investigate the domain-swapped nature of the variants to determine whether this may play a role in the increased activity. NMR-based experiments may allow us to see long-range interactions that are consistent with domain-swapping. Alternatively, we could introduce a protease cleavage site in the linker and determine whether after cleavage the protein is still dimeric or shifts to a monomeric molecular weight.

In addition to potential differences in domain swapping, the simple increase in carbohydrate binding sites may increase the avidity of the CVN-gp120 interaction. WT CVN itself has a very high affinity for gp120,^{15,29} but an increase in avidity in the CVN₂ variants may provide an extra force to prevent possible dissociation and escape of the virus. An alternate mechanism for increased neutralization is that the CVN₂s, with binding sites that are further apart than in WT CVN, are able to crosslink glycosylation sites on a single gp120 or crosslink multiple gp120 subunits on an envelope spike or, less likely, multiple spikes. This crosslinking would sterically hinder more gp120 subunits from binding to CD4 than would be blocked by WT CVN, thus decreasing the IC₅₀. An interesting note is that in the domain-swapped structure of WT CVN, every pair of carbohydrate binding sites is approximately 30 to 40 Å apart (Figure 2-8). The neutralizing antibody 2G12, which also binds the glycosylation site on gp120 and is also domain-swapped, has carbohydrate binding sites that are also approximately 35 Å apart.¹⁴ Perhaps by stabilizing the domain-swapped structure of CVN, the carbohydrate binding sites of the CVN₂ variants are optimally positioned to interact with gp120 and neutralize the virus.

While addition of a second CVN molecule increases the efficacy of HIV neutralization significantly, the addition of a third CVN repeat (CVN₃) does not significantly increase it further. Although the mechanism for enhanced activity is not yet understood, if increased domain-swapping is involved, an unpaired, third CVN may not significantly increase the neutralization. To test this hypothesis, tetrameric CVN molecules can be engineered and tested for increased HIV neutralization. Alternatively, due to the close proximity of the N- and C-termini in the WT structure and their proximity to the low affinity carbohydrate binding site, the third CVN molecule may sterically prohibit access to some of the carbohydrate binding sites in the molecule, rendering those sites nonfunctional and therefore not conveying any additional effect.

WT CVN and the CVN₂ molecules show excellent cross-clade and cross-strain reactivity. This property is promising for the development of these or other variants for therapeutic use as they can be used potentially throughout the world. In addition, CVN variants could be used in combination therapy to direct gp120 evolution toward decreased glycosylation. Glycosylation itself has been shown to be important in the folding and function of viral glycoproteins³³ and in the case of HIV, deglycosylation of gp120 diminishes the binding to CD4, making the virus less infective.^{34,35} Alternatively, deglycosylation of gp120 could merely reveal more protein epitopes that can be recognized by the adaptive immune system, allowing our own bodies to fight off infection more effectively.

Although these initial results are promising, more work needs to be done to elucidate a mechanism for the increased efficacy that we see for the dimeric and trimeric variants. It is probable that multiple mechanisms are at work to affect the neutralization

activity. Toward this goal, we have solved crystal structures of CVN₂ L0, CVN₂ L1, and CVN₂ L10 to determine whether any structural differences can account for changes in activity (Chapter 3 of this thesis). We also hope to employ a biophysical-based assay such as surface plasmon resonance or isothermal calorimetry to determine whether there are specific differences in the carbohydrate binding affinities between the CVN₂ variants and WT that explain the differences in HIV neutralization.

Additionally, we are testing our variants against other enveloped viruses including influenza, Andes virus (a highly lethal hantavirus), and vaccinia virus (a small pox model) to determine whether the increase in efficacy for the dimeric variants applies to all or most enveloped viruses or is specific to HIV.

Acknowledgements

I would like to thank Priyanthi Peiris for her time and expertise in running the majority of the HIV neutralization assays presented here as well as for helpful discussions. I would also like to express gratitude to Dr. Anthony West for organizing the collaboration with Prof. Michael Seaman and the CAVD Neutralizing Antibody Laboratory and thank Prof. Michael Seaman and his lab for assaying our variants against their wide range of viral envelopes. John Yong generated many of the CVN₂ constructs and expressed and purified them during a rotation in the Mayo lab. Noreen Tiangco provided the gp120 used in the Biacore studies. The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: SVPB8 (Drs. David Montefiori and Feng Gao); pSG3^{Δenv} (Drs. John C. Kappes and Xiaoyun Wu); Tzm-B1 cells (Dr. John C. Kappes, Dr. Xiaoyun Wu, and Tranzyme, Inc.)

References

1. Zhu, P., Liu, J., Bess, J., Jr., Chertova, E., Lifson, J. D., Grise, H., Ofek, G. A., Taylor, K. A. & Roux, K. H. (2006). Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441**, 847-52.
2. Burton, D. R., Stanfield, R. L. & Wilson, I. A. (2005). Antibody vs. HIV in a clash of evolutionary titans. *Proc Natl Acad Sci U S A* **102**, 14943-8.
3. Zhou, T., Xu, L., Dey, B., Hessel, A. J., Van Ryk, D., Xiang, S. H., Yang, X., Zhang, M. Y., Zwick, M. B., Arthos, J., Burton, D. R., Dimitrov, D. S., Sodroski, J., Wyatt, R., Nabel, G. J. & Kwong, P. D. (2007). Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**, 732-7.
4. Sattentau, Q. J. & Moore, J. P. (1991). Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J Exp Med* **174**, 407-15.
5. Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. J., Thomas, J. N. & Gregory, T. J. (1990). Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem* **265**, 10373-82.
6. Li, M., Gao, F., Mascola, J. R., Stamatatos, L., Polonis, V. R., Koutsoukos, M., Voss, G., Goepfert, P., Gilbert, P., Greene, K. M., Bilska, M., Kothe, D. L., Salazar-Gonzalez, J. F., Wei, X. P., Decker, J. M., Hahn, B. H. & Montefiori, D. C. (2005). Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* **79**, 10108-10125.
7. Li, M., Salazar-Gonzalez, J. F., Derdeyn, C. A., Morris, L., Williamson, C., Robinson, J. E., Decker, J. M., Li, Y., Salazar, M. G., Polonis, V. R., Mlisana, K., Karim, S. A., Hong, K., Greene, K. M., Bilska, M., Zhou, J., Allen, S., Chomba, E., Mulenga, J., Vwalika, C., Gao, F., Zhang, M., Korber, B. T., Hunter, E., Hahn, B. H. & Montefiori, D. C. (2006). Genetic and neutralization properties of subtype C human immunodeficiency virus type 1 molecular env clones from acute and early heterosexually acquired infections in Southern Africa. *J Virol* **80**, 11776-90.
8. Trkola, A., Pomales, A. B., Yuan, H., Korber, B., Maddon, P. J., Allaway, G. P., Katinger, H., Barbas, C. F., 3rd, Burton, D. R., Ho, D. D. & et al. (1995). Cross-clade neutralization of primary isolates of human immunodeficiency virus type 1 by human monoclonal antibodies and tetrameric CD4-IgG. *J Virol* **69**, 6609-17.
9. Binley, J. M., Wrin, T., Korber, B., Zwick, M. B., Wang, M., Chappey, C., Stiegler, G., Kunert, R., Zolla-Pazner, S., Katinger, H., Petropoulos, C. J. & Burton, D. R. (2004). Comprehensive cross-clade neutralization analysis of a

panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *J Virol* **78**, 13232-52.

10. Karlsson Hedestam, G. B., Fouchier, R. A., Phogat, S., Burton, D. R., Sodroski, J. & Wyatt, R. T. (2008). The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. *Nat Rev Microbiol* **6**, 143-55.
11. Ofek, G., Tang, M., Sambor, A., Katinger, H., Mascola, J. R., Wyatt, R. & Kwong, P. D. (2004). Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. *J Virol* **78**, 10724-37.
12. Cardoso, R. M., Zwick, M. B., Stanfield, R. L., Kunert, R., Binley, J. M., Katinger, H., Burton, D. R. & Wilson, I. A. (2005). Broadly neutralizing anti-HIV antibody 4E10 recognizes a helical conformation of a highly conserved fusion-associated motif in gp41. *Immunity* **22**, 163-73.
13. Trkola, A., Purtscher, M., Muster, T., Ballaun, C., Buchacher, A., Sullivan, N., Srinivasan, K., Sodroski, J., Moore, J. P. & Katinger, H. (1996). Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol* **70**, 1100-8.
14. Calarese, D. A., Scanlan, C. N., Zwick, M. B., Deechongkit, S., Mimura, Y., Kunert, R., Zhu, P., Wormald, M. R., Stanfield, R. L., Roux, K. H., Kelly, J. W., Rudd, P. M., Dwek, R. A., Katinger, H., Burton, D. R. & Wilson, I. A. (2003). Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300**, 2065-71.
15. Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T., Gulakowski, R. J., Wu, L., Rivera, M. I., Laurencot, C. M., Currens, M. J., Cardellina, J. H., 2nd, Buckheit, R. W., Jr., Nara, P. L., Pannell, L. K., Sowder, R. C., 2nd & Henderson, L. E. (1997). Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother* **41**, 1521-30.
16. Bewley, C. A. (2001). Solution structure of a cyanovirin-N:Man alpha 1-2Man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* **9**, 931-40.
17. Barrientos, L. G., O'Keefe, B. R., Bray, M., Sanchez, A., Gronenborn, A. M. & Boyd, M. R. (2003). Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral Res* **58**, 47-56.
18. O'Keefe, B. R., Smee, D. F., Turpin, J. A., Saucedo, C. J., Gustafson, K. R., Mori, T., Blakeslee, D., Buckheit, R. & Boyd, M. R. (2003). Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. *Antimicrob Agents Chemother* **47**, 2518-25.

19. Labrijn, A. F., Poignard, P., Raja, A., Zwick, M. B., Delgado, K., Franti, M., Binley, J., Vivona, V., Grundner, C., Huang, C. C., Venturi, M., Petropoulos, C. J., Wrin, T., Dimitrov, D. S., Robinson, J., Kwong, P. D., Wyatt, R. T., Sodroski, J. & Burton, D. R. (2003). Access of antibody molecules to the conserved coreceptor binding site on glycoprotein gp120 is sterically restricted on primary human immunodeficiency virus type 1. *J Virol* **77**, 10557-65.
20. Dacheux, L., Moreau, A., Ataman-Onal, Y., Biron, F., Verrier, B. & Barin, F. (2004). Evolutionary dynamics of the glycan shield of the human immunodeficiency virus envelope during natural infection and implications for exposure of the 2G12 epitope. *J Virol* **78**, 12625-37.
21. Wei, X. P., Decker, J. M., Wang, S. Y., Hui, H. X., Kappes, J. C., Wu, X. Y., Salazar-Gonzalez, J. F., Salazar, M. G., Kilby, J. M., Saag, M. S., Komarova, N. L., Nowak, M. A., Hahn, B. H., Kwong, P. D. & Shaw, G. M. (2003). Antibody neutralization and escape by HIV-1. *Nature* **422**, 307-312.
22. West, A. P., Jr., Galimidi, R. P., Foglesong, C. P., Gnanapragasam, P. N., Huey-Tubman, K. E., Klein, J. S., Suzuki, M. D., Tiangco, N. E., Vielmetter, J. & Bjorkman, P. J. (2008). Design and expression of a dimeric form of the Human Immunodeficiency Virus Type 1 antibody 2G12 with increased neutralization potency. *J Virol*.
23. Wolbank, S., Kunert, R., Stiegler, G. & Katinger, H. (2003). Characterization of human class-switched polymeric (immunoglobulin M [IgM] and IgA) anti-human immunodeficiency virus type 1 antibodies 2F5 and 2G12. *J Virol* **77**, 4095-103.
24. Yang, F., Bewley, C. A., Louis, J. M., Gustafson, K. R., Boyd, M. R., Gronenborn, A. M., Clore, G. M. & Wlodawer, A. (1999). Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J Mol Biol* **288**, 403-12.
25. Barrientos, L. G., Lasala, F., Delgado, R., Sanchez, A. & Gronenborn, A. M. (2004). Flipping the switch from monomeric to dimeric CV-N has little effect on antiviral activity. *Structure* **12**, 1799-807.
26. Kelley, B. S., Chang, L. C. & Bewley, C. A. (2002). Engineering an obligate domain-swapped dimer of cyanovirin-N with enhanced anti-HIV activity. *J Am Chem Soc* **124**, 3210-1.
27. Stemmer, W. P., Cramer, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49-53.
28. Barrientos, L. G., Louis, J. M., Hung, J., Smith, T. H., O'Keefe, B. R., Gardella, R. S., Mori, T., Boyd, M. R. & Gronenborn, A. M. (2002). Design and initial characterization of a circularly permuted variant of the potent HIV-inactivating protein cyanovirin-N. *Proteins* **46**, 153-60.

29. Colleluori, D. M., Tien, D., Kang, F., Pagliei, T., Kuss, R., McCormick, T., Watson, K., McFadden, K., Chaiken, I., Buckheit, R. W., Jr. & Romano, J. W. (2005). Expression, purification, and characterization of recombinant cyanovirin-N for vaginal anti-HIV microbicide development. *Protein Expr Purif* **39**, 229-36.
30. Mori, T., Barrientos, L. G., Han, Z., Gronenborn, A. M., Turpin, J. A. & Boyd, M. R. (2002). Functional homologs of cyanovirin-N amenable to mass production in prokaryotic and eukaryotic hosts. *Protein Expr Purif* **26**, 42-9.
31. Mori, T., Shoemaker, R. H., Gulakowski, R. J., Krepps, B. L., McMahon, J. B., Gustafson, K. R., Pannell, L. K. & Boyd, M. R. (1997). Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV (human immunodeficiency virus)-inactivating protein. *Biochem Biophys Res Commun* **238**, 218-22.
32. Esser, M. T., Mori, T., Mondor, I., Sattentau, Q. J., Dey, B., Berger, E. A., Boyd, M. R. & Lifson, J. D. (1999). Cyanovirin-N binds to gp120 to interfere with CD4-dependent human immunodeficiency virus type 1 virion binding, fusion, and infectivity but does not affect the CD4 binding site on gp120 or soluble CD4-induced conformational changes in gp120. *J Virol* **73**, 4360-71.
33. Vigerust, D. J. & Shepherd, V. L. (2007). Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol* **15**, 211-8.
34. Fenouillet, E., Gluckman, J. C. & Bahraoui, E. (1990). Role of N-linked glycans of envelope glycoproteins in infectivity of human immunodeficiency virus type 1. *J Virol* **64**, 2841-8.
35. Montefiori, D. C., Robinson, W. E., Jr. & Mitchell, W. M. (1988). Role of protein N-glycosylation in pathogenesis of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* **85**, 9248-52.
36. Seaman, M. (2008). Bjorkman Designed Reagents NAb results. Email communication edit.
37. Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H., Boyd, M. R. & Wlodawer, A. (2002). Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem* **277**, 34336-42.

Table 2-1. CVN₂ and CVN₃ linker sequences.

Variant	Linker Sequence
CVN ₂ L0	N/A
CVN ₂ L1	G
CVN ₂ L3	GSG
CVN ₂ L5	GGSGG
CVN ₂ L6	GSGGSG
CVN ₂ L7	(GGS) ₂ G
CVN ₂ L8	(GGS) ₂ GG
CVN ₂ L9	GGSGGGSGG
CVN ₂ L10	(GGSGG) ₂
CVN ₂ L11	(GGS) ₃ GG
CVN ₂ L13	GGS(GGGS) ₂ GG
CVN ₂ L15	(GGSGG) ₃
CVN ₂ L17	GGS(GGGS) ₃ GG
CVN ₂ L20	(GGSGG) ₄
CVN ₃ L0	N/A
CVN ₃ L5	GGSGG
CVN ₃ L10	(GGSGG) ₂

Table 2-2. IC₅₀S (nM) of CVN and HIV neutralizing antibodies^a against various envelopes in HIV clades A, B, and C.

Clade	Envelope	4 E10	2G12	2F5	IgG1b12	CV-N	CVN ₂ L0	CVN ₂ L10
Clade A^b								
	DJ263.8	N/A	N/A	N/A	N/A	7.48	0.46	0.48
	Q23.17	108.75	>300	46.25	>300	15.4	1.35	2.59
	Q842.d12	87.5	>300	53.75	>300	19.1	2.83	3.34
	Q259.d2.17	89.375	>300	66.25	>300	162.6	13.79	41.1
	3718.v3.c11	71.875	>300	21.25	>300	64.33	4.01	8.89
	0330.v4.c3	36.25	4.375	59.375	>300	2.69	0.11	0.25
	3415.v1.c1	146.875	13.125	227.5	156.25	5.69	0.28	0.44
Clade B^c								
	SF163.LS	1.875	3.75	0.625	0.0625	16.43	1.05	1.69
	PV0.4	40.625	7.5	>300	>300	4.09	0.23	0.49
	CAAN5342.A2	16.875	>300	22.5	>300	34.29	9.49	14.61
	WITO4160.33	1.875	6.875	3.75	19.375	1.85	0.1	0.12
	AC10.2.29	1.875	>300	8.125	11.875	5.01	0.27	0.79
	SC422661.8	5.625	13.125	4.375	1.25	4.21	0.24	0.42
	6535.3	1.25	12.5	11.875	8.75	18.26	1.35	2.56
	THRO4156.18	1.875	>300	>300	3.125	7.75	0.59	0.74
	REJO4541.67	4.375	>300	3.75	4.375	11.48	0.5	0.68
	TRJO4551.58	28.125	>300	>300	>300	4.48	0.13	0.35
	QH0692.42	8.75	17.5	6.25	1.875	14.32	1.02	2.39
	TRO.11	1.875	2.5	>300	>300	9.57	0.53	0.85
	RHPA4259.7	43.125	>300	75	0.625	11.57	1.14	2.04
Clade C^d								
	MW965.26	N/A	N/A	N/A	N/A	7.06	0.4	0.89
	ZM197M.PB7	3.125	>300	76.875	124.375	4.34	0.41	0.49
	ZM249.PL1	13.125	>300	>300	20	22.93	1.93	2.31
	ZM53M.PB12	43.75	>300	>300	161.875	19.24	1.41	3.42
	ZM214M.PL15	25	>300	>300	18.75	29.35	1.46	2.59
	Du156.12	1.25	>300	>300	5	24.32	1.99	3.8
	Du442.1	4.375	>300	>300	1.25	5.02	0.28	0.48
	Du172.17	1.875	>300	>300	6.25	3.31	0.33	0.38
	CAP45.2.00.G3	16.25	>300	>300	4.375	1.21	0.17	0.41
	CAP210.2.00.E8	7.5	>300	>300	127.5	16.75	1.43	1.03
	ZM233M.PB6	7.5	>300	>300	>300	4.56	0.24	0.29
	ZM109F.PB4	3.75	>300	>300	>300	18	2.77	5.49
	ZM135M.PL10a	3.75	>300	>300	>300	16.79	1.89	3.08

^a A molecular weight of 160,000 g/mol was used to convert neutralizing antibody data (4E10, 2G12, 2F5, IgG1b12) from µg/mL to nM.

^b Clade A neutralizing antibody data, personal communication, Prof. Michael Seaman.³⁶

^c Clade B neutralizing antibody data from Li *et al.*, 2005.⁶

^d Clade C neutralizing antibody data from Li *et al.*, 2006.⁷

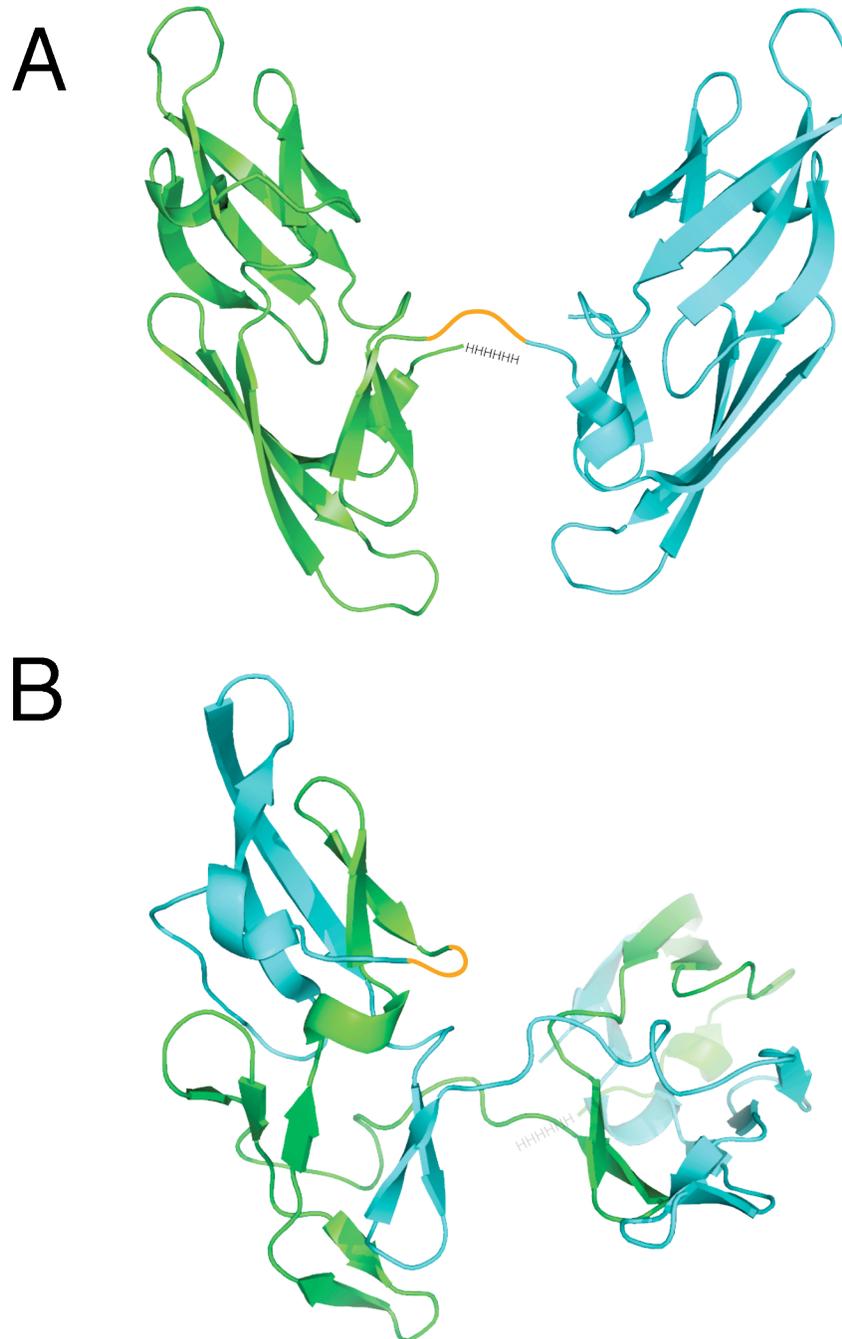


Figure 2-1. Model of generic CVN₂ protein. The CVN repeats are shown in green and cyan and the flexible polypeptide linker is shown in orange. The N-terminal His-tag is depicted as HHHHHH. The CVN₂ structures may adopt a linked monomer structure (A) or a linked domain-swapped structure (B). These representations were created using solved WT CVN structures^{16,24} and are not based on any structural data (see Chapter 3 for structural information).

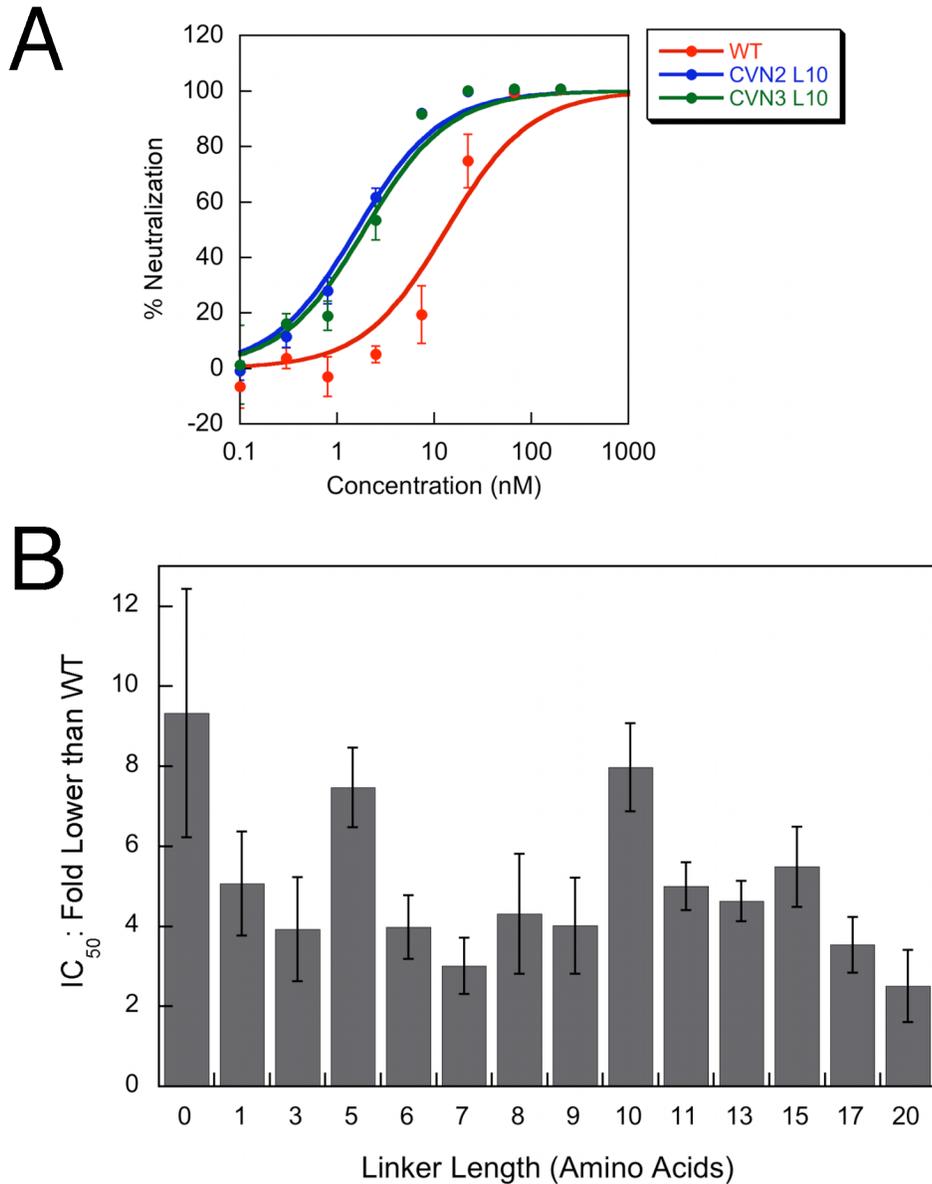


Figure 2-2. HIV neutralization assay results. (A) Typical neutralization data for WT CVN and two variants run on the same plate. The data are analyzed and fit as described in the methods. (B) Summary of IC_{50} s from various CVN₂s of differing linker lengths as compared to WT on the same plate. CVN₂ L0, CVN₂ L5, and CVN₂ L10 show the largest increase in efficacy over WT. All linked dimers, however, are at least two-fold more effective than WT.

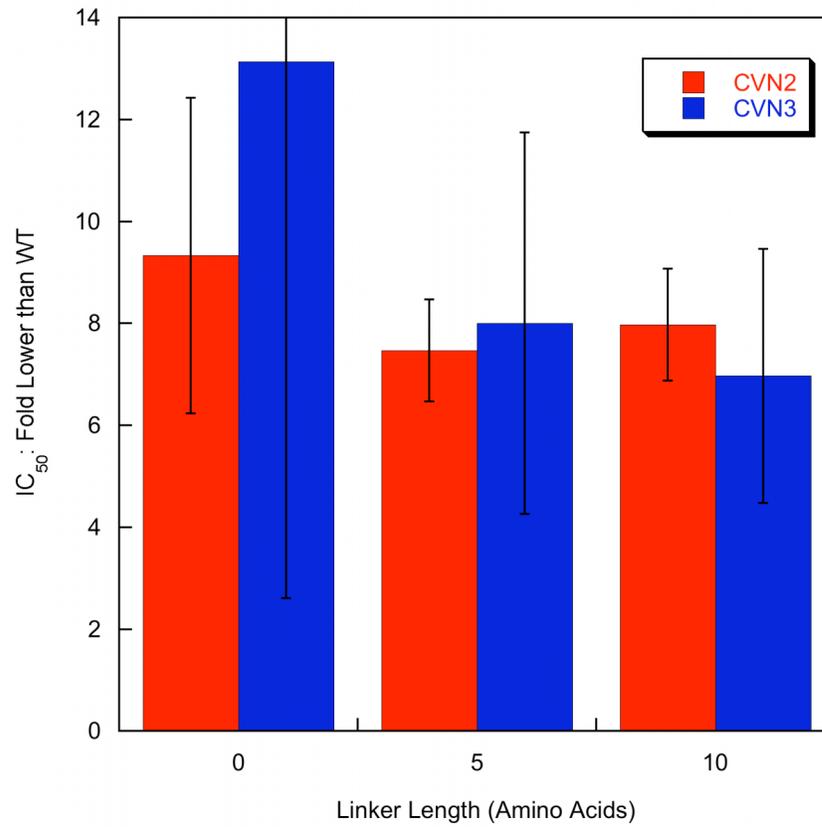


Figure 2-3. CVN₃ HIV neutralization data. The CVN₃ variants are all significantly more effective at neutralizing HIV than WT, but there is no significant difference between CVN₂s (red bars) and CVN₃s (blue bars) of these linker lengths.

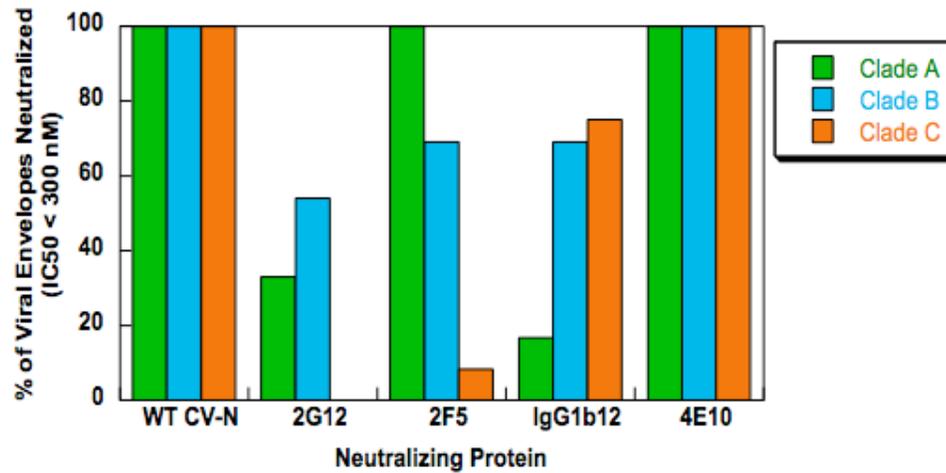


Figure 2-4. WT CVN cross-clade reactivity compared to broadly HIV neutralizing antibodies. WT CVN effectively neutralized all of the HIV pseudoviruses from clades A, B, and C. The 2G12 HIV neutralizing antibody neutralized some clade A and clade B viruses, but is not effective against any clade C viruses. The 2F5 neutralizing antibody works against only a few clade C envelopes and the IgG1B12 antibody is not fully effective against clade A envelopes. 4E10 is the only HIV neutralizing antibody with comparative cross-clade reactivity to WT CVN, neutralizing all viruses tested.

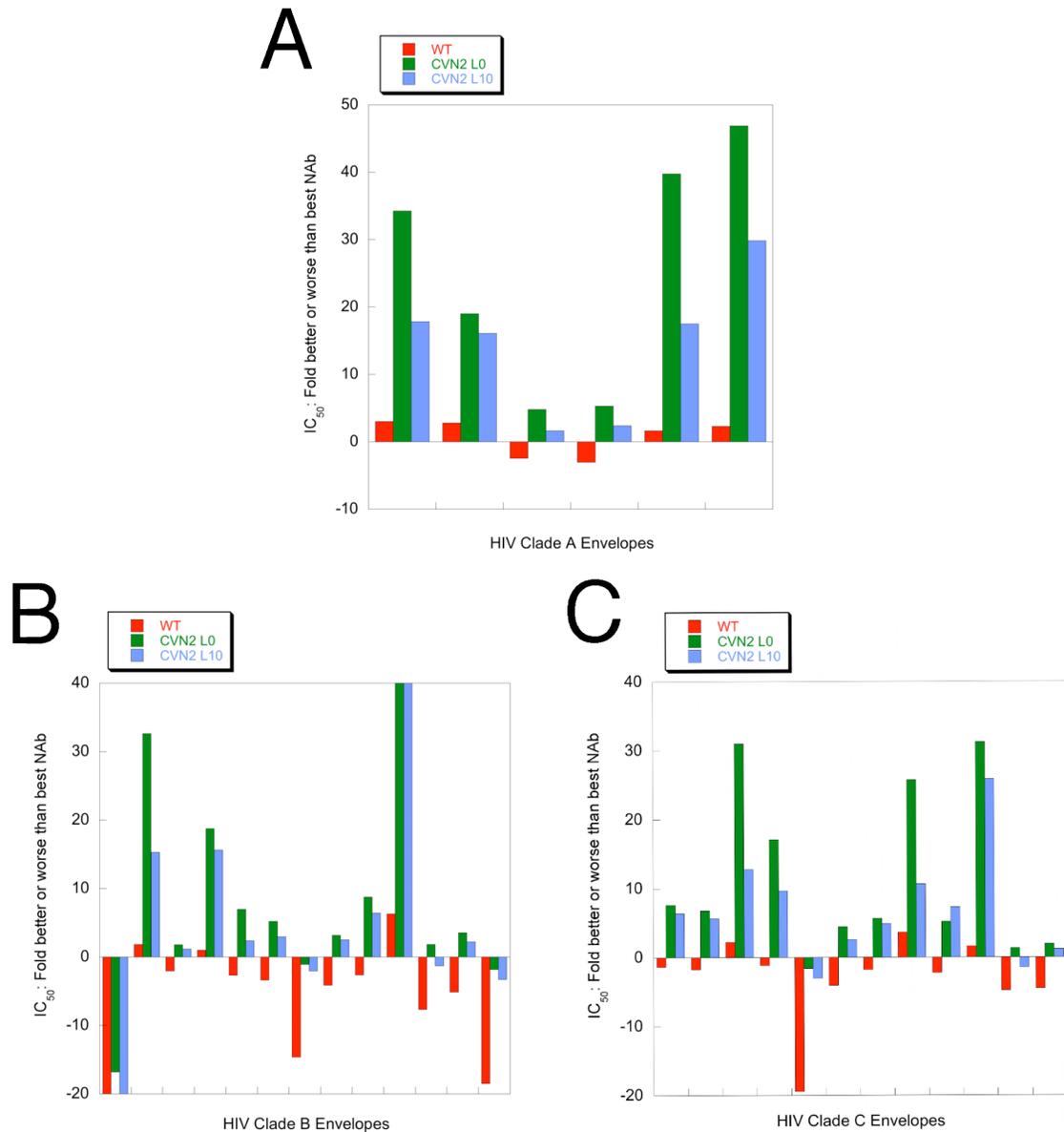


Figure 2-5. Engineered CVN₂ variants neutralize most HIV pseudoviruses with a lower IC₅₀ compared to the most effective broadly neutralizing antibody (NAb). For each envelope, the neutralizing antibody with the lowest IC₅₀ was chosen for comparison (see Table 2-2). CVN, CVN₂ L0, and CVN₂ L10 were evaluated against this best NAb from each envelope. CVN IC₅₀s that were lower than the NAb IC₅₀ were treated as described in the methods (“Fold lower than WT”). For variants with higher IC₅₀s than the NAb, the “Fold worse than best NAb” is the negative inverse of “Fold lower than WT” to provide clarity in the plot. Variants that have IC₅₀s lower than the best NAb are shown with positive bars and those with IC₅₀s higher than the best NAb are shown with negative bars. (A) Clade A envelopes. (B) Clade B envelopes. (C) Clade C envelopes.

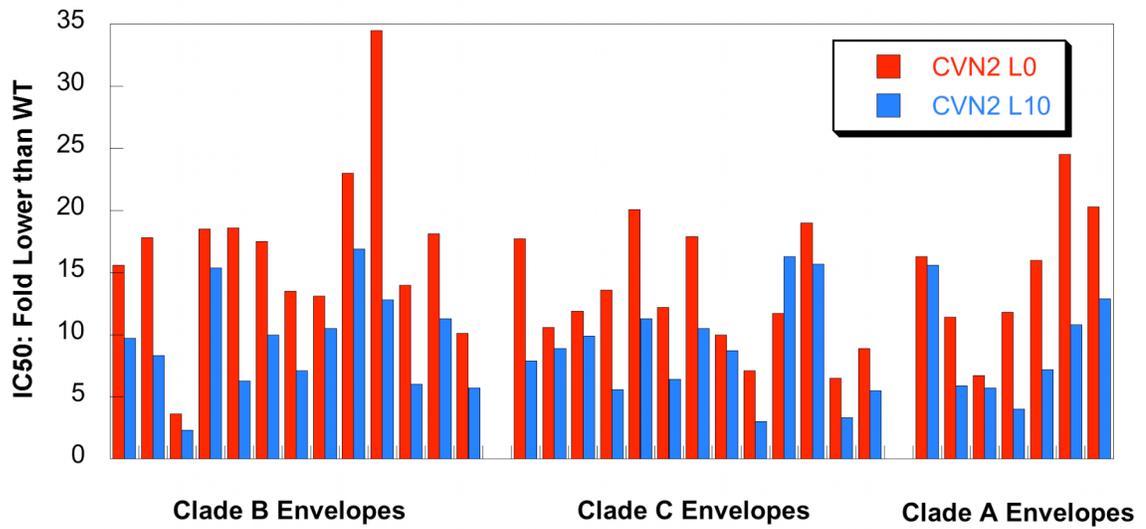


Figure 2-6. Engineered CVN variants are more effective at neutralizing various HIV pseudoviruses than WT CVN. For every virus tested, CVN₂ L10 neutralized with a lower IC₅₀ than WT CVN. For all viruses except one (a clade C envelope), CVN₂ L0 neutralized with a lower IC₅₀ than CVN₂ L10. CVN₂ L0 was up to 35-fold better at neutralizing HIV as compared to WT. CVN₂ L10 was at most 15-fold better neutralizing than WT.

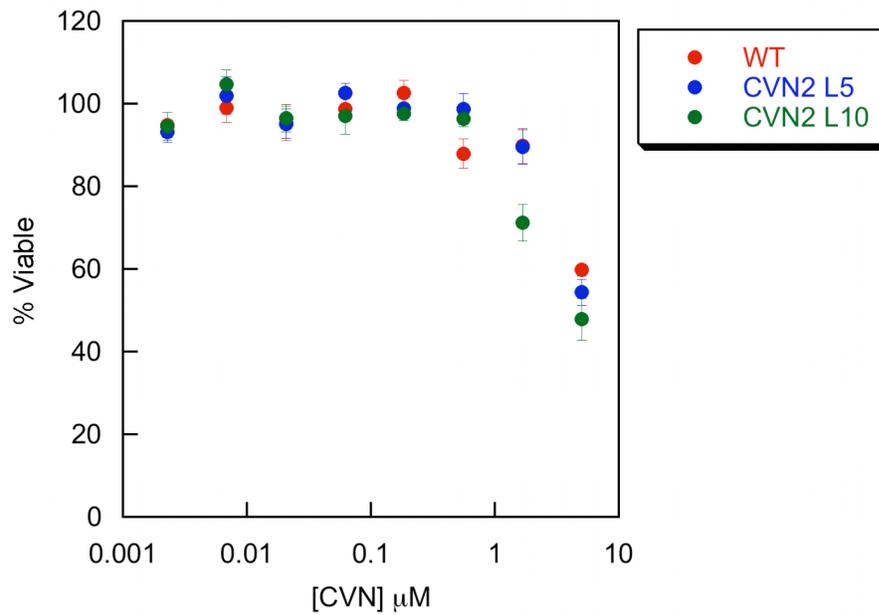


Figure 2-7. Cellular toxicity assay of CVN and CVN₂s. An XTT-based assay was used to determine whether CVN and CVN₂ variants are toxic to Tzm-B1 cells in culture. The CVNs are not toxic at the concentrations used in the HIV neutralization assay (up to 200 nM), but some toxicity was observed at higher concentrations.

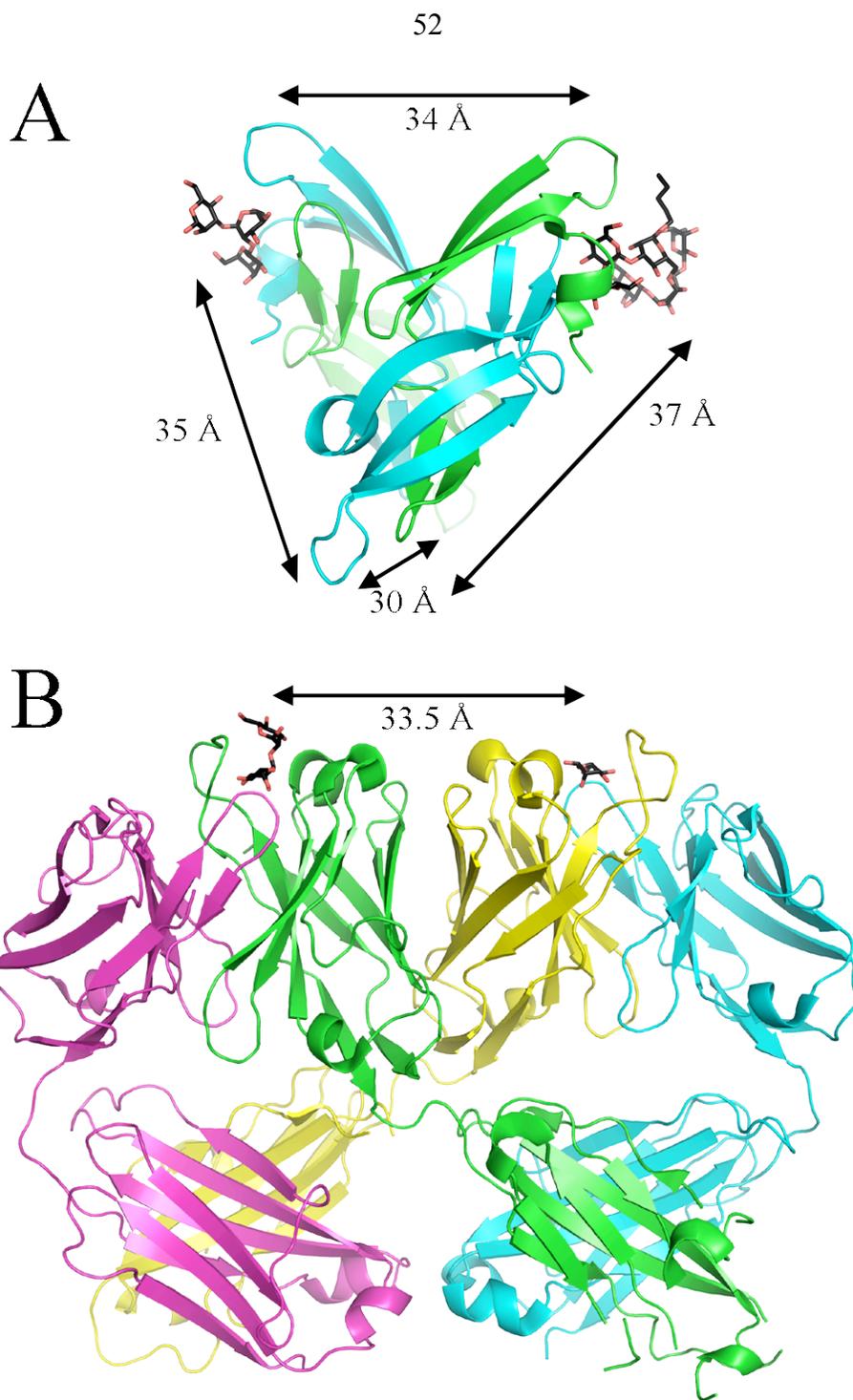


Figure 2-8. Carbohydrate binding site spacing in CVN and the 2G12 anti-HIV Fab. (A) Each of the four carbohydrate binding sites in the WT CVN crystal structure (P4₁2₁2 space group)³⁷ is approximately 30 to 40 Å from the other sites. (B) The 2G12 Fab, which is specific to carbohydrates on gp120 and is broadly neutralizing, has an unusual domain-swapped form in the crystal structure.¹⁴ This domain-swapping rigidifies the carbohydrate binding sites with respect to each other and holds them approximately 35 Å apart.

Chapter 3

Structural characterization of engineered cyanovirin-N dimers

Abstract

Cyanovirin-N (CVN) is an antiviral protein that has broad specificity for enveloped viruses. It specifically binds carbohydrates expressed on envelope proteins and prevents the virus-host cell interactions that are necessary for infection. In the previous chapter, we described dimeric CVN variants (CVN₂s) that contain a tandem repeat of CVN linked through a flexible polypeptide linker. These variants showed up to 10-fold greater efficacy against HIV in a cell-based HIV neutralization assay and also showed broad specificity for different strains and clades of HIV. To investigate the mechanism for this increase in biological function, we have solved four crystal structures of three of these dimeric variants. The structures show that CVN₂ L0, CVN₂ L1, and CVN₂ L10 are all intramolecularly domain-swapped and have a great degree of similarity to wild-type (WT) CVN domain-swapped structures. Although we were able to see some density for the linker in each case, the structures were complicated by the fact that only half of the CVN₂ dimer was in the asymmetric unit in three out of four cases. In the fourth case, a CVN₂ L1 structure, we could clearly distinguish the free termini from the linked termini. Unfortunately, there were no major differences between the linked dimers and WT CVN, and therefore major structural changes do not contribute to the increase in HIV neutralization. We can, however, conclude that linking the termini of two CVN molecules does not adversely affect the structure or function and that the molecules are intramolecularly domain-swapped. Additionally, in the domain-swapped form, the distance between carbohydrate binding sites in CVN become more consistent and these distances are similar to the distance between the binding sites in the broadly neutralizing

anti-HIV antibody, 2G12. By stabilizing this domain-swapped form, we may be positioning the binding sites at the ideal geometric position to best neutralize HIV.

Introduction

Cyanovirin-N (CVN) is a lectin originally isolated from the cyanobacterium *Nostoc ellipsosporum* that has been shown to effectively neutralize various enveloped viruses, including HIV,¹ influenza,² and Ebola.³ CVN specifically binds α 1-2 linked oligomannose glycosylation⁴ sites on the envelope proteins of these viruses and blocks critical interactions between the virus and the host cell, thus preventing infection.⁵

CVN exists in solution both as a monomer and a domain-swapped dimer (Figure 3-1). The monomer consists of two pseudo-domains that display high sequence homology. The first domain contains residues 1-39 and 90-101, and the second domain contains residues 39-89.⁶ CVN includes a three-stranded antiparallel β sheet and a β hairpin in each pseudo-domain. The pseudo-domains are connected through two helical turns. CVN also contains two native disulfide bridges: between residues 8 and 22, and between residues 58 and 73. These two disulfide bridges are located near each end of the molecule and anchor the first strand of the β sheet to the second strand. The dimer contains the same topology, but is domain-swapped at residues 51-53.⁷ In this case, the first domain of one chain (A) forms a “monomer-like” structure with the second domain of the other chain (B’) in an almost symmetric domain swapping (Figure 3-1B). The two quasi-monomers can sample different orientations relative to each other due to the flexibility of the domain-swap region, and the orientation appears to be pH dependent in crystal structures.⁸

CVN contains two carbohydrate-binding sites of differing affinities for α 1-2 mannose: one at each end of the molecule (Figure 3-1A). The high affinity site, located distal from the N- and C-termini, has a dissociation constant in the low nanomolar range,

whereas the lower affinity site has approximately 10-fold weaker affinity.⁹ To date, no crystal structures have been solved of CVN with carbohydrate bound to the high affinity site. This is probably due to crystallographic packing, which obstructs this binding site. However, numerous structures have been solved with carbohydrate bound to the low affinity sites, and these structures are very similar to the NMR structures.^{10,11}

In solution, CVN exists mainly as a monomer, and NMR structures have been solved of monomeric CVN, both free⁶ and bound to carbohydrates.⁴ However, all crystal structures of the wild-type (WT) protein to date have yielded domain-swapped structures, a proposed artifact of the crystallographic process.^{7,10,12,13} The domain-swapped form is metastable in solution and rapidly converts to the more stable monomer at physiological temperatures, but is stable for long periods of time at low temperature.¹² However, in the crystallographic conditions of high protein concentration, extreme pH, and high precipitant concentration, the equilibrium is shifted from the purified monomer to the crystallized dimer. Solution structures of isolated dimer are similar to those solved using crystallography.¹² Various constructs have been engineered to modulate the domain-swapping, including variants that preferentially form monomers¹² and those that form dimers in solution.^{12,14,15} One of these variants, a five-fold mutant including the P51G mutation, which stabilizes the monomeric state, was solved recently as a monomer using crystallography.¹¹

Although a great deal of work has been done to change the preference for monomeric or dimeric protein, there is still a controversy about the effect of dimerization on the antiviral activity of CVN. Because WT domain-swapped dimer converts to monomer during the course of a viral neutralization assay, it is difficult to assay the effect

of dimerization directly. Therefore, various mutants have been generated to try to elucidate the relationship between oligomerization and activity. Kelley *et al.* created an obligate domain-swapped dimer by deleting one of the residues in the domain-swap region and observed a 3.5-fold reduction in the concentration at which half of the viral particles are neutralized (IC_{50}) of HIV fusion. They also showed a similar 3.5-fold reduction in the IC_{50} for purified WT dimer.¹⁴ However, Barrientos *et al.* tested WT monomer, WT dimer, and various engineered mutants and found that regardless of the oligomerization state, all molecules had essentially the same activity against both HIV and Ebola Zaire.¹⁵ Differences in the incubation time and assay conditions could explain the discrepancy, but the question still remains whether dimeric CVN is more effective at neutralizing viruses than monomeric CVN.

In Chapter 2, I described the creation and characterization of obligate dimeric CVN molecules (CVN_2 s). These proteins consist of two copies of WT CVN linked with a flexible linker of varying length. We tested linker lengths ranging from 0 amino acids (L0) to 20 amino acids (L20) and found that some of the variants displayed up to 10-fold better HIV neutralization than WT. Specifically, CVN_2 L0, CVN_2 L5, and CVN_2 L10 were significantly more effective than WT. By adding only one amino acid between the N- and C-termini, however, the CVN_2 L1 variant displayed only approximately five-fold better neutralization than WT. While this difference could easily be explained by experimental error in the biological assay, we wanted to investigate the mechanism for the increase in efficacy for CVN_2 L0, CVN_2 L1, and CVN_2 L10 and determine whether there was a structural explanation for the increase in activity. We therefore solved crystal structures of these three variants. We hypothesize that differences in the structures of the

proteins, including potential domain-swapping differences, could account for the changes in activity.

Methods

Protein expression and purification. CVN₂ L0, CVN₂ L1, and CVN₂ L10 were expressed and purified as described in Chapter 2. After gel filtration, the proteins were concentrated using 5,000 MWCO Amicon Ultra concentrators (Millipore) to 25-30 mg/mL.

Crystallization. Crystallization conditions were set up using a Mosquito automated nanoliter pipettor (TTP Labtech) in the Molecular Observatory at Caltech. Screening was done with 480 conditions in 96-well sitting drop plates using 0.3×0.3 μ L drops. Each protein crystallized under many conditions, and suitable crystals were found for data collection from these initial screens. The best diffracting CVN₂ L0 crystals were grown in 0.1 M sodium HEPES pH 7.5, 0.8 M potassium dihydrogen phosphate, 0.8 M sodium dihydrogen phosphate. CVN₂ L1 data sets were collected on crystals from 0.1 M phosphate-citrate pH 4.2, 2.0 M sodium/potassium phosphate (P3₂21 structure) and from 0.1 M CHES pH 9.5, 0.2 M lithium sulfate and 1 M potassium/sodium tartrate (P4₁2₁2 structure). The CVN₂ L10 data set was collected on a crystal grown in 0.2 M sodium fluoride and 20% PEG-3350.

Data collection and refinement. All crystals except the CHES CVN₂ L1 crystal were cryoprotected in TMP oil. CVN₂ L1 crystals grown in the CHES/tartrate condition

were cryoprotected using the reservoir condition including 20% glycerol. Data for the CVN₂ L0 and L1 structures were collected using a MicroMax-007HF X-ray generator with an RAXIS IV++ detector (Rigaku Corp.). The CVN₂ L10 data set was collected on the 12-2 beam line at the Stanford Synchrotron Radiation Lightsource (SSRL). All data were processed using CrystalClear (Rigaku Corp.) and Mosflm.¹⁶ The indexed and scaled data were further evaluated using CCP4i.¹⁷ The molecular replacement for data sets indexed to the P3₂21 space group were done using 3EZM as the starting model.⁷ The molecular replacement of the CVN₂ L1 data in the P4₁2₁2 space group was done using 2Z21 as the starting model.¹¹ Phaser version 1.3.3 was used for the molecular replacement.¹⁸ Further refinement was done using Coot¹⁹ and Refmac²⁰ and omit maps were created using CNS.^{21,22} Figures were made using PyMOL.²³

Results

Crystallization. CVN₂ L0, CVN₂ L1, and CVN₂ L10 were chosen for structural characterization. As described in Chapter 2, CVN₂ L0 and CVN₂ L10 were the most active of the engineered dimers, while CVN₂ L1 was less active than CVN₂ L0 due to the single amino acid linker. We therefore solved crystal structures of these proteins to determine whether any major structural differences could explain the changes in the HIV neutralization activity that we observed. All of the proteins were crystallized in 96-well plates with 0.6 μ L drops. CVN₂ L0 crystallized in approximately 20 out of the 480 conditions tested. Most of these conditions contained sulfate or phosphate as the precipitant and low pH buffers, although the protein also crystallized well in 20% PEG 3350. CVN₂ L10 crystallized in similar conditions to CVN₂ L0, and crystals were

observed in approximately 35 conditions. CVN₂ L1 crystallized in approximately 35 conditions as well, but in addition to crystal forms seen for CVN₂ L0 or CVN₂ L10, new crystal forms were observed in high pH conditions. Approximately half of the crystal conditions for CVN₂ L0 were above neutral pH, while the CVN₂ L10 conditions were only about one-third above neutral pH. The structure for CVN₂ L1 was therefore determined from both a low pH condition (P3₂21 space group) and a high pH condition (P4₁2₁2 space group). CVN₂ L0 and CVN₂ L10 structures were solved using the only well-diffracting crystals available, which were from low pH conditions (P3₂21 space group).

Crystal structure refinement. All of the P3₂21 space group structures were solved using the domain-swapped WT structure 3EZM.pdb⁷ as the model for molecular replacement (Table 3-1). This WT structure was solved from the same space group, and the molecular replacement provided a good initial model. Further refinement on all structures yielded domain-swapped models that fit the density well. Omit maps were calculated for each of the structures, which agreed well with the models. The omit maps did not indicate any major differences for either the backbone or the side chains of the structure. Solvent molecules, including waters and sodium ions, were added to each structure when there were appropriate electron density and hydrogen bonding partners.

The structures solved from P3₂21 space group crystals contained only half of the CVN₂ dimer in the asymmetric unit, and the second tandem repeat of CVN was generated through crystallographic symmetry. However, because the two copies of CVN are covalently linked through a flexible polypeptide linker, this caused difficulty in the

refinement. In order to properly model the termini as well as the linkage, the linker residues are at 50% occupancy. This is because half of the proteins are oriented in a way that the free termini are in a specific location, while the other half of the proteins have the linkage in that same location. The protein crystallized in both orientations with approximately the same frequency, resulting in symmetry with 50% occupancy of the termini. In the case of CVN₂ L0, as described below, we have not yet been able to fully resolve the linkage. We suspect that the free termini and the linked termini have significantly different conformations that are difficult to distinguish in the density.

In addition to the low pH structure, the CVN₂ L1 structure at high pH (P4₁2₁2 space group) was also determined (Table 3-1). A molecular replacement with 3EZM.pdb was suboptimal, giving a solution and electron density map that did not correlate. However, replacement with a monomeric five-fold variant of WT CVN (2Z21.pdb¹¹) gave a model and map that were reasonable. Upon inspection of the map, it was clear that the structure was in fact domain-swapped, similarly to the P3₂21 structures. The structure model was modified to reflect the domain-swapping of the electron density, and solvent molecules were added.

To confirm that the crystallized protein in each case was in fact CVN₂ and not contamination from WT CVN, we ran an SDS-PAGE gel on crystals grown in the same conditions as those the data sets were collected on. The crystals were rinsed to remove any non-crystallized protein before being denatured. The gel showed that all of the crystals were indeed CVN₂ with no WT contamination (data not shown), allowing us to conclusively determine that indeed only half of the molecule was present in the

asymmetric unit in the P3₂21 space group cases and that it was reasonable to model in the linker at 50% occupancy.

CVN₂ L0 structure. CVN₂ L0 is a domain-swapped dimer under low pH crystallographic conditions (Figure 3-2). Its structure is remarkably similar to WT CVN, with an RMSD of 0.239 Å. Although the refinement is not complete, we can state with certainty that there are no major disruptions of the structure by directly linking two CVNs together without a linker. There are, however, minor differences in the domain-swapped area compared to WT CVN.

Because only one half of the CVN₂ L0 was in the asymmetric unit, as described above, the electron density at the termini is a composite from both the free and the linked termini. A view of the electron density fit to a free N- and C-terminus shows positive electron density between them (Figure 3-3A), indicating the model does not fit well in this area. We hypothesize that the free termini are in a significantly different conformation from the termini that are linked and that the electron density is a combination of these. Future rounds of refinement will be done to model both conformations separately at 50% occupancy in order to fit the experimental data.

We can clearly see from the electron density that CVN₂ L0 is a domain-swapped dimer under these conditions (Figure 3-3B). Molecular replacement with a monomeric WT CVN resulted in density that also showed clear domain-swapping, indicating that model bias is not responsible for this density.

CVN₂ L1 structures. CVN₂ L1 crystallized with two major morphologies. Because of the difference in the shape of the crystals and because they were indexed to different space groups, we solved two structures of CVN₂ L1: one from the P3₂21 space group and the other from the P4₁2₁2 space group. Upon molecular replacement, it became clear that both structures were WT-like domain-swapped structures with slightly different orientations of the domains relative to each other (Figure 3-4). It has been shown previously that WT CVN crystallizes in different space groups and different morphologies depending on the pH of the crystallization condition.⁸ This appears to be the case here as well.

The P3₂21 structure, solved from a low pH condition, is very similar to the WT low pH structure, with an RMSD of 0.283 Å. Like the other P3₂21 structures, there are minor deviations in the domain-swap area, but overall there are no major structural perturbations by linking two termini. The P4₁2₁2 also does not appear to be significantly different from WT CVN when compared to a structure solved at high pH (Figure 3-4B).¹² The RMSD between WT CVN and the P4₁2₁2 is 0.407 Å. In this case as well, there are no major structural changes to CVN₂ L1 that would explain the vast increase in biological activity.

Clear density was visible for a single glycine linker between the termini of the P3₂21 structure (Figure 3-5A). However, due to the fact that only half of the molecule is in the asymmetric unit, we have modeled the linker in at 50% occupancy. We believe that the free termini have significantly different conformations from the termini that are linked. We have not yet modeled the free termini and trust that this will increase the

reliability of the model. Forty-five water molecules and two sodium ions were placed in the structure with high confidence.

In the P4₁2₁2 structure, 104 water molecules were added. There was also a CHES molecule near the high affinity carbohydrate binding site of one CVN domain, which broke the symmetry of the molecule, causing the entire CVN₂ L1 to be in the asymmetric unit, unlike the P3₂21 structures. The free termini and the linked termini were clearly distinguishable in the initial electron density map, and therefore, we were able to model them separately (Figure 3-6). In addition to different conformations for the linked residues, there was also clear density for the single glycine in the linker.

In both crystal forms, CVN₂ L1 forms a domain-swapped dimer similar to WT CVN (Figure 3-5B, Figure 3-6C). The conformations of residues involved in the swap in the P4₁2₁2 structure are almost identical to the WT structure, but there are small differences in the conformations from the P3₂21 structure compared to WT. While these differences may be real, we do not expect that they explain the significant increase in activity of CVN₂ L1 over WT CVN in the neutralization assay, as described in Chapter 2.

Both CVN₂ L1 structures have unexpectedly high R_{free} values (Table 3-1). Omit maps on each structure indicate that there were no major problems with either the backbone or the side chains. Because the termini and the linkage are somewhat unstructured and the model is not perfectly matched to the electron density in this area, we may be getting some model bias. However, we do not expect these small deviations to have a significant impact on the overall structures. The addition of more water and solvent molecules may also decrease the R_{free} and give more reasonable statistics.

CVN₂ L10 structure. The crystal structure of CVN₂ L10 indicates that this protein is also very similar to WT CVN (Figure 3-7). The two structures have an RMSD of 0.353 Å and the only significant differences between the two structures are in the domain-swap area and some of the backbone phi and psi angles of the β strands. The CVN₂ L10 structure forms more optimal β strands in many cases than previously solved WT structures.

Like all of the other P3₂21 structures, only half of the CVN₂ L10 molecule is in the asymmetric unit. However, because the 10 amino acid linker is long and flexible, the N- and C-termini appear to be identical when free and when linked. We were therefore able to fit four of the linker residues (two glycines on each terminus) at 50% occupancy while leaving Leu1 and Glu101 as 100% occupied (Figure 3-8A). There was no clear density for the six residues in the center of the linker, so they are not included in the structure.

CVN₂ L10 is also definitively domain-swapped and contains several structural water molecules to stabilize this conformation (Figure 3-8B). The CVN₂ L10 data set was molecularly replaced with a monomeric CVN model, and domain-swapping density was clear, indicating in this case, as in the previous cases, the domain swapping is not an artifact of the replacement and refinement process.

Conclusions

We solved four crystal structures of three different CVN₂ variants with linkers containing zero, one, and ten amino acids. We had hoped that significant structural changes or differences in the domain-swapping of the variants would partially explain the

increase in biological activity described in Chapter 2. However, the structures are all remarkably similar to domain-swapped WT CVN crystal structures. All four structures are intramolecularly domain-swapped and show varying amounts of density for the flexible polypeptide linker. The RMSDs for the structures as compared to WT were all less than 0.5 Å and the minor differences were typically observed in the domain-swap region.

Complicating the structure refinement was the fact that three of the four structures contained only half of the CVN₂ molecule in the asymmetric unit. In these cases, the molecule could align in two possible orientations: one with the free termini in a given location and another with the linked termini in the same location. This led to an additional plane of symmetry where the two halves of the CVN₂ were generated by crystallographic symmetry and the free and linked termini were each represented by the same density at 50% occupancy each.

The crystal structures of three CVN₂ molecules showed no major differences from WT CVN. Because the structures are remarkably similar, we can be confident that linking two repeats of CVN does not negatively affect the structure and does not cause any major perturbations. In addition, the linkage stabilizes the domain-swapped form over the monomeric form due to the steric restraints provided by a short linker and the increase in local concentration and therefore the proteins form obligate domain-swapped dimers. All of the structures are intramolecularly domain-swapped, and we feel confident that this is the biologically relevant conformation in solution. If the crystals had contained intermolecularly domain-swapped CVN₂ protein, we could have inferred that the molecule in solution was monomeric-like. We know by gel filtration that the protein is

monomeric or intramolecularly domain-swapped because we can effectively separate intermolecularly domain-swapped CVN₂. Although we are confident that the proteins are indeed domain-swapped in solution, additional studies are necessary to confirm this finding. We hope that NMR experiments will help determine the solution state of the molecules. In addition, we are looking into cleavage assays in which the protein is cleaved at the linker and then run on a gel filtration column. If the protein is indeed forming a domain-swapped dimer in solution, as expected, we would see dimer elute from a gel filtration column. If, on the other hand, the protein contains two independent monomer-like domains, we would expect that after cleavage, each would elute as a monomer on a gel filtration column.

By stabilizing the domain-swapped dimer, we believe we are rigidifying the distances between the four carbohydrate binding sites in CVN₂. Although the two domains may sample various conformations with respect to each other, as demonstrated by the differences between the P3₂21 and P4₁2₁2 structures, the carbohydrate binding sites in both cases are brought together and held in close proximity in the domain-swapped form. For example, in the domain-swapped dimer crystal structure of WT CVN in the P4₁2₁2 space group, each of the four carbohydrate binding sites is approximately 30 to 40 Å from the other sites (Figure 3-9). This geometry may be ideal for interacting with gp120 glycosylation, and by stabilizing this form we may be increasing the affinity of interaction. Additional support for this mechanism comes from the crystal structure of 2G12, a broadly neutralizing anti-HIV antibody that is also specific to the glycosylation on gp120.²⁴ Unlike the standard “Y” shaped antibodies, 2G12 contains a domain swap in the Fab region, which brings the two carbohydrate binding sites approximately 35 Å

apart, a similar distance to the distances in domain-swapped CVN (Figure 3-9B).²⁵ Instead of being highly flexible, the antigen binding region of 2G12 is fixed to enhance the interaction with gp120. The similarity in the spacing between binding sites in domain-swapped CVN and 2G12 indicates that 30 to 40 Å spacing may be biologically relevant for gp120 glycosylation binding and that efficacy may be increased by stabilizing the domain-swapped form of CVN.

Acknowledgements

I would like to thank Len Thomas for help with data collection and refinement and Pavle Nikolovski and the Molecular Observatory for setting up crystal trays. I would also like to thank Professors Pamela Bjorkman, Doug Rees, and Bil Clemons for helpful discussions regarding structure refinement. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

References

1. Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T., Gulakowski, R. J., Wu, L., Rivera, M. I., Laurencot, C. M., Currens, M. J., Cardellina, J. H., 2nd, Buckheit, R. W., Jr., Nara, P. L., Pannell, L. K., Sowder, R. C., 2nd & Henderson, L. E. (1997). Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother* **41**, 1521-30.
2. O'Keefe, B. R., Smee, D. F., Turpin, J. A., Saucedo, C. J., Gustafson, K. R., Mori, T., Blakeslee, D., Buckheit, R. & Boyd, M. R. (2003). Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. *Antimicrob Agents Chemother* **47**, 2518-25.
3. Barrientos, L. G., O'Keefe, B. R., Bray, M., Sanchez, A., Gronenborn, A. M. & Boyd, M. R. (2003). Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral Res* **58**, 47-56.
4. Bewley, C. A. (2001). Solution structure of a cyanovirin-N:Man alpha 1-2Man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* **9**, 931-40.
5. Mori, T. & Boyd, M. R. (2001). Cyanovirin-N, a potent human immunodeficiency virus-inactivating protein, blocks both CD4-dependent and CD4-independent binding of soluble gp120 (sgp120) to target cells, inhibits sCD4-induced binding of sgp120 to cell-associated CXCR4, and dissociates bound sgp120 from target cells. *Antimicrob Agents Chemother* **45**, 664-72.
6. Bewley, C. A., Gustafson, K. R., Boyd, M. R., Covell, D. G., Bax, A., Clore, G. M. & Gronenborn, A. M. (1998). Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nat Struct Biol* **5**, 571-8.
7. Yang, F., Bewley, C. A., Louis, J. M., Gustafson, K. R., Boyd, M. R., Gronenborn, A. M., Clore, G. M. & Wlodawer, A. (1999). Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *J Mol Biol* **288**, 403-12.
8. Botos, I. & Wlodawer, A. (2003). Cyanovirin-N: a sugar-binding antiviral protein with a new twist. *Cell Mol Life Sci* **60**, 277-87.
9. Bewley, C. A. & Otero-Quintero, S. (2001). The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120. *J Am Chem Soc* **123**, 3892-902.
10. Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H., Boyd, M. R. & Wlodawer, A. (2002). Structures of the complexes of a

potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem* **277**, 34336-42.

11. Fromme, R., Katiliene, Z., Giomarelli, B., Bogani, F., Mc Mahon, J., Mori, T., Fromme, P. & Ghirlanda, G. (2007). A monovalent mutant of cyanovirin-N provides insight into the role of multiple interactions with gp120 for antiviral activity. *Biochemistry* **46**, 9199-207.
12. Barrientos, L. G., Louis, J. M., Botos, I., Mori, T., Han, Z., O'Keefe, B. R., Boyd, M. R., Wlodawer, A. & Gronenborn, A. M. (2002). The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure* **10**, 673-86.
13. Botos, I., Mori, T., Cartner, L. K., Boyd, M. R. & Wlodawer, A. (2002). Domain-swapped structure of a mutant of cyanovirin-N. *Biochem Biophys Res Commun* **294**, 184-90.
14. Kelley, B. S., Chang, L. C. & Bewley, C. A. (2002). Engineering an obligate domain-swapped dimer of cyanovirin-N with enhanced anti-HIV activity. *J Am Chem Soc* **124**, 3210-1.
15. Barrientos, L. G., Lasala, F., Delgado, R., Sanchez, A. & Gronenborn, A. M. (2004). Flipping the switch from monomeric to dimeric CV-N has little effect on antiviral activity. *Structure* **12**, 1799-807.
16. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Prot. Crystallography* **26**.
17. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3.
18. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61**, 458-64.
19. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32.
20. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55.
21. Brunger, A. T. (2007). Version 1.2 of the Crystallography and NMR system. *Nat Protoc* **2**, 2728-33.
22. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R.

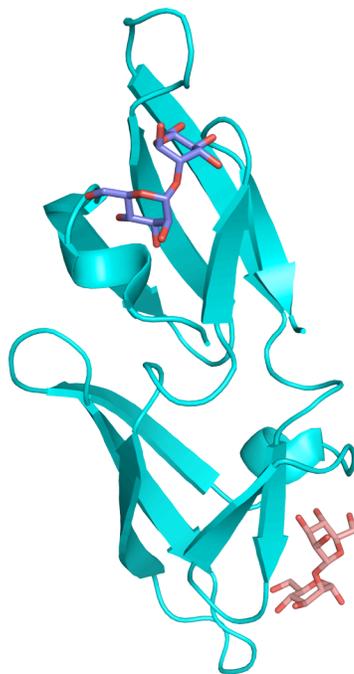
- J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-21.
23. DeLano, W. L. (2008). The PyMOL Molecular Graphics System. <http://www.pymol.org>, DeLano Scientific, Palo Alto, CA.
24. Trkola, A., Purtscher, M., Muster, T., Ballaun, C., Buchacher, A., Sullivan, N., Srinivasan, K., Sodroski, J., Moore, J. P. & Katinger, H. (1996). Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol* **70**, 1100-8.
25. Calarese, D. A., Scanlan, C. N., Zwick, M. B., Deechongkit, S., Mimura, Y., Kunert, R., Zhu, P., Wormald, M. R., Stanfield, R. L., Roux, K. H., Kelly, J. W., Rudd, P. M., Dwek, R. A., Katinger, H., Burton, D. R. & Wilson, I. A. (2003). Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300**, 2065-71.

Table 3-1. Crystallographic statistics.

	CVN ₂ L0	CVN ₂ L1	CVN ₂ L1	CVN ₂ L10
Data collection				
Space group	P3 ₂ 21	P3 ₂ 21	P4 ₁ 2 ₁ 2	P3 ₂ 21
Cell dimensions				
a,b,c (Å)	47.9, 47.9, 78.7	47.5, 47.5, 78.6	60.6, 60.6, 147.6	48.0, 48.0, 79.3
α,β,γ (deg)	90, 90, 120	90, 90, 120	90, 90, 90	90, 90, 120
Resolution (Å) *	2.0 (2.11 - 2.0)	2.0 (2.11 - 2.0)	2.1 (2.21 - 2.1)	1.75 (1.84 - 1.75)
No. reflections	42309	50188	206257	59304
Unique reflections	7456	7347	16878	11158
R _{merge} (%) *	5.1 (22.8)	3.7 (10.8)	4.3 (27.6)	10.3 (38.4)
I/ σ I *	23.6 (5.6)	33.3 (13.5)	42.9 (9.2)	12.7 (4.2)
Completeness (%) *	100.0 (100.1)	99.9 (99.9)	100.0 (100.0)	100.0 (100.1)
Redundancy *	5.7 (5.7)	6.8 (6.7)	12.2 (12.3)	5.3 (5.2)
Refinement				
Resolution (Å)	23.9 - 2.0	28.4 - 2.0	27.1 - 2.1	36.8 - 1.75
No. reflections				
Working set	6687	6642	15109	10060
Test set	341	340	851	528
R _{work} /R _{free}	22.0 / 24.9%	21.8 / 27.4%	22.0 / 28.8%	18.8 / 21.2%
No. atoms				
Protein	796	785	1567	823
Solvent	60	47	118	89
B-factors				
Protein	31.2	31.0	32.4	19.3
Water	34.0	34.5	34.9	28.6
R.m.s. deviations				
Bond lengths (Å)	0.012	0.011	0.011	0.011
Bond angles	1.375	1.307	1.215	1.351
Ramachandran plot				
Favored (%)	89	90.2	88.3	98.2
Allowed (%)	11	9.8	12	10.8
Generously allowed (%)	0	0	0	0
Disallowed (%)	0	0	0	0

* Last shell is shown in parentheses.

A



B

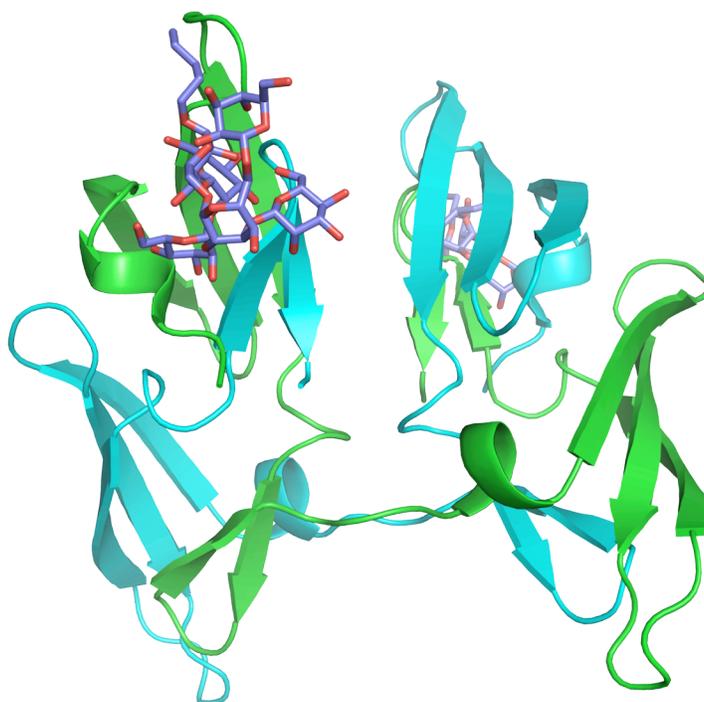


Figure 3-1. Wild-type CVN structures. In solution, CVN can exist as either a monomer (A)⁴ or a domain-swapped dimer (B).¹⁰ CVN is shown in green and cyan ribbons to indicate protein chains. Carbohydrates bound to the high affinity site are shown with orange carbons (present only in A) and carbohydrates bound to the low affinity site are shown with blue carbons (present in both A and B). The monomer and the left half of the dimer are in approximately the same orientation.

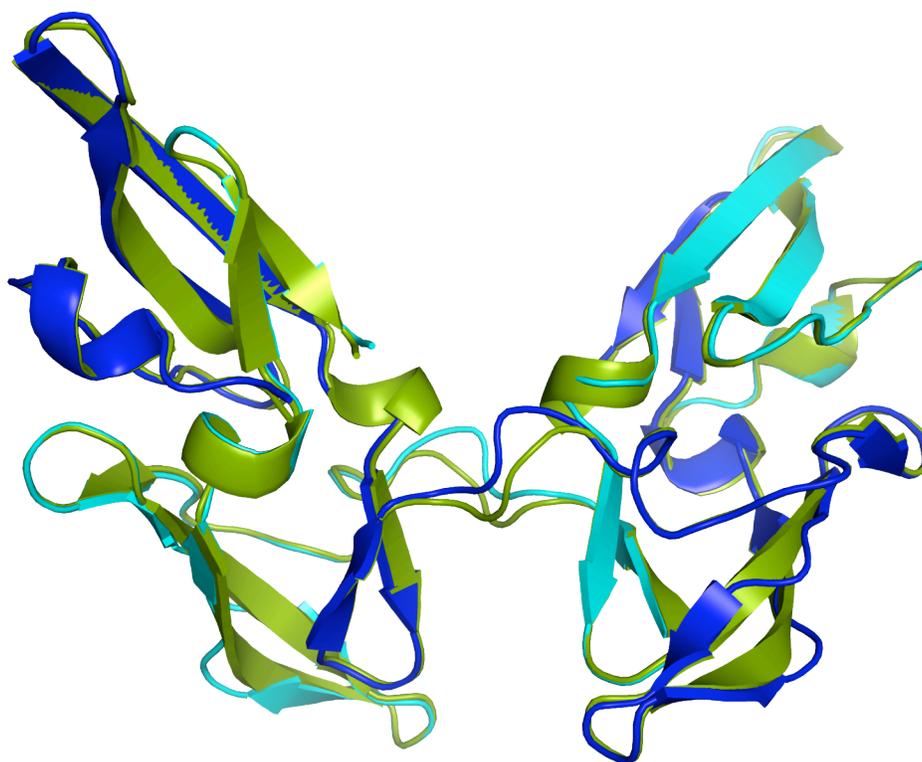


Figure 3-2. CVN₂ L0 crystal structure compared to WT CVN. The CVN₂ L0 crystal structure (green) and domain-swapped WT CVN (3EMZ.pdb⁷) (blue and cyan) have an RMSD of 0.239 Å. The structures are very similar with small deviations in the β strands and the domain swap area.

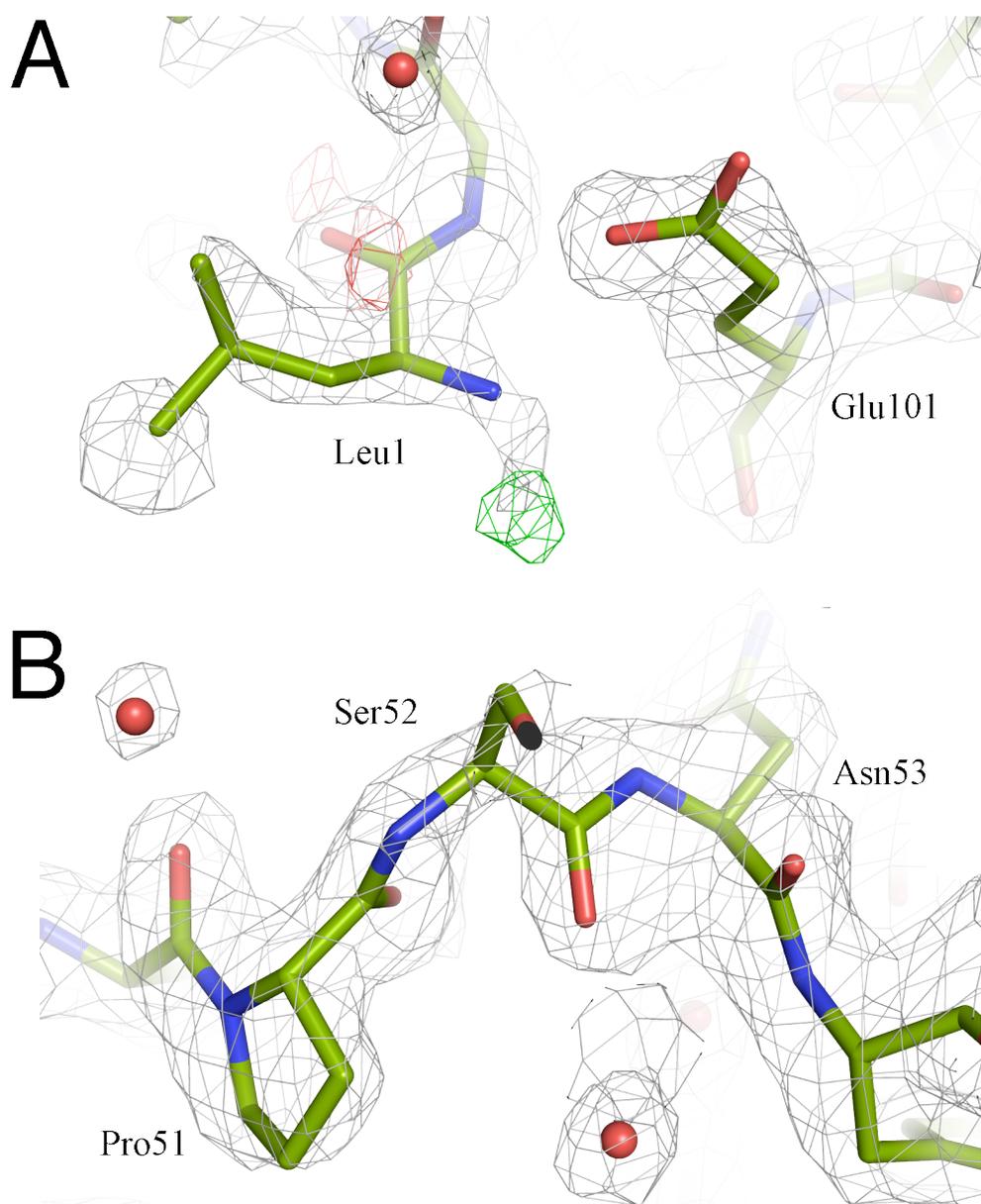


Figure 3-3. CVN₂ L0 structure. Only half of the CVN₂ dimer was found in the asymmetric unit. The other half is generated through crystallographic symmetry, and the linker and free termini are each 50% occupied. The 2Fo-Fc electron density map is shown as gray mesh. (A) The termini of CVN₂ L0 are not well defined. There is positive density where the linkage may occur. Further refinement is necessary to fit the model to the data. Additionally, the free N- and C-terminal residues probably have very different conformations from those that are linked. Positive density from the Fo-Fc map is shown in green, and negative density is shown in red. (B) CVN₂ L0 is domain-swapped in the crystal structure. The density in the swapped area is clear and definitive.

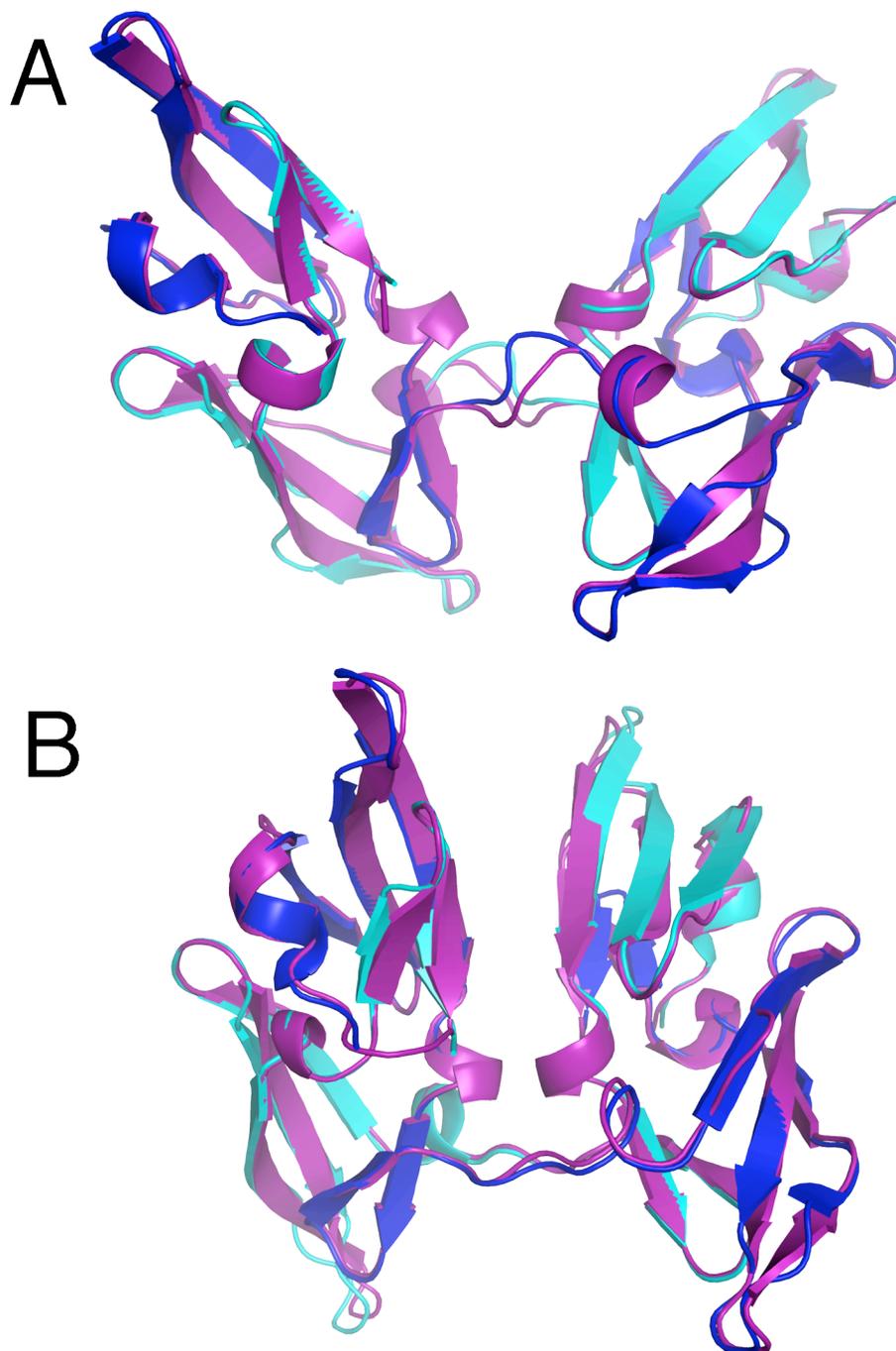


Figure 3-4. CVN₂ L1 crystal structures compared to WT CVN. CVN₂ L1 structures are shown in purple and the WT CVN structures are shown in blue and cyan. (A) CVN₂ L1 in the P3₂21 space group, solved in low pH conditions, overlaid with domain-swapped WT CVN (3EZM.pdb).⁷ The RMSD of these structures is 0.283 Å, and the major differences are seen in the domain swap area. (B) CVN₂ L1 in the P4₁2₁2 space group, solved in high pH conditions, overlaid with domain-swapped WT CVN (1L5B.pdb).¹² These structures have a 0.407 Å RMSD. The one residue linker can be seen in the P4 structure in the domain on the left.

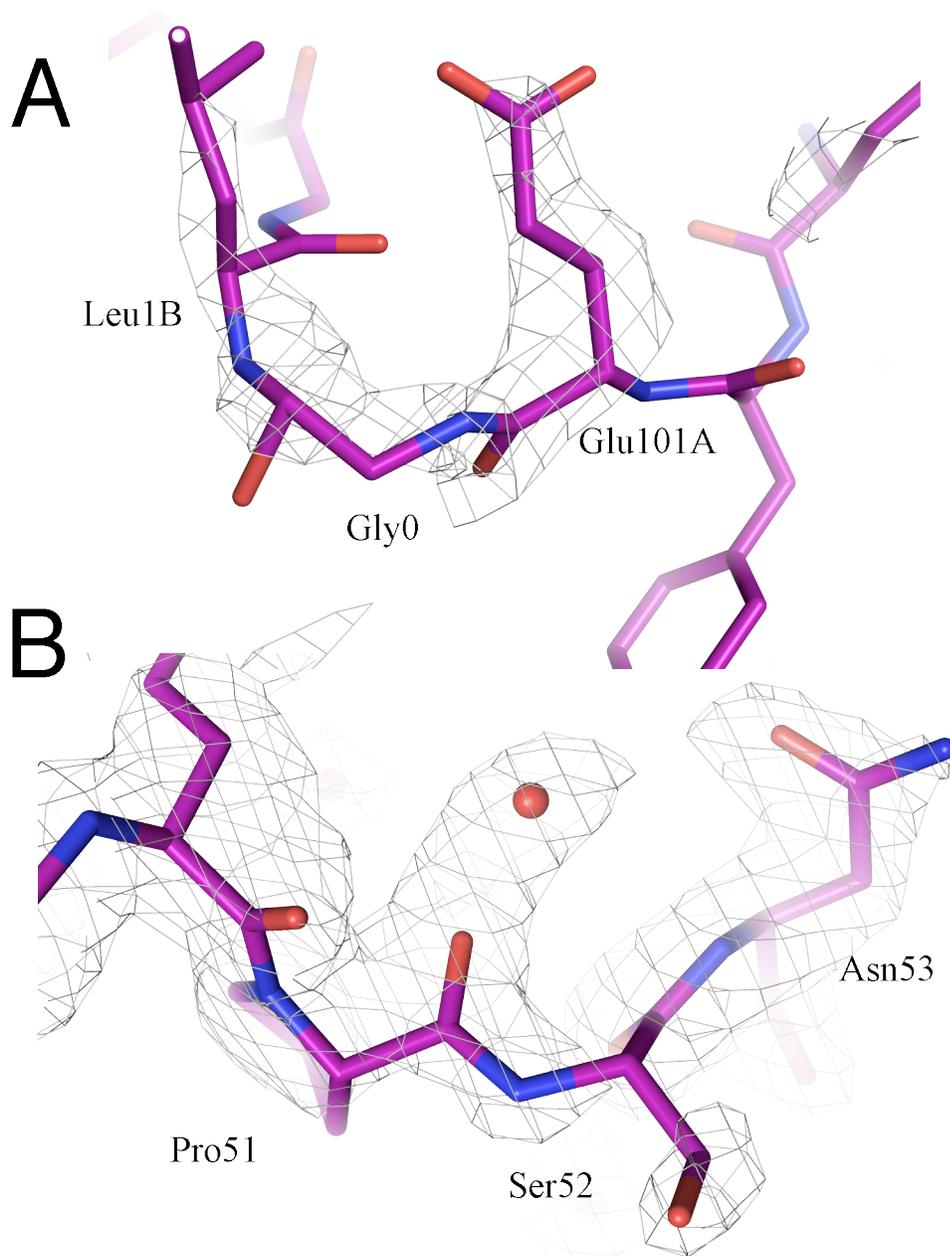


Figure 3-5. CVN₂ L1 P3₂21 structure. Only half of the CVN₂ molecule was in the asymmetric unit. The free termini and the linked termini are both represented with the same density, at 50% occupancy each. (A) The N- and C-termini of CVN₂ L1 with a 2Fo-Fc electron density map contoured at 1 σ showing clear density for the single glycine residue linker. Gly0 shown in the figure is only 50% occupied. (B) The CVN₂ L1 in this crystal structure is clearly domain-swapped as evidenced by clear electron density in the domain swap region.

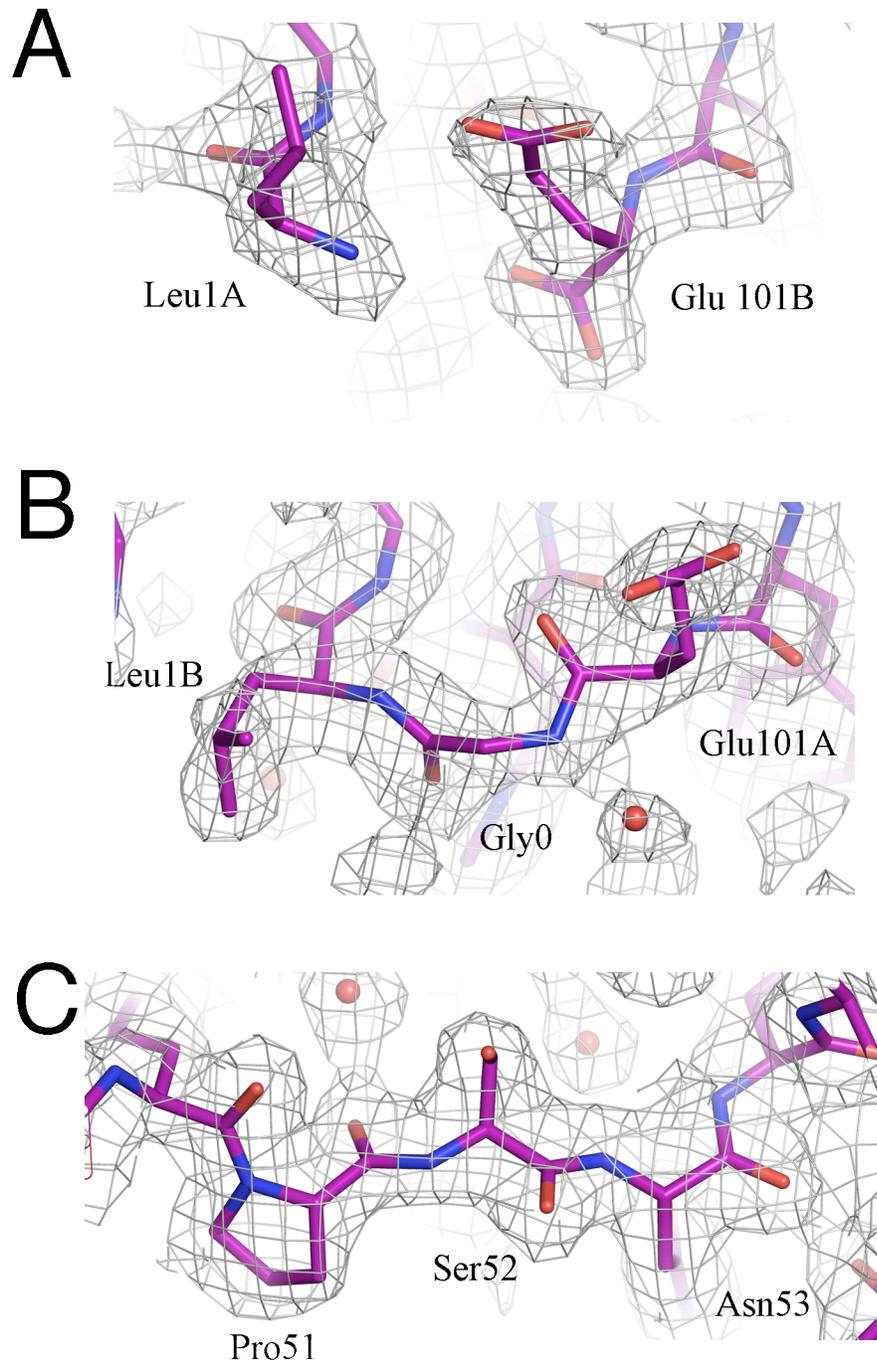


Figure 3-6. CVN₂ L1 P4₁2₁2 structure. The entire CVN₂ L1 dimer was within the asymmetric unit in the P4₁2₁2 crystal structure. The free termini were clearly distinguishable from the linkage as determined by the 2Fo-Fc electron density contoured to 1.0 σ . (A) The free N- and C-termini of CVN₂ L1. (B) The linked N- and C-termini of CVN₂ L1 and clear density for the single glycine linker. (C) Well defined density is seen in both chains for the domain swap residues 51 through 53.

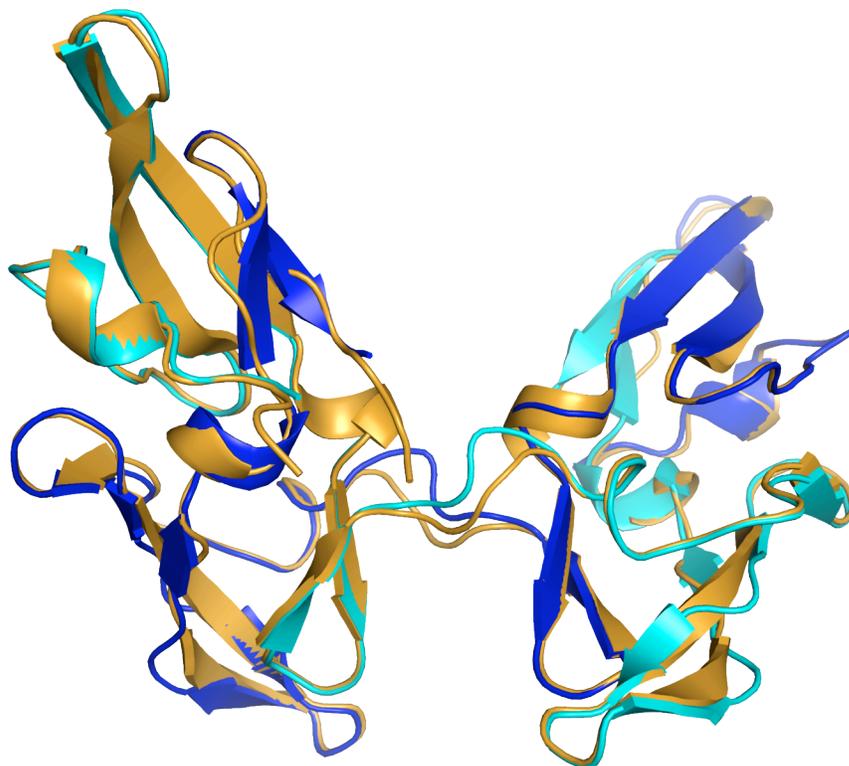


Figure 3-7. CVN₂ L10 structure compared to WT CVN. CVN₂ L10 is shown in orange and the domain-swapped WT CVN (3EZM.pdb)⁷ is shown in blue and cyan. These structures have an RMSD of 0.353 Å. Four of the ten linker residues have electron density and are shown in the left domain of the structure. The other six residues are presumed to be disordered and are therefore not modeled in this structure.

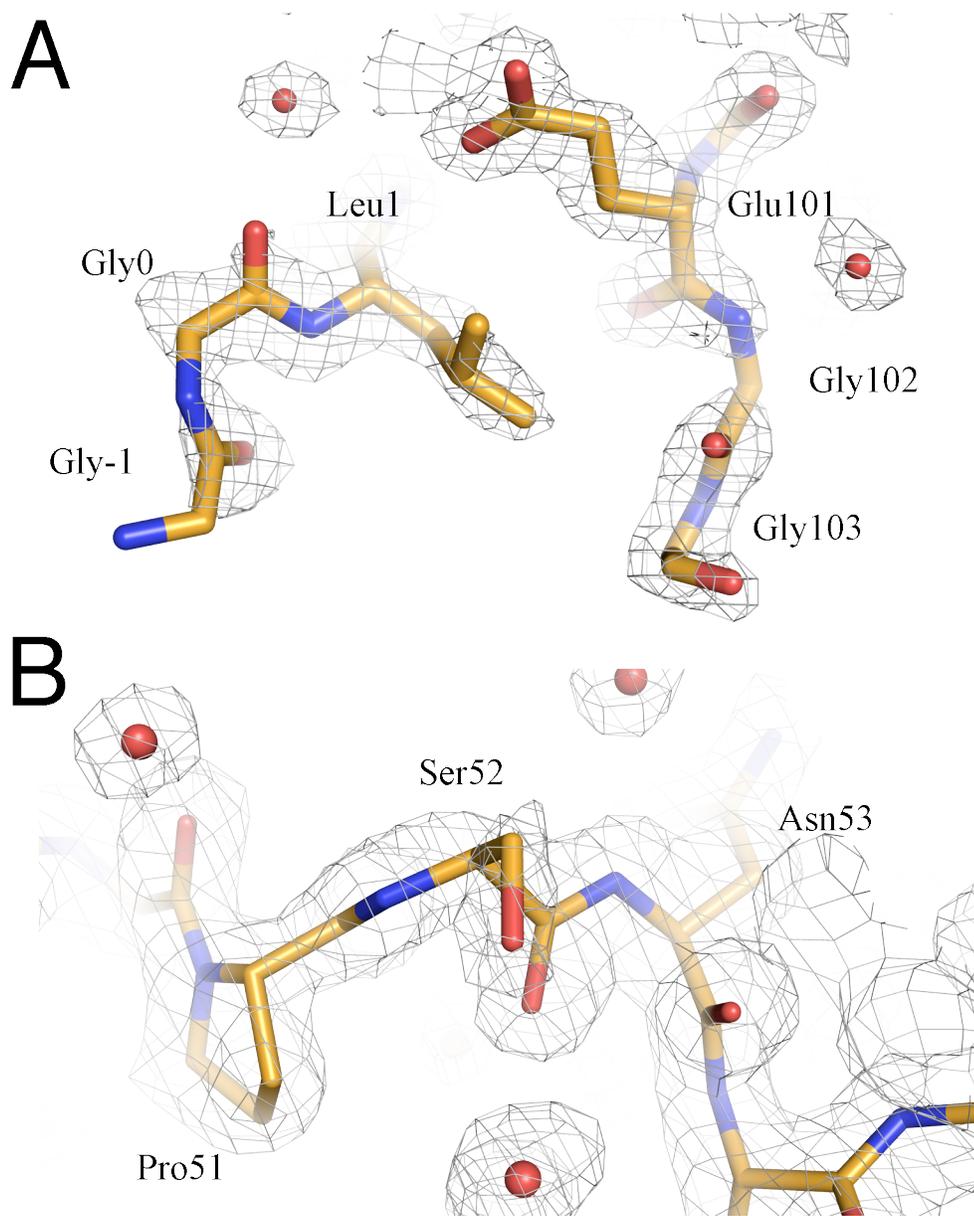


Figure 3-8. CVN₂ L10 structure. Only one copy of CVN was in the asymmetric unit of this structure, therefore the linker residues are only 50% occupied. (A) The termini of the crystal structure with the 2Fo-Fc map contoured to 1.0 σ . Leu1 and Glu101 are both occupied at 100%, whereas the four linker residues with visible density are at 50% occupancy. (B) This structure is domain-swapped as evidenced by the clear electron density in the domain-swapping region (residues 51-53).

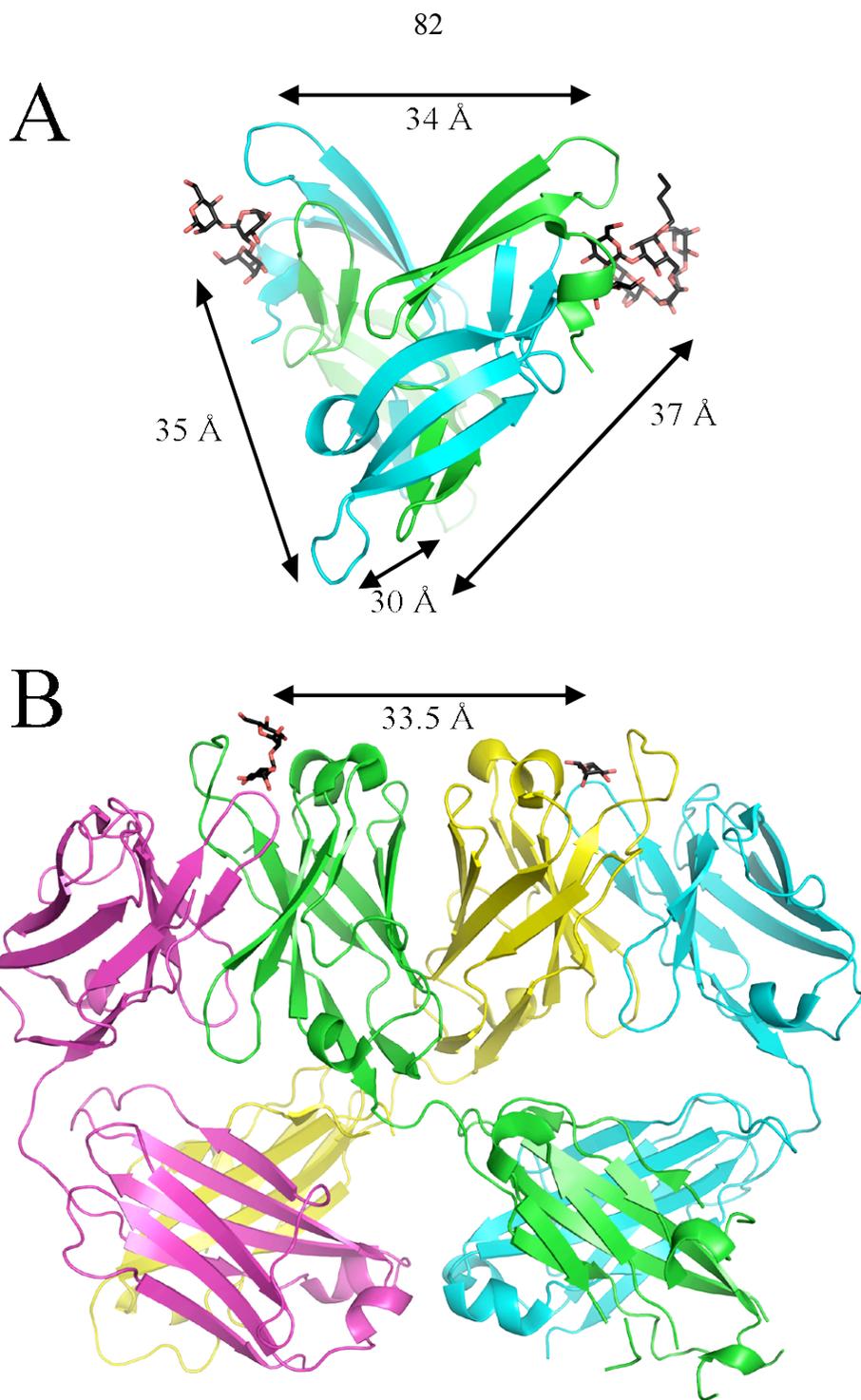


Figure 3-9. Carbohydrate binding site spacing in CVN and the 2G12 anti-HIV Fab. (A) Each of the four carbohydrate binding sites in the WT CVN crystal structure (P4₁2₁2 space group)¹⁰ is approximately 30 to 40 Å from the other sites. (B) The 2G12 Fab, which is specific to carbohydrates on gp120 and is broadly neutralizing, has an unusual domain-swapped form in the crystal structure.²⁵ This domain-swapping rigidifies the carbohydrate binding sites with respect to each other and holds them approximately 35 Å apart.

Chapter 4

Lectibody: Design and characterization of a cyanovirin-N – Fc chimera

Abstract

Cyanovirin-N (CVN) is a protein that is a broadly potent inhibitor of many enveloped viruses, including HIV, Ebola, and influenza. It acts to neutralize these viruses by binding to glycoproteins on the viral envelope and preventing viral fusion to the host cell. Although CVN has already been shown to be quite effective against these viruses, we hope to make a variant that has more potential therapeutic value by recruiting activities of the human adaptive immune system. We present here a CVN-Fc chimeric fusion protein. This protein, termed a “lectibody” for its fusion of a lectin (CVN) and the constant region (Fc) of an antibody, is designed to incorporate the viral neutralization properties of CVN with Fc-mediated effector functions, such as antibody-dependent cell-mediated cytotoxicity (ADCC), complement-dependent cytotoxicity (CDC), increased serum half-life, and antibody-dependent cell-mediated phagocytosis (ADCP). Here, we show that a CVN lectibody has similar neutralization activity to wild-type CVN in an anti-HIV assay and that there is significant higher order oligomerization of the protein that is due in some part to domain swapping of CVN. This new class of antiviral protein could act to neutralize free viral particles as well as invoke an immune response surrounding virus-infected cells.

Introduction

Antibodies are a vital component of the mammalian adaptive immune system. They are responsible for neutralizing infectious particles by binding to them and directly inhibiting them as well as by recruiting other components of the immune system, including macrophages, neutrophils, and natural killer (NK) cells, to the site of an infection.^{1,2} An antibody consists of two major regions, the variable region (Fab) and the constant region (Fc). The Fab portion of the antibody is highly variable and is specific to the antigen, whereas the relatively conserved Fc portion contains binding sites for Fc receptors (FcRs) and engages the immune effector functions. There are five major isotypes of antibodies: IgM, IgA, IgD, IgE, and IgG; these differ in their heavy chain sequence and oligomerization state and mediate different responses. While all of these isotypes are important in an immune response, IgG is the most abundant antibody type found in humans, has the longest serum half-life, and is involved in most of the major effector functions.¹ For these reasons, the Fc of IgG1 was chosen for this study.

In addition to direct neutralization of potential pathogens via the Fab regions of an antibody, effector functions mediated through Fc binding are vital to a normally functioning immune system. The Fc of IgG1 specifically interacts with FcRn and Fc receptors specific to the γ chain (Fc γ R), including Fc γ RI, Fc γ RII, and Fc γ RIII.^{1,3-5} These receptors act as messengers, linking antibody-mediated responses to cellular responses. The interaction between Fc and FcRn is involved in recycling antibodies, thereby extending their lifetime *in vivo*, and in transporting antibodies across epithelial barriers.^{4,6-8} The other Fc γ Rs, when complexed with antigen-bound IgG1, can mediate antibody-dependent cell-mediated cytotoxicity (ADCC), antibody-dependent cell-

mediated phagocytosis (ADCP), and endocytosis.³⁻⁵ In addition to FcR-mediated cellular responses, Fc can also activate the complement pathway, which leads to cell lysis or phagocytosis.^{1,9}

Due to their simple protein A-based purification, extended *in vivo* lifetime, and Fc-mediated effector functions, Fc fusion proteins have become increasingly popular.¹⁰⁻¹² In a recent review, Jazayeri and Carroll report that at least six Fc fusion proteins are currently used clinically for several indications, including asthma, psoriasis, and rheumatoid arthritis.¹² In addition, countless other fusions have been made for both pharmaceutical and basic research purposes. While many researchers are interested in the increased lifetime of small, soluble proteins that is conferred by addition of an Fc,^{13,14} Fc fusions have also been used to display Fc in a reverse orientation in order to study Fc-mediated effector functions,¹⁵ to investigate protein-protein interactions,¹⁶ and as potential therapeutics for various diseases or conditions.¹⁰⁻¹² One particularly relevant Fc fusion is CD4-Fc.^{17,18} Various constructs combining the soluble portion of the HIV receptor CD4 with the Fc domain of an antibody were investigated for inhibition of HIV *in vivo*. Unfortunately, the results of clinical trials on these specific constructs were disappointing, but the constructs were able to induce ADCC of HIV-infected cells in culture and were efficiently transferred across the placenta in non-human primates.¹⁹

The ability to incorporate extended *in vivo* lifetimes and activation of cell-mediated effector functions is a very compelling reason to engineer Fc fusion proteins. Additionally, these Fc-mediated functions can be modulated through mutations in the Fc region to either increase or abrogate the response, providing more flexibility to the system.^{20,21} Various studies have indicated that single point mutations^{22,23} or changes in

the Fc-linked carbohydrate composition^{9,24,25} can dramatically increase the ADCC response by increasing the affinity for FcγR. Engineered mutations in the Fc have been shown to increase activation of the complement pathway.²⁶ Alternatively, Lazar *et al.* showed that a point mutation could destroy the ability for an Fc to activate complement-dependent cytotoxicity (CDC) while retaining or enhancing ADCC and other effector functions.²² Extending the lifetime of Fc fusions has also been extensively studied. Even though Fc-fused proteins often already have longer *in vivo* lifetimes than the unfused molecule, any improvements in the circulatory half-life of a molecule is a possible benefit for potential therapeutics. A 2- to 2.5-fold increase in the half-life of Fc fusions was accomplished by either a single or double point mutation in the Fc.^{27,28} The incorporation of one or more of these mutations allows researchers to specifically study the effects of ADCC, complement, and half-life on a particular system.

The protein of interest in this study, cyanovirin-N (CVN), is a small cyanobacterially-derived protein that inhibits infection by various enveloped viruses including HIV,²⁹ Ebola,³⁰ and influenza.³¹ CVN is a lectin that specifically binds α1-2 linked high-mannose molecules.³²⁻³⁵ This type of carbohydrate linkage is found in high concentrations on the envelope proteins of these viruses, including gp120 on HIV.^{34,36} CVN effectively neutralizes HIV by binding with high affinity and avidity to the glycosylation on gp120 and blocking interactions with the host cell receptor, CD4, and coreceptors.³⁷

Here, we report the generation of a CVN-Fc fusion that retains wild-type (WT) CVN-like HIV neutralization activity. We have termed this construct a “lectibody” as it is a fusion between a lectin (CVN) and the Fc of an antibody (Figure 4-1). Similarly to the

CVN₂ dimers discussed in Chapter 2, we hope that we can improve the efficacy of HIV neutralization by dimerizing the CVN through the Fc domain. The CVN lectibody also has the potential for Fc-mediated effector functions as described above. Previous studies have shown that ADCC plays a role in protection against HIV³⁸ and that ADCC and other FcR-mediated effector functions provide some protection against viruses³⁹⁻⁴² even when associated with non-neutralizing antibodies.^{43,44} We also anticipate that this lectibody construct will have a longer half-life *in vivo*. A study on CVN showed that after subcutaneous injection in mice, WT CVN was mostly cleared from the bloodstream after 7 to 24 hours.⁴⁵ Since a daily injection to maintain therapeutic levels would most likely not be feasible, a variant with a longer half-life would make a potential therapeutic more practical. This construct would also benefit from a potential pulmonary delivery route, as Fc fusions have been shown to be effectively transported across the pulmonary epithelial barrier in both humans and non-human primates through FcRn-mediated transcytosis.⁶⁻⁸ Here, we present initial data showing the viability of a CVN-Fc fusion: a lectibody.

Methods

Construct generation. Lectibody constructs were created by subcloning the WT CVN sequence or the CVN₂ L0 sequence described in Chapter 2 including DNA that encodes a five-amino acid linker (GGSGG) between CVN and the Fc of human IgG1 into the baculovirus expression vector pAc-κ-Fc using the XhoI and SpeI restriction sites (Progen Biotechnik). Sequencing on this construct revealed that the Fc portion was missing the last eight residues and included two point mutations. To rectify this, the last eight Fc residues were added during the second cloning step in which the secretion

signal, CVN, and Fc were subcloned using PCR-based techniques into the mammalian expression vector pcDNA3.1 (Invitrogen) or pTT5 (NRC Biotechnology Research Institute), and the mutations were reversed to give the WT Fc sequence. Human-codon optimized CVN sequences were determined using the Custom Gene Synthesis program from IDT (Integrated DNA Technologies, Inc). The optimized gene was assembled via recursive PCR⁴⁶ and ligated into the pcDNA3.1 or pTT5 vector already containing the secretion leader sequence and the Fc sequence. Point mutations were introduced into the lectibody constructs using the QuikChange Site-Directed Mutagenesis kit (Stratagene). All constructs were verified through DNA sequencing.

Bacterially expressed constructs were created as described in Chapter 2. Point mutations for bacterially expressed variants were introduced using the QuikChange Site-Directed Mutagenesis kit (Stratagene).

Expression and purification. Lectibody constructs were expressed in transiently transfected, suspended HEK293-T or HEK293-6E cells (NRC Biotechnology Research Institute). The cells were transfected with 1 mg of plasmid DNA per liter of culture using a polyethylenimine-mediated transfection protocol (PEI). The secreted protein was harvested from the cell supernatants after 6-8 days and buffer exchanged into 100 mM sodium phosphate buffer pH 7.5, 150 mM NaCl. The protein was purified on a Protein A column, eluted in pH 3.0 elution buffer (Pierce) and immediately neutralized with Tris-base. A second purification step on a Superdex-200 gel filtration column (GE Healthcare) in 25 mM sodium phosphate pH 7.4, 150mM NaCl was used to separate high molecular

weight aggregates from smaller species. Protein was stored as eluted or concentrated in a 10,000 MWCO centrifugal concentrator (Millipore) then kept at 4°C.

Deglycosylation of lectibody proteins was accomplished using PNGase F (New England Biolabs). The protein was denatured, then PNGase F was added according to the manufacturer's protocol. Complete deglycosylation was achieved after 1-2 hours. After removing the carbohydrates, the apparent molecular weights of the proteins were assessed by SDS-PAGE.

Bacterial expression and purification of non-Fc fusion constructs were performed as described in Chapter 2 of this thesis.

Circular dichroism. Circular dichroism (CD) spectra were obtained on an Aviv 62DS spectrometer with a 1 mm path length cell. Samples were 50 μ M protein in 25 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl. Wavelength scans were collected at various temperatures between 200 and 250 nm with a 1 nm step size. A single scan was collected for each variant with an averaging time of 5 sec. Temperature denaturation was monitored at 233 nm from 1°C to 99°C. The sample was equilibrated at each temperature for a minimum of 2 minutes before the data was averaged for 30 seconds and recorded. The denaturation curves were not reversible and therefore thermodynamic parameters could not be determined. Instead, the data were fit to a two-state model⁴⁷ to estimate the midpoint of thermal denaturation (T_m), an estimate of thermal stability.

Neutralization assays. Neutralization assays were performed as described in Chapter 2 of this thesis.⁴⁸ All variants were tested against strain SC422661.8 from clade

B and compared to WT CVN from the same 96-well plate unless otherwise noted. Due to the low concentrations of various constructs, some assays were performed with twice the standard volume of protein to increase the final concentration in the well.

Surface plasmon resonance (SPR). SPR experiments were conducted on a T100 instrument from Biacore. Approximately 30 response units (RUs) of bacterially expressed WT CVN were immobilized on a CM5 chip using standard amine coupling. All assays were conducted in HBS-EP buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 0.0005% v/v Surfactant P20, 1 mM EDTA; Biacore). Various analytes were injected over the surface for 60 seconds at a flow rate of 30 μ L/min. The chip was regenerated with two pulses of 50 mM NaOH. Complete regeneration was not achieved after lectibody variants were analyzed and therefore proteins injected later may have exhibited binding to the unregenerated surface and not to the surface itself. We therefore repeated the assay on a new surface and tested the samples in reverse order to confirm the results of the first experiment. The data were analyzed for binding or lack of binding based on the sensorgram.

Results

Mammalian expression. All CVN-Fc (lectibody) constructs were expressed and secreted in mammalian cell culture. Yields were typically low for the pcDNA constructs with *Escherichia coli*-optimized CVN sequences (between 100 and 500 μ g protein per L of cell culture). For comparison, a similar construct containing only the expression leader sequence and Fc expressed approximately 4 mg/L. To try to resolve this problem, we

made various constructs intended to increase protein expression. We found that changing the vector from pcDNA3.1 to pTT5 did not significantly improve the expression and in multiple trials actually produced a larger fraction of degradation product. We did find, however, that changing the codons of the CVN gene to correspond with optimal human codon usage produced an approximately 10% increase in soluble expression. These yields were sufficient for the assays in this study, but are still much lower than desired.

Glycosylation. After Protein A purification, the initial lectibody construct, CVN-Fc, appeared to migrate much slower on an SDS-PAGE gel than expected (data not shown). Therefore, we deglycosylated the protein to confirm the expected molecular weight. However, upon deglycosylation, it became clear that the protein contained two separate N-linked glycosylation sites instead of only the expected site on the Fc. The NetNGlyc 1.0 Server⁴⁹ was used to predict potential N-linked glycosylation sites and found a highly probable site at position 30 of the CVN sequence in addition to the known glycosylation site in the Fc. This potential glycosylation site in CVN is located on the surface of the protein and has the sequence N-T-S, which is consistent with the N-X-(S/T) consensus sequence for N-linked glycosylation (where X is any amino acid except proline).⁵⁰ Visual inspection of the NMR and crystal structures indicated that glycosylation of residue 32 may interfere with substrate binding since this residue is near one of the binding sites of CVN. This result was confirmed by HIV neutralization assays, which showed that CVN-Fc had no neutralization activity (data not shown).

To remove the non-native glycosylation site in CVN, we constructed four variants in the bacterially expressed WT background to assess their effect on the structure and

function of CVN. Both N30 and S32 make side chain-backbone hydrogen bonds in the crystal structure, so we constructed two variants for each position, an Ala mutation that deleted the side chain and a polar mutation that may be able to satisfy the hydrogen bond (N30S and S32N). An S32T mutation would have possibly satisfied the hydrogen bond requirement, but it would have also met the glycosylation consensus and therefore would not have destroyed the site.

The four glycosylation deletion variants were assayed for changes in their secondary structure and thermal stability by CD, and their HIV neutralization abilities were compared to WT (Figure 4-2). No significant differences were seen in the CD wavelength scans of the four variants compared to WT CVN, indicating that the secondary structure was not affected by the mutation (Figure 4-2A). We did see slight differences in the midpoint of thermal denaturation (T_m) of the variants, however (Figure 4-2B). WT and the two N30 mutants had T_m s that were within experimental error (48.8°C to 49.8°C), whereas S32A and S32N were destabilized by approximately 5°C and 8°C, respectively. The results indicate that the physiological temperatures at which the HIV neutralization assays are performed may partially denature the S32N variant, making it less than ideal for our purposes. The neutralization assays showed that the N30A variant was slightly less active than WT, whereas the other three variants were WT-like in their HIV neutralization (Figure 4-2C). All this data together indicated that N30S was the best mutation to incorporate into the lectibody construct. N30S in the background of WT CVN had WT-like HIV neutralization activity, secondary structure, and thermal stability. Additionally, mutation at N30 guarantees the elimination of the N-linked glycosylation,

whereas mutation at position 32 leaves the Asn to which glycosylation would be attached intact, giving rise to a small possibility that glycosylation could still occur.

CVN-Fc N30S. After determining the ideal mutation to remove the non-native glycosylation site from CVN-Fc, we expressed and purified CVN-Fc N30S. This variant, similarly to the WT lectibody, had a significantly higher apparent molecular weight than expected as assayed by gel filtration chromatography, due to higher order oligomers or to aggregation (Figure 4-3). The expected elution volume for dimeric lectibody was approximately 0.60 CV. Although the majority of the protein elutes in the void volume, there was a small peak approximately corresponding to dimeric lectibody. Although Protein A purified protein and fractions containing high molecular weight species showed WT-like HIV neutralization activity, this fraction contained no activity (Figure 4-3B). We therefore sought to solve this unwanted higher order oligomerization problem to obtain monodispersive samples for assaying.

We hypothesized that the low pH elution from the Protein A column may cause some partial denaturation of the CVN portion of the lectibody. We therefore assessed the secondary structure, potential changes in oligomerization, and HIV neutralization of WT bacterially-expressed CVN at various pHs (data not shown). These experiments showed no significant differences between protein in pH 7.4 buffer and protein in buffers down to pH 2.0, including the actual Protein A elution buffer (Pierce). We therefore conclude that WT CVN does not show a pH dependence for the general secondary structure, HIV neutralization, or oligomerization.

Another possibility for the higher order oligomers formed by the lectibodies was that CVN, a carbohydrate binding protein, was binding the glycosylation on Fc and therefore causing large complexes of protein specifically bound to other lectibodies. To test this hypothesis, we expressed the lectibody with an additional mutation (N181A, equivalent to position 297 in a full length heavy chain) that eliminates the native Fc glycosylation site. This variant (CVN-Fc noglycos) behaved similarly to CVN-Fc N30S, and most of the protein eluted near the void volume of the gel filtration column, indicating it was almost completely composed of higher order oligomers. There was no apparent molecular weight shift upon deglycosylating this sample, indicating that we did, indeed remove all of the N-linked glycosylation sites. Although almost entirely oligomerized, CVN-Fc noglycos that was eluted from the Protein A column had approximately WT-like activity (as compared to bacterially expressed WT CVN) in the HIV neutralization assay, indicating that glycosylation is not necessary for the proper folding of the protein or for the activity, as expected. We also tested whether glycosylated Fc could bind WT CVN in an SPR assay (Figure 4-4). We saw no evidence of binding to immobilized CVN and therefore concluded that CVN does not bind the glycosylation on Fc. Interestingly, as seen in Figure 4-4, we saw significant amounts of binding of CVN-Fc N30S and CVN-Fc noglycos to the WT CVN surface. Because we know that the Fc is not responsible for the binding, we deduce that the CVN component of the lectibody is aggregating on the surface. WT CVN, on the other hand, shows no evidence of binding the CVN surface. This evidence suggests that the lectibody, although it contains some active and therefore properly folded protein, probably contains some misfolded protein, which has a tendency to aggregate. Additionally, the lectibody could have alternate

domain-swapping properties for the CVN component, leading to intermolecular domain swapping, with either WT CVN or another lectin protein.

Domain-swapping variant lectin bodies. To assess whether domain swapping of CVN is contributing to the formation of higher order oligomers, we created and assayed two new constructs, CVN₂ L0-Fc and N30S/P51G-Fc. In Chapter 2 of this thesis, I describe the dimeric variants of CVN that we created to test the effects of oligomerization on the efficacy of HIV neutralization. We hypothesize that by covalently linking the termini of two copies of CVN we are stabilizing the domain-swapped dimeric form of CVN, which in the context of WT is only metastable.⁵¹ If this hypothesis is true, the CVN₂ L0 variant described in Chapter 2 should be stably domain-swapped and should not interact with other molecules to form intermolecularly domain-swapped complexes. While the CVN₂ L0-Fc variant showed a significantly lower proportion of high molecular weight species, this protein was not active against HIV. When we added the N30S mutation to this construct, the majority of protein was shifted to high molecular weight and it remained inactive in the HIV neutralization assay. The second domain-swapping variant appears to hold more promise. In this case, the P51G mutation was added to CVN-Fc N30S. P51G has been shown to shift the equilibrium toward monomeric protein and destabilize the domain-swapped form.⁵¹ N30S/P51G-Fc expressed much more readily in the mammalian expression system, although when it was neutralized after the Protein A column, a significant amount of protein precipitated and was lost. The remaining protein, when separated on a gel filtration column, produced a broad peak around 0.44 CV that contained N30S/P51G-Fc as assayed by SDS-PAGE and

showed WT-like activity in the HIV neutralization assay (Figure 4-5). A second large peak at 0.58 CV was attributed to contamination by BSA from the expression process. This peak contained no anti-HIV activity. Although the N30S/P51G-Fc did not elute at the expected volume, it is not forming the very high order oligomers of previous constructs. This indicates that domain swapping is a concern in the lectibody constructs and is an issue that must be overcome.

Discussion

We have successfully created a chimeric CVN-Fc variant that shows WT-like anti-HIV activity. We have shown that a non-native glycosylation site is present in CVN and glycosylated in mammalian tissue culture and that that site must be removed for efficient viral neutralization activity. In addition, we have shown that the lectibody constructs are prone to formation of higher order oligomers, which can in part be prevented by using a variant that stabilizes the monomeric state of CVN over the domain-swapped dimer.⁵¹ Although more work is required to create completely monodispersed lectibody, we have shown that it is possible to modulate the oligomerization through simple mutation. Additional obstacles to overcome include solving a potential misfolding problem that allows the lectibodies to bind to WT CVN, as evidenced by the Biacore experiments.

Unlike in the case of the dimeric CVN molecules (CVN₂s) described in Chapter 2 of this thesis, we do not see a significant increase in the anti-HIV activity of the lectibody as compared to WT CVN attributable to the dimerization. This may be due to the fact that we have yet to isolate pure dimeric lectibody and the samples may be significantly

contaminated by partially or fully unfolded, nonfunctional protein. Additionally, some of the carbohydrate binding sites on CVN could be sterically inhibited by the high order oligomerization that we are seeing. By generating a variant that is monodispersed and dimeric, we hope to resolve these issues. It is also possible that the conformation of the Fc does not allow the CVNs to interact in a way that they are able to in the pure CVN₂ samples, perhaps not allowing the proper domain-swapping interactions.

Our success in generating a functional lectibody is only partially complete. In order to fully realize the goals of this project, we must assay the Fc effector function and *in vivo* half-life. Although we expect the lectibody to exhibit all the potential functions of the Fc, we hope to assay both antibody-dependent cell-mediated cytotoxicity (ADCC) and complement functions in the near future to verify.

Acknowledgements

I would like to thank Jost Vielmetter and Michael Anaya of the Protein Expression Center at Caltech for expression and initial purification of all the Fc fusion proteins. I would also like to acknowledge Priyanthi Peiris for running the HIV neutralization assays that are presented in this work. Josh Klein supplied the initial Fc sequence as well as helpful advice at all stages of this project. The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: SVPB8 (Drs. David Montefiori and Feng Gao); pSG3^{Aenv} (Drs. John C. Kappes and Xiaoyun Wu); Tzm-B1 cells (Dr. John C. Kappes, Dr. Xiaoyun Wu, and Tranzyme, Inc.)

References

1. Janeway, C., Travers, P., Walport, M. & Sholmchik, M. J. (2005). *Immunobiology: the immune system in health and disease*. 6th edit, Garland Science Publishing, New York, NY.
2. Huber, M. & Trkola, A. (2007). Humoral immunity to HIV-1: neutralization and beyond. *J Intern Med* **262**, 5-25.
3. Daeron, M. (1997). Fc receptor biology. *Annu Rev Immunol* **15**, 203-34.
4. Raghavan, M. & Bjorkman, P. J. (1996). Fc receptors and their interactions with immunoglobulins. *Annu Rev Cell Dev Biol* **12**, 181-220.
5. Gessner, J. E., Heiken, H., Tamm, A. & Schmidt, R. E. (1998). The IgG Fc receptor family. *Ann Hematol* **76**, 231-48.
6. Bitonti, A. J., Dumont, J. A., Low, S. C., Peters, R. T., Kropp, K. E., Palombella, V. J., Stattel, J. M., Lu, Y., Tan, C. A., Song, J. J., Garcia, A. M., Simister, N. E., Spiekermann, G. M., Lencer, W. I. & Blumberg, R. S. (2004). Pulmonary delivery of an erythropoietin Fc fusion protein in non-human primates through an immunoglobulin transport pathway. *Proc Natl Acad Sci U S A* **101**, 9763-8.
7. Dumont, J. A., Bitonti, A. J., Clark, D., Evans, S., Pickford, M. & Newman, S. P. (2005). Delivery of an erythropoietin-Fc fusion protein by inhalation in humans through an immunoglobulin transport pathway. *J Aerosol Med* **18**, 294-303.
8. Low, S. C., Nunes, S. L., Bitonti, A. J. & Dumont, J. A. (2005). Oral and pulmonary delivery of FSH-Fc fusion proteins via neonatal Fc receptor-mediated transcytosis. *Hum Reprod* **20**, 1805-13.
9. Jefferis, R., Lund, J. & Pound, J. D. (1998). IgG-Fc-mediated effector functions: molecular definition of interaction sites for effector ligands and the role of glycosylation. *Immunol Rev* **163**, 59-76.
10. Ashkenazi, A., Capon, D. J. & Ward, R. H. (1993). Immunoadhesins. *Int Rev Immunol* **10**, 219-27.
11. Chamow, S. M. & Ashkenazi, A. (1996). Immunoadhesins: principles and applications. *Trends Biotechnol* **14**, 52-60.
12. Jazayeri, J. A. & Carroll, G. J. (2008). Fc-based cytokines : prospects for engineering superior therapeutics. *BioDrugs* **22**, 11-26.
13. Dwyer, M. A., Huang, A. J., Pan, C. Q. & Lazarus, R. A. (1999). Expression and characterization of a DNase I-Fc fusion enzyme. *J Biol Chem* **274**, 9738-43.

14. Shapiro, R. I., Wen, D., Levesque, M., Hronowski, X., Gill, A., Garber, E. A., Galdes, A., Strauch, K. L. & Taylor, F. R. (2003). Expression of Sonic hedgehog-Fc fusion protein in *Pichia pastoris*. Identification and control of post-translational, chemical, and proteolytic modifications. *Protein Expr Purif* **29**, 272-83.
15. Stabila, P. F., Wong, S. C., Kaplan, F. A. & Tao, W. (1998). Cell surface expression of a human IgG Fc chimera activates macrophages through Fc receptors. *Nat Biotechnol* **16**, 1357-60.
16. Gurbaxani, B. M. & Morrison, S. L. (2006). Development of new models for the analysis of Fc-FcRn interactions. *Mol Immunol* **43**, 1379-89.
17. Capon, D. J., Chamow, S. M., Mordenti, J., Marsters, S. A., Gregory, T., Mitsuya, H., Byrn, R. A., Lucas, C., Wurm, F. M., Groopman, J. E. & et al. (1989). Designing CD4 immunoadhesins for AIDS therapy. *Nature* **337**, 525-31.
18. Langner, K. D., Niedrig, M., Fultz, P., Anderson, D., Reiner, G., Repke, H., Gelderblom, H., Seed, B., Hilfenhaus, J. & Zettlmeissl, G. (1993). Antiviral effects of different CD4-immunoglobulin constructs against HIV-1 and SIV: immunological characterization, pharmacokinetic data and in vivo experiments. *Arch Virol* **130**, 157-70.
19. Byrn, R. A., Mordenti, J., Lucas, C., Smith, D., Marsters, S. A., Johnson, J. S., Cossum, P., Chamow, S. M., Wurm, F. M., Gregory, T. & et al. (1990). Biological properties of a CD4 immunoadhesin. *Nature* **344**, 667-70.
20. Presta, L. G. (2005). Selection, design, and engineering of therapeutic antibodies. *J Allergy Clin Immunol* **116**, 731-6; quiz 737.
21. Presta, L. G. (2008). Molecular engineering and design of therapeutic antibodies. *Curr Opin Immunol* **20**, 460-70.
22. Lazar, G. A., Dang, W., Karki, S., Vafa, O., Peng, J. S., Hyun, L., Chan, C., Chung, H. S., Eivazi, A., Yoder, S. C., Vielmetter, J., Carmichael, D. F., Hayes, R. J. & Dahiyat, B. I. (2006). Engineered antibody Fc variants with enhanced effector function. *Proc Natl Acad Sci U S A* **103**, 4005-10.
23. Shields, R. L., Namenuk, A. K., Hong, K., Meng, Y. G., Rae, J., Briggs, J., Xie, D., Lai, J., Stadlen, A., Li, B., Fox, J. A. & Presta, L. G. (2001). High resolution mapping of the binding site on human IgG1 for Fc gamma RI, Fc gamma RII, Fc gamma RIII, and FcRn and design of IgG1 variants with improved binding to the Fc gamma R. *J Biol Chem* **276**, 6591-604.
24. Shields, R. L., Lai, J., Keck, R., O'Connell, L. Y., Hong, K., Meng, Y. G., Weikert, S. H. & Presta, L. G. (2002). Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fc gamma RIII and antibody-dependent cellular toxicity. *J Biol Chem* **277**, 26733-40.

25. Shinkawa, T., Nakamura, K., Yamane, N., Shoji-Hosaka, E., Kanda, Y., Sakurada, M., Uchida, K., Anazawa, H., Satoh, M., Yamasaki, M., Hanai, N. & Shitara, K. (2003). The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. *J Biol Chem* **278**, 3466-73.
26. Idusogie, E. E., Wong, P. Y., Presta, L. G., Gazzano-Santoro, H., Totpal, K., Ultsch, M. & Mulkerrin, M. G. (2001). Engineered antibodies with increased activity to recruit complement. *J Immunol* **166**, 2571-5.
27. Hinton, P. R., Johlf, M. G., Xiong, J. M., Hanestad, K., Ong, K. C., Bullock, C., Keller, S., Tang, M. T., Tso, J. Y., Vasquez, M. & Tsurushita, N. (2004). Engineered human IgG antibodies with longer serum half-lives in primates. *J Biol Chem* **279**, 6213-6.
28. Hinton, P. R., Xiong, J. M., Johlf, M. G., Tang, M. T., Keller, S. & Tsurushita, N. (2006). An engineered human IgG1 antibody with longer serum half-life. *J Immunol* **176**, 346-56.
29. Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T., Gulakowski, R. J., Wu, L., Rivera, M. I., Laurencot, C. M., Currens, M. J., Cardellina, J. H., 2nd, Buckheit, R. W., Jr., Nara, P. L., Pannell, L. K., Sowder, R. C., 2nd & Henderson, L. E. (1997). Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother* **41**, 1521-30.
30. Barrientos, L. G., O'Keefe, B. R., Bray, M., Sanchez, A., Gronenborn, A. M. & Boyd, M. R. (2003). Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral Res* **58**, 47-56.
31. O'Keefe, B. R., Smee, D. F., Turpin, J. A., Saucedo, C. J., Gustafson, K. R., Mori, T., Blakeslee, D., Buckheit, R. & Boyd, M. R. (2003). Potent anti-influenza activity of cyanovirin-N and interactions with viral hemagglutinin. *Antimicrob Agents Chemother* **47**, 2518-25.
32. Bewley, C. A. (2001). Solution structure of a cyanovirin-N:Man alpha 1-2Man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure* **9**, 931-40.
33. Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H., Boyd, M. R. & Wlodawer, A. (2002). Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem* **277**, 34336-42.
34. Bolmstedt, A. J., O'Keefe, B. R., Shenoy, S. R., McMahon, J. B. & Boyd, M. R. (2001). Cyanovirin-N defines a new class of antiviral agent targeting N-linked,

- high-mannose glycans in an oligosaccharide-specific manner. *Mol Pharmacol* **59**, 949-54.
35. O'Keefe, B. R., Shenoy, S. R., Xie, D., Zhang, W., Muschik, J. M., Currens, M. J., Chaiken, I. & Boyd, M. R. (2000). Analysis of the interaction between the HIV-inactivating protein cyanovirin-N and soluble forms of the envelope glycoproteins gp120 and gp41. *Mol Pharmacol* **58**, 982-92.
 36. Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. J., Thomas, J. N. & Gregory, T. J. (1990). Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem* **265**, 10373-82.
 37. Dey, B., Lerner, D. L., Lusso, P., Boyd, M. R., Elder, J. H. & Berger, E. A. (2000). Multiple antiviral activities of cyanovirin-N: blocking of human immunodeficiency virus type 1 gp120 interaction with CD4 and coreceptor and inhibition of diverse enveloped viruses. *J Virol* **74**, 4562-9.
 38. Hessel, A. J., Hangartner, L., Hunter, M., Havenith, C. E., Beurskens, F. J., Bakker, J. M., Lanigan, C. M., Landucci, G., Forthal, D. N., Parren, P. W., Marx, P. A. & Burton, D. R. (2007). Fc receptor but not complement binding is important in antibody protection against HIV. *Nature* **449**, 101-4.
 39. Florese, R. H., Van Rompay, K. K., Aldrich, K., Forthal, D. N., Landucci, G., Mahalanabis, M., Haigwood, N., Venzon, D., Kalyanaraman, V. S., Marthas, M. L. & Robert-Guroff, M. (2006). Evaluation of passively transferred, nonneutralizing antibody-dependent cellular cytotoxicity-mediating IgG in protection of neonatal rhesus macaques against oral SIVmac251 challenge. *J Immunol* **177**, 4028-36.
 40. Gomez-Roman, V. R., Patterson, L. J., Venzon, D., Liewehr, D., Aldrich, K., Florese, R. & Robert-Guroff, M. (2005). Vaccine-elicited antibodies mediate antibody-dependent cellular cytotoxicity correlated with significantly reduced acute viremia in rhesus macaques challenged with SIVmac251. *J Immunol* **174**, 2185-9.
 41. Huber, V. C., Lynch, J. M., Bucher, D. J., Le, J. & Metzger, D. W. (2001). Fc receptor-mediated phagocytosis makes a significant contribution to clearance of influenza virus infections. *J Immunol* **166**, 7381-8.
 42. Baldrige, J. R. & Buchmeier, M. J. (1992). Mechanisms of antibody-mediated protection against lymphocytic choriomeningitis virus infection: mother-to-baby transfer of humoral protection. *J Virol* **66**, 4252-7.
 43. Holl, V., Peressin, M., Decoville, T., Schmidt, S., Zolla-Pazner, S., Aubertin, A. M. & Moog, C. (2006). Nonneutralizing antibodies are able to inhibit human

immunodeficiency virus type 1 replication in macrophages and immature dendritic cells. *J Virol* **80**, 6177-81.

44. Holl, V., Peressin, M., Schmidt, S., Decoville, T., Zolla-Pazner, S., Aubertin, A. M. & Moog, C. (2006). Efficient inhibition of HIV-1 replication in human immature monocyte-derived dendritic cells by purified anti-HIV-1 IgG without induction of maturation. *Blood* **107**, 4466-74.
45. Bringans, S. D., O'Keefe, B. R., Bray, M., Whitehouse, C. A. & Boyd, M. R. (2004). Development of a fluorescent microplate assay for determining cyanovirin-N levels in plasma. *Anal Bioanal Chem* **380**, 269-74.
46. Stemmer, W. P., Cramer, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49-53.
47. Becketl, W. J. & Schellman, J. A. (1987). Protein stability curves. *Biopolymers* **26**, 1859-77.
48. Li, M., Gao, F., Mascola, J. R., Stamatatos, L., Polonis, V. R., Koutsoukos, M., Voss, G., Goepfert, P., Gilbert, P., Greene, K. M., Bilaska, M., Kothe, D. L., Salazar-Gonzalez, J. F., Wei, X. P., Decker, J. M., Hahn, B. H. & Montefiori, D. C. (2005). Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* **79**, 10108-10125.
49. Gupta, R., Jung E., & Brunak, S. (2004). Prediction of N-glycosylation sites in human proteins. *In preparation*. <http://www.cbs.dtu.dk/services/NetNGlyc/>.
50. Imperiali, B. & O'Connor, S. E. (1999). Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol* **3**, 643-9.
51. Barrientos, L. G., Louis, J. M., Botos, I., Mori, T., Han, Z., O'Keefe, B. R., Boyd, M. R., Wlodawer, A. & Gronenborn, A. M. (2002). The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure* **10**, 673-86.
52. Saphire, E. O., Parren, P. W., Pantophlet, R., Zwick, M. B., Morris, G. M., Rudd, P. M., Dwek, R. A., Stanfield, R. L., Burton, D. R. & Wilson, I. A. (2001). Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *Science* **293**, 1155-9.

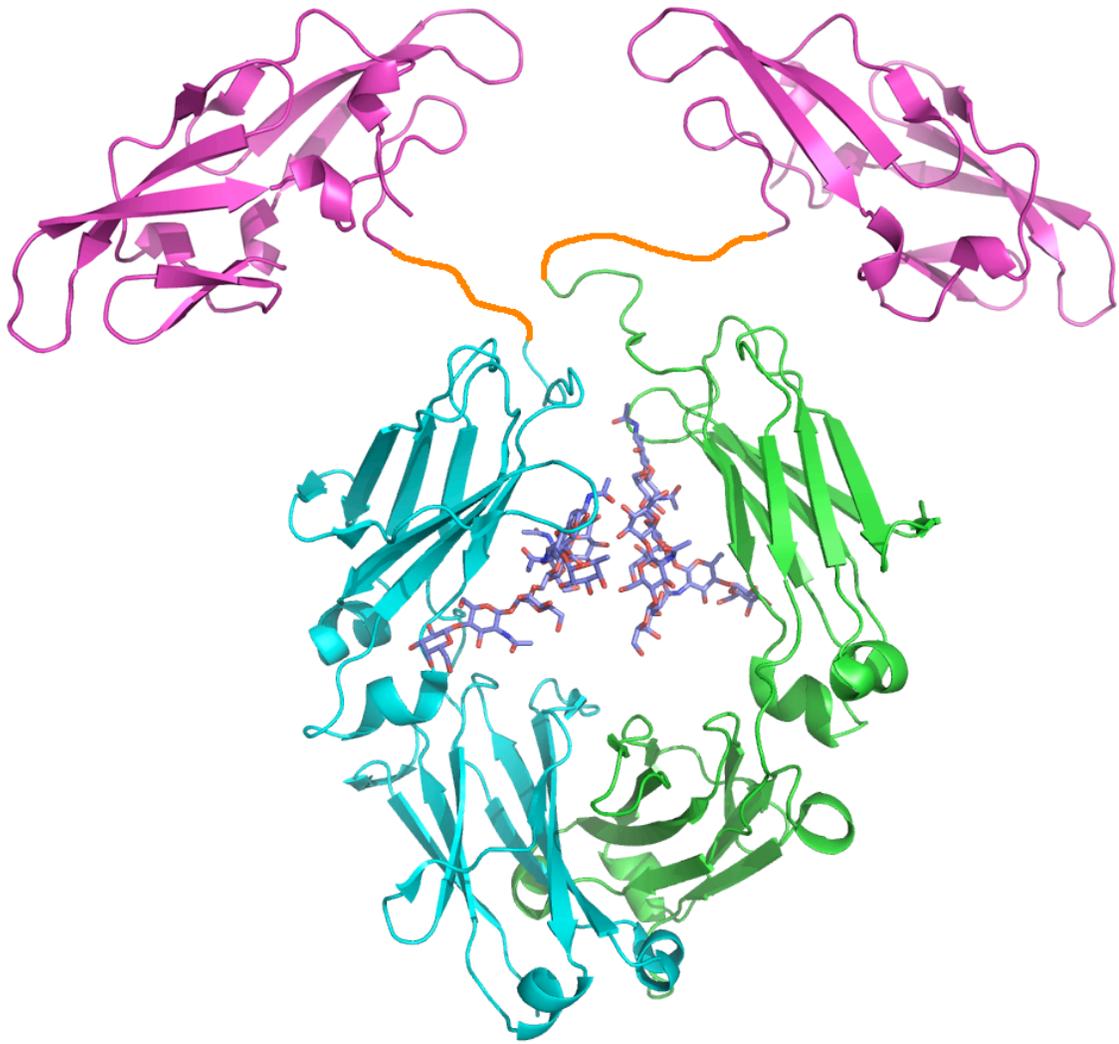


Figure 4-1. Model of the CVN-Fc lectibody. The CVN monomers are shown in magenta attached to the Fc (cyan and green) through flexible polypeptide linkers shown in orange. The Fc glycosylation is shown in stick representation with blue carbons. This model was created by combining a monomeric NMR structure of CVN³² and the Fc from the IgG1b12 crystal structure (1HZH)⁵² in Adobe Photoshop and is not a solved structure of this variant.

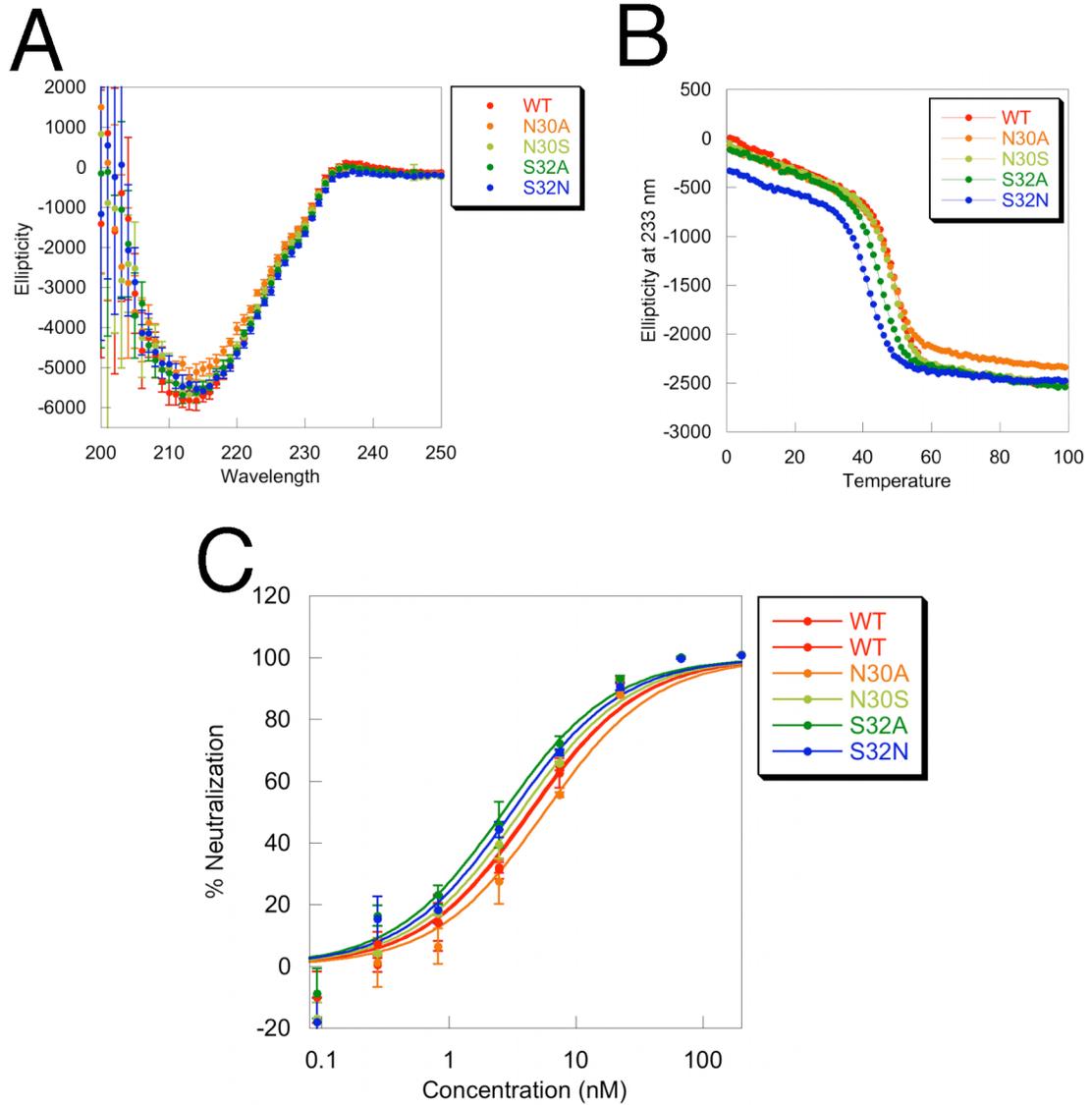


Figure 4-2. Assessment of glycosylation site deletion variants. (A) CD wavelength scans of the four variants compared to WT CVN. (B) Thermal denaturation of WT and the variants monitored by CD at 233 nm. (C) HIV neutralization curves of glycosylation site variants and WT.

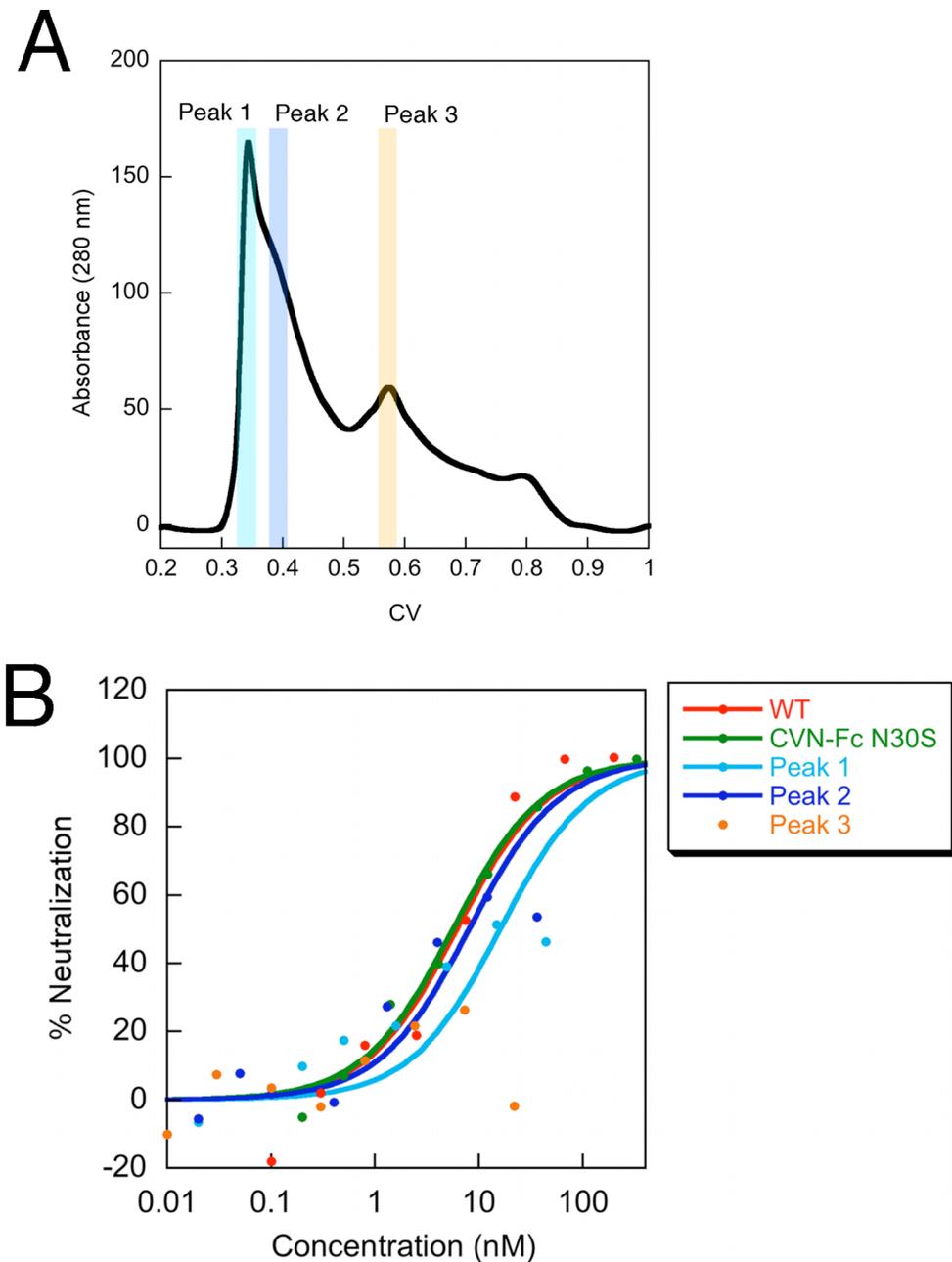


Figure 4-3. CVN-Fc N30S purification and activity. (A) A gel filtration trace of CVN-Fc N30S shows that the majority of the protein forms high order oligomers. (B) CVN-Fc N30S has WT-like HIV neutralization activity, but the active protein is the high molecular weight species and not from Peak 3 which corresponds to dimeric lectibody. No curve fit is shown for Peak 3 due to the low neutralization activity.

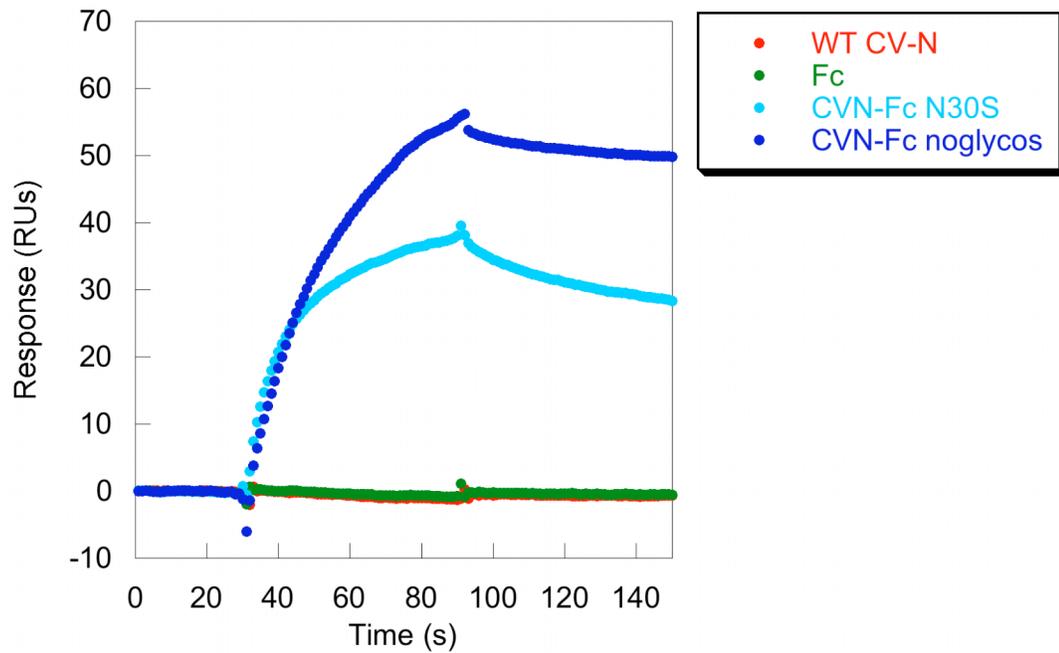


Figure 4-4. Surface plasmon resonance assays of lectin bodies and Fc. WT CVN was immobilized on the surface and various proteins were analyzed for binding. WT CVN and human glycosylated Fc did not bind to the surface. However, both lectin body constructs (CVN-Fc N30S and CVN-Fc noglycos) showed significant interaction with the WT CVN surface that could not be regenerated.

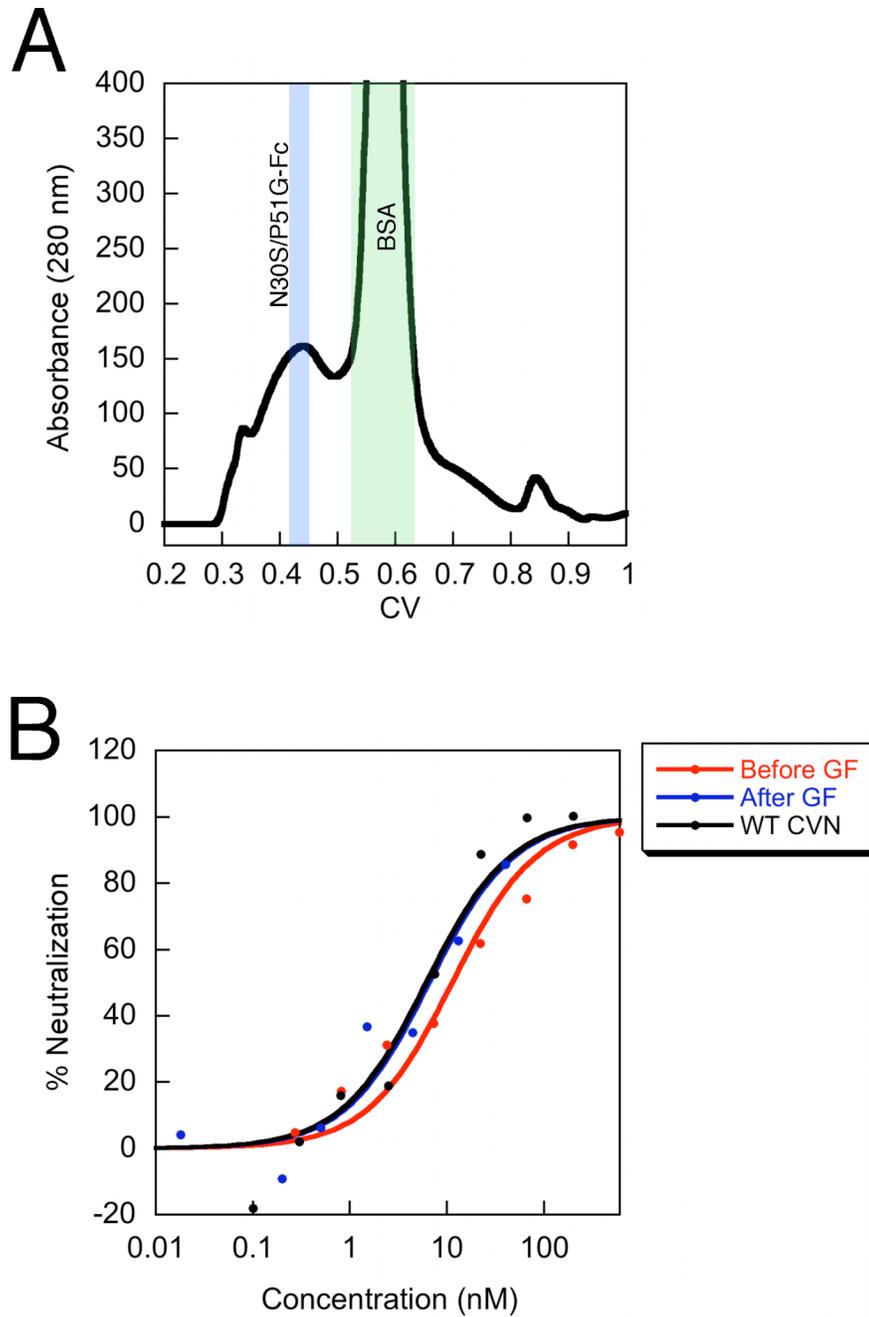


Figure 4-5. N30S/P51G-Fc purification and activity. (A) Gel filtration of N30S/P51G-Fc shows a broad peak corresponding to the lectibody around 0.44 CV. This sample was heavily contaminated by BSA from the expression (0.58 CV). (B) N30S/P51G-Fc is active both before (red data) and after (blue data) gel filtration. Due to lack of space on the assay plate, WT CVN was not run on this plate. A reference curve from a previous assay for WT is shown in black. The small void volume peak and the BSA peak both showed no anti-HIV activity (data not shown).

Appendix A

**Toward computationally designed calmodulin variants with
enhanced peptide binding specificity**

Abstract

Calmodulin (CaM) is a second messenger protein that binds a wide variety of natural protein and peptide substrates. Although these interactions are of high affinity, CaM shows little specificity among its native substrates. We used computational design including explicit negative design to engineer CaM variants that preferentially bind one natural peptide substrate over another. By specifically modeling the structures in complex with both the desired and undesired substrates, negative design methods allow us to predict variants that will bind the desired substrate (smMLCK peptide) with high affinity and the undesired substrate (CaMKI peptide) with lower affinity. We found that one of our variants (M51Y) showed a 10-fold preference for smMLCK_p over CaMKI_p in a fluorescence assay, as predicted from the negative design calculation. Additionally, we show that a surface plasmon resonance (SPR)-based method is more effective at determining the dissociation constants for these high affinity interactions than the commonly used fluorescence-based assay. These initial studies provide a strong foundation for future work on the CaM-peptide system.

Introduction

Protein-protein interactions. Protein-protein interactions are necessary for every cellular process: from signal transduction to regulation of gene expression to apoptosis to intracellular transport. The protein-protein interfaces involved in these interactions are, by necessity, specific to the binding partners and exhibit vast diversity in their size, types of interactions, and shape of the interaction surface. To better understand cellular processes and the proteome, several groups have focused on modeling and predicting protein-protein interactions.¹⁻³ A complementary approach is to design novel or improved protein-protein interactions to test our knowledge and understanding of the forces that are important for both affinity and specificity.

The ability to design protein-protein interfaces with high affinity and specificity could be valuable in developing new or improved protein therapeutics as well as in advancing basic scientific research. Novel interfaces have already been generated to create restriction enzymes with new DNA recognition sequences,^{4,5} new Ca²⁺ sensors that can be used *in vivo* to study variations in calcium concentration,⁶ and new orthogonal protein-peptide pairs that can provide an alternative for current affinity purification and pull-down assays.⁷

Design of protein-protein interactions. The importance of protein-protein interactions in understanding biological systems has stimulated a vast amount of research aimed at designing and modulating protein-protein interfaces. Various techniques have been employed to optimize or alter protein-protein contacts, including selection schemes to assay libraries of variants,⁸⁻¹⁰ structure-based rational design,¹¹ and computational

methods. Computational design methods have become increasingly valuable for modulating protein-protein interfaces because they allow researchers to screen many more sequences than would be realistic using experimental methods. This technique has led to the success of several designed protein-protein interfaces, including engineered obligate heterodimeric endonucleases that recognize novel DNA targets,^{4,5} new orthogonal interacting pairs with calmodulin⁶ and PDZ domains,⁷ protein-inhibitor interfaces with redesigned specificities,¹²⁻¹⁴ and novel protein-protein interfaces from monomeric proteins.¹⁵

Calmodulin-peptide complexes have proven to be useful systems for studying specificity^{16,17} as well as for designing orthogonal interfaces.^{6,18,19} Previous work from our group showed that modest specificity enhancements could be obtained by stabilizing one CaM-peptide structure without taking any negative design into account.¹⁶ Other researchers showed that by designing both the peptide and the calmodulin surface, highly specific interactions could be formed.⁶ However, not all studies have found calmodulin specificity to be easily modulated. Green *et al.* found that although they could design new calmodulin-peptide pairs that display high affinity for each other, the designed molecules did not show enhanced specificity.¹⁹

Various strategies for optimizing or modulating interfaces have been employed in computational protein interface design. A “knob-and-hole” strategy in which unfavorable mutations are introduced into one binding partner and the other binding partner is designed with compensating mutations is very common.¹³ This is an efficient method for developing novel interfaces, but is less useful for altering specificity among native substrates. Additionally, burial of more hydrophobic surface area²⁰ or increasing the

amount of surface area in the interface²¹ have been valuable approaches for enhancing the binding affinities of protein-protein or protein-peptide interactions; however, they do not necessarily provide specificity of interaction.

Although much work has already been done, we still face a number of challenges in the design of protein-protein interactions. While standard optimization methods can lead to stabilized complexes and potentially slight increases in specificity, designs are often more successful when negative design is included.^{22,23} Bolon *et al.* showed that although heterodimers could be designed using only positive design methods, the resulting molecules were more specific if unfavorable structures (homodimers) were also modeled and specifically selected against during the optimization.²³

In this study, we investigated the specificity of the native calmodulin system. In a continuation of previous results from our group,^{16,17} we hoped to modulate the specificity of calmodulin between two native peptide substrates by explicitly modeling both bound states in our calculations. This work differs from previous studies in that our intention was not to generate a new interface, but to impart specificity between two substrates that already bind to the target protein with high affinity.

Computational protein design using ORBIT. ORBIT (Optimization of Rotamers By Iterative Techniques) is a fully automated computational protein design software suite developed in the Mayo laboratory. This method has been used with great success to address a number of protein design problems. ORBIT was effectively used to design a zinc finger motif without zinc, and the predicted sequence was subsequently shown to adopt the expected fold.²⁴ Many proteins were designed to be more thermodynamically

stable than their wild-type counterparts.^{25,26} More recently, ORBIT has been used to modulate the function of proteins. In 2002, Datta *et al.* used ORBIT to design a tRNA-synthetase that allows an unnatural amino acid to be incorporated into proteins *in vivo*.²⁷ Modest success has been made in the area of enzyme design. ORBIT was used to design an active site onto an inert protein scaffold and generate an “enzyme-like protein” that exhibits catalytic activity,²⁸ and a more efficient variant was created by optimizing the residues surrounding the active site of a native enzyme.²⁹

To reduce the combinatorial complexity inherent in computational design, ORBIT optimizations begin with a fixed backbone taken from a high-resolution crystal structure. In addition, discrete rotamers or conformers are used to model the flexibility of the side chains within the protein. Rotamers are idealized versions of the average conformation of rotamers found in protein structures in the Protein Data Bank (PDB,³⁰ whereas conformers are side-chain conformations that are the closest to the average conformation within a defined area of phi/psi space.³¹ These conformers are actual 3D coordinates from structures and not idealized averages, and therefore may represent more realistic conformations for side chains. A backbone-dependent or backbone-independent rotamer library can be used to limit the computation time and select for more physically relevant side-chain conformations.

The ORBIT scoring function is based on the DREIDING force field³²; it calculates the energy of a given conformation of a protein sequence, and incorporates four physical forces that are believed to be important in protein design and folding. The first component is a scaled van der Waals interaction energy that includes both a short-range repulsive component as well as a long-range attractive component.³³ Hydrogen

bonding energies are calculated using a distance-, angle-, and hybridization-dependent hydrogen bond scoring function. The third component is a distance-dependent electrostatic potential.³³ Finally, a surface area-based solvation potential is included that applies a benefit for burial of non-polar surface area, a penalty for burial of polar surface area, and a penalty for exposure of non-polar surface area.³⁴ After the energies are calculated using the scoring function, various deterministic³⁵⁻³⁷ and stochastic³⁸⁻⁴¹ search algorithms are used to find either the best sequence (the global minimum energy conformation or GMEC) or other low energy sequences that are consistent with the design and backbone.

Previous studies using ORBIT, including those on calmodulin,^{16,17} were performed using only positive computational design. That is, a sequence is determined using the scoring function and the desired structure in order to stabilize the fold specified. In the case of calmodulin designs, this was a successful strategy for creating a variant that showed specificity between some peptides, but not all of those tested. In this study, we added a negative design component to the computational design in which we explicitly model the undesired state (the protein bound to a different peptide) and select for sequences that favor the desired structure while disfavoring the undesired one. By incorporating this additional information, we hoped to validate our negative design methods as well as develop methods for incorporating specificity into a protein system.

Calmodulin-peptide system. Calmodulin (CaM) is a highly conserved, ubiquitous protein found in all eukaryotes. It is a calcium-dependent regulatory protein that binds to and modulates the function of many cellular proteins including those involved in

phosphorylation, muscle contraction, cell proliferation, and metabolism.^{42,43} CaM regulates proteins in multiple ways, the most common of which is by relieving autoinhibition of the target by binding to the autoinhibitory domain (AID).⁴⁴ Although CaM binds with broad specificity to a vast array of different proteins, many of these targets contain a basic amphiphilic α -helical peptide to which CaM can bind with high affinity.⁴⁵⁻⁴⁷ This natural high affinity and broad specificity makes CaM an ideal candidate for studying protein-protein interactions, protein specificity, and negative design.

Currently, over 15 high-resolution structures are available of CaM bound to peptides in the most common conformation in which the two domains of CaM wrap around the helical peptide (Figure A-1). When bound to a target peptide in this manner, CaM forms a compact α -helical structure bound to four Ca^{2+} ions which envelopes the amphiphilic peptide. Within the protein-peptide interface, eight methionine residues and a flexible central helix are thought to provide a great deal of conformational elasticity that may allow the CaM to adjust to many different peptide substrates.^{46,48}

For this study, we chose to investigate protein-peptide specificity using two CaM-peptide structures: smMLCK_p-CaM⁴⁹ and CaMKI_p-CaM.⁵⁰ These high-resolution crystal structures have very similar backbone orientations, with an RMSD of 1.4 Å. However, the peptides show little sequence homology with the exception of the tryptophan, which anchors the peptide and is common in CaM-binding peptides (see Table A-1). Additionally, it has been shown that CaM can distinguish between these and other peptides by binding them with slight backbone differences to accommodate the peptide side chains.^{48,51} These minor differences in structure lead us to believe that we can

computationally design variants to show specificity between the two high affinity substrates smMLCK_p and CaMKI_p by using negative design techniques that explicitly model both complexed structures.

Methods

ORBIT calculations. ORBIT optimization calculations were run on the protein-peptide interface in both the smMLCK_p-CaM cocrystal structure (1CDL)⁴⁹ and the CaMKI_p-CaM cocrystal structure (1MXE).⁵⁰ To prepare the structures for calculations, water molecules and calcium ions were removed from the pdb files for each structure and Reduce, a program in the MolProbity suite, was used to add hydrogen atoms.⁵² A side-chain placement calculation was run to add side chains for residues 77 and 115, which were missing from the 1CDL structure. In addition, 99 Phe and 143 Thr were placed in the 1CDL structure to mutate the sequence from the human CaM to Drosophila CaM and therefore be consistent with the 1MXE sequence. All four of these residues are distal from the protein-peptide interface and were therefore not expected to have an impact in the ultimate design. 1CDL and 1MXE were minimized for 50 steps using the DREIDING force field.³² Each residue was characterized as surface, boundary, or core based on the distance of its C α and C β to the solvent-accessible surface.²⁴

Positive designs were first conducted on both structures. Residues that were within 4 Å of the peptide in both structures, were defined as core, and were either hydrophobic or glutamate (Glu) in the wild-type protein were allowed to vary in the calculations. Residues 11, 12, 15, 18, 19, 32, 39, 55, 68, 84, 88, 91, 92, 105, 108, 112, and 128 were allowed to sample Ala, Val, Leu, Ile, Phe, Tyr, and Glu amino acids.

Residues 36, 51, 71, 72, 109, 124, 144, and 145 are methionine in the wild-type protein and therefore were allowed to sample conformations of Met in addition to the amino acids listed above. Wild-type CaM contains no tryptophans and Trp was not included in the designs to simplify the binding analysis by Trp fluorescence. The side chains in the peptides were allowed to change conformation but not identity.

Two positive designs were carried out on each structure. In the first set of designs, a backbone-dependent rotamer library in which each rotamer was expanded by 1 standard deviation around χ_1 and χ_2 was used (e2).³⁰ These results are designated “peptide-e2.” In the second set of designs, a backbone-independent conformer library was used.³¹ Results from these calculations are designated “peptide-conf.”

Standard ORBIT parameters were used for these calculations^{33,34} with the exception of a decreased polar burial penalty (0.03 kcal/mol/Å²) to not overly penalize Glu rotamers, and a 100 kcal/mol cutoff energy. The increased cutoff energy allowed us to keep rotamers that might be necessary in the negative design calculation. To place more emphasis on the interactions between CaM and the peptides, we introduced a scale factor to bias the intermolecular interaction energies.¹⁷ This scale factor resulted in recovery of Glu84, which forms favorable interactions with the peptides but has unfavorable electrostatics with the rest of CaM. The scale factors used were 1.3 (conf designs) and 1.1 (e2 designs). Optimization of the positive designs was accomplished using a modification of the FASTER algorithm, giving a single low-energy sequence, the FMEC.^{38,39}

Negative designs were scored using a simple scoring function (Equation A-1), where ΔE_{pos} and ΔE_{neg} are the energies of a given sequence in the context of the desired

complex and the undesired complex, respectively. W is an arbitrary weighting factor used to balance the desire for a good energy for the positive structure and a bad energy for the negative structure. The weighting factor for these calculations was 0.2.

$$Score = \Delta E_{pos} - w * \Delta E_{neg} \quad (A-1)$$

The negative design calculations were performed using the 1CDL structure (CaM-smMLCK_p) as the desired complex and the 1MXE structure (CaM-CaMKI_p) as the undesired complex. The energy calculations from the positive designs (described above) were used as input and the score was minimized according to Equation A-1 using FASTER³⁹ with 200 trajectories. The single sequence with the lowest score was investigated further. This protocol was completed for both the conformer library designs and the standard rotamer library designs.

Construct generation. Wild-type *Drosophila* CaM and CaM8 (an eight-fold mutant previously shown to have enhanced specificity) constructs in pET15b were obtained from Julia M. Shifman.^{16,17} Variant constructs containing a single mutation were created using inverse PCR mutagenesis,⁵³ and constructs with two or more mutations were generated using the Quikchange multi-site mutagenesis kit (Stratagene). All constructs were verified through DNA sequencing.

Protein expression and purification. Wild-type CaM and all variants were expressed in *E. coli* BL21(DE3). All growth was conducted in LB supplemented with 100 µg/mL ampicillin. Expression was induced at OD₆₀₀ = 0.6 with 1 mM IPTG and cells

were grown for 2-4 hours at 37 °C. Cells were harvested at 5,000 × g for 10 minutes in a centrifuge and the cell pellet resuspended in 25 mL of Lysis Buffer (50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 5 mM CaCl₂). Resuspended cells were frozen at -20 °C overnight. After thawing, 1 mM PMSF, DNase, and RNase were added and the cells were lysed using an Emulsiflex (Avestin) for 3-5 minutes. The lysate was clarified by centrifuging for 30 minutes at 21,000 × g and filtered through a 0.22 μm filter before chromatography was conducted. The clarified lysate containing the CaM was first purified on a Phenyl Sepharose HP HiLoad 16/10 column (General Electric) preequilibrated with Buffer A (50 mM Tris-HCl pH 7.5, 1 mM CaCl₂), essentially as described.⁵⁴ Briefly, the clarified lysate was loaded on the column, washed with 4 CV of Buffer A, 2 CV of Buffer A containing 500 mM NaCl, then an additional 2 CV of Buffer A. CaM was eluted with Buffer C (50 mM Tris-HCl pH 7.5, 1 mM EDTA). The phenyl sepharose elution was concentrated to approximately 0.5 mL using 5,000 MWCO Amicon Ultra concentrators (Millipore), then loaded onto a SDX-75 column (General Electric) preequilibrated with 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM CaCl₂. Pure CaM eluted at approximately 0.53 CV and was collected and stored at 4 °C. Wild type (WT) CaM and the variants were purified to > 95% purity as determined by SDS-PAGE. In addition, the molecular weight for all proteins was verified through mass spectrometry.

CaM concentrations were calculated one of two ways. The first method measures the absorbance of tyrosine in denatured CaM. A 1/10 dilution of protein into 8 M GnHCl was incubated for 10 minutes before the absorbance at 278 nm was taken. The extinction coefficient at 278 nm for WT and variants containing one tyrosine was 1400 M⁻¹cm⁻¹ and 2800 M⁻¹cm⁻¹ for variants containing two tyrosines.⁵⁵ The second method was used to

calculate the concentration of protein under native conditions. Typically 1/3 dilutions of CaM were made into 50 mM Tris-HCl pH 7.5, 1 mM CaCl₂ and the A₂₇₆ was measured. An extinction coefficient of 1576 M⁻¹cm⁻¹ was used for proteins, including WT, that contain only one tyrosine.⁵⁶ The concentrations from both methods are typically within 5% of each other.

CaM-binding peptides. Unmodified peptides corresponding to the peptides bound to CaM in the crystal structures 1CDL (smMLCK_p) and 1MXE (CaMKI_p) were purchased from PeptidoGenic Research and Co., Inc. or from Sigma Genosys at greater than 95% purity. Biotinylated smMLCK and CaMKI peptides were purchased from AnaSpec, Inc. at greater than 95% purity. The N-terminal biotin moiety was separated from the peptide sequence by two 6-aminohexanoic acid linkers (LC) and a Gly-Gly-Ser-Gly-Gly peptide linker. The sequences for all peptides used in this study can be found in Table 5-1.

The molecular weight of each peptide was verified through mass spectrometry, and amino acid analysis was performed on unmodified smMLCK and CaMKI peptides (Jinny Johnson, Texas A&M University) to confirm the composition. The extinction coefficient A₂₈₀ = 5690 M⁻¹cm⁻¹ was used to determine the concentration of all peptides under denaturing conditions.⁵⁵

Fluorescence binding assays. Fluorescence assays were conducted at 25 °C on a QuantaMaster UV-Vis fluorimeter (Photon Technology International) essentially as described.¹⁶ Samples (1 mL) were prepared with varying amounts of CaM (0 to 2.0 μM)

and 1 μM peptide (unless otherwise noted) in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM CaCl_2 . Samples were allowed to equilibrate a minimum of 1 hour before measurements were taken. Each sample was excited at 295 nm and the emission spectrum was recorded between 310 nm and 460 nm. A 1 nm step size and 1 second averaging time were used in all assays. The slit width on the excitation source and emission detectors were adjusted to give a maximum signal near the limit of the detector, typically between 0.9 and 1.1 nm.

The fluorescence at 318 nm (unless otherwise noted) as a function of the CaM concentration was plotted in Kaleidograph (Synergy Software). Data were fit to a 1:1 binding model to obtain the dissociation binding constant (K_D) (Equation A-2) where ϵ_{free} (the extinction coefficient for free peptide), ϵ_{bound} (the extinction coefficient for bound peptide), and K_D are parameters that are fit during optimization, and $[P]$ and $[CaM]$ are the concentration of peptide and CaM, respectively. In addition, a CaM concentration factor was fit during the curve fitting procedure to correct for inaccuracies in the determination of CaM concentrations. This fit value typically ranged between 0.9 and 1.1, indicating that our CaM concentrations were off by 10% or less.

$$\text{Fluorescence} = \epsilon_{\text{free}} * [P] + (\epsilon_{\text{bound}} - \epsilon_{\text{free}}) * 0.5 * (K_D + [P] + [CaM]) \pm \sqrt{(K_D + [P] + [CaM])^2 - 4 * [CaM] * [P]} \quad (\text{A-2})$$

Surface Plasmon Resonance (SPR) assays. SPR (Biacore) experiments were performed on a T100 instrument (Biacore). Approximately 3000-5000 response units of streptavidin were immobilized to all four flow cells of a CM5 chip through standard amine coupling. Biotinylated peptide was captured by the streptavidin surface on flow cells 2-4. Flow cell 1 was reserved as a control. All assays were conducted in HBS-P

buffer (Biacore) with 1 mM CaCl₂ (10 mM HEPES pH 7.4, 150 mM NaCl, 0.0005% v/v Surfactant P20, 1 mM CaCl₂). The chip was regenerated with two 30-second injections of 10 mM EGTA unless otherwise noted.

Kinetics experiments for wild-type CaM were conducted on a chip with between 5 and 15 RUs of immobilized peptide. Samples containing serial dilutions between 2 μ M and 6.3 pM WT CaM were injected over the surface of the chip for 180 seconds at 100 μ L/min then allowed to dissociate for 300 seconds before the surface was regenerated. One sample in the series was repeated to confirm that the immobilized surface was unchanged during the experiment. Sensorgrams were analyzed using the BiaEvaluation software package (Biacore). Various fitting models were used to try to fit the data accurately, including a 1:1 model, a multivalent model, and a mass-transport limited model.

Equilibrium experiments were performed with approximately 20 RUs of peptide immobilized on the streptavidin surface. Samples used to create binding curves were either from a 2X dilution series containing 20 nM to 9.8 pM CaM in running buffer or a 24-point semi-log dilution series with the highest concentration of 50 μ M. The flow rate and contact time for each injection was varied in order to reach equilibrium binding. The flow rates varied from 20 to 8 μ L/min and the contact times ranged between 420 and 2500 sec. All concentration series were run from low to high concentration without regeneration in order to expedite the equilibrium. After a complete concentration series, the chip was regenerated with 5 pulses of 10 mM EGTA. In addition, a 2.5 nM WT CaM reference sample was run after every two proteins in order to verify the stability of the surface. The data were analyzed by plotting the equilibrium RU value as a function of the

concentration of CaM injected. Data are shown as unfit curves and can only be ranked to determine the relative affinities of the variants due to poor fits to a 1:1 model.

Approximately 120 RUs of biotinylated smMLCK peptide was immobilized to a streptavidin CM5 chip for the competition assays. Samples containing 100 nM CaM (WT or variant) and either no peptide or between 250 nM and 3.3 nM non-biotinylated peptide were injected onto the chip at 50 $\mu\text{L}/\text{min}$ with a 30 second association time and no dissociation time. Two samples were repeated during the run to ensure that the streptavidin/peptide surface was not being damaged or degraded. The surface was regenerated after each injection. Data was exported from BiaEvaluation software and the initial rate of signal increase for each sensorgram was determined by fitting the initial seconds of injection to a linear regression in Excel (Microsoft). The dissociation binding constant (K_D) was determined by plotting the initial slopes as a function of the log of the concentration of competitive (non-biotinylated) peptide. Kaleidograph was used to solve for K_D by fitting the data to Equation A-3 where $[A_0]$ is the CaM concentration, R_0 is the initial rate with no competing receptor present, X is the $\log[L_0]$, where $[L_0]$ is the concentration of competitive peptide, and R is the initial rate at $[L_0]$.⁵⁷

$$R = \frac{R_0}{2[A_0]} \left([A_0] - 10^x - K_D \right) + \sqrt{(K_D)^2 + 2(10^x)(K_D) + (10^x)^2 + 2[A_0]K_D - 2[A_0]10^x + [A_0]^2} \quad (\text{A-3})$$

Results

ORBIT calculations. We chose the structures for the smMLCK-CaM complex (pdb ID 1CDL)⁴⁹ and the CaMKI-CaM complex (pdb ID 1MXE)⁵⁰ to incorporate explicit negative design into our computational protocol. These complexes were chosen because both have high-resolution structures available, both have very high CaM-peptide binding

affinities, and smMLCK_p and CaMKI_p both contain a tryptophan residue. As CaM contains no Trps, the change in fluorescence of the peptide Trp can be monitored to determine binding.¹⁶ In addition, these two structures are very similar: their RMSD is only 1.4 Å and the peptide is bound in a very similar orientation. Previous work in the Mayo lab indicated that CaM can be engineered to have altered specificity without including explicit negative design.¹⁶ We hoped to expand on this work and create a CaM variant with increased specificity for the smMLCK peptide relative to the CaMKI peptide.

The ORBIT protein design software suite was used to both optimize the peptide-protein interface in the two CaM-peptide input structures as well as predict sequences that would stabilize the smMLCK_p-bound structure (1CDL) and destabilize the CaMKI_p structure (1MXE). The resulting sequences are shown in Table A-2. The sequences are split into two categories depending on the rotamer library used, as described in the methods.

The positive designs for the smMLCK_p-CaM structure (smMLCK-conf and smMLCK-e2) contain four and five mutations, respectively, from wild-type CaM. V55I, V91I, and V108I are found in both designs. These are very conservative mutations with the extra methyl group predicted to fill small spaces in the protein-peptide interface. Positions 18 and 51 were also mutated in one or both of these sequences; these mutations will be discussed in the negative design results as they are also found in the negative design sequences.

The positive designs for the CaMKI_p-CaM structure (CaMKI-conf and CaMKI-e2) contain five and three mutations, respectively. Many of the mutations are valine to

phenylalanine or tyrosine. The CaMKI_p-CaM structure is not as packed at the protein-peptide interface as the smMLCK_p-CaM structure, so there is additional room for larger side chains to pack into the interface at positions 55, 91, and 108. Position 84 is also mutated from a Glu to an Ile or Leu in these designs. In the smMLCK_p-CaM structure, Glu 84 makes electrostatic interactions with two arginine residues in the peptide. However, in the CaMKI peptide these arginines have been replaced by an alanine and a histidine that is not within hydrogen bonding distance. Therefore, in the CaMKI_p-CaM optimization, the electrostatics interactions with E84 are less stabilizing than the van der Waals interactions between the hydrophobic Ile or Leu and the peptide.

Because the goal of the negative designs was to create a CaM variant that binds smMLCK_p with high affinity and CaMKI_p with reduced affinity, it was no surprise that the sequences resulting from the negative design protocol are very similar to the sequences for the smMLCK_p-CaM positive designs. There are, however, two notable exceptions that may convey specificity to the negative design variants. Those mutations are L18I, found in both negative design solutions as well as in smMLCK-e2, and M51Y, which was selected in the negative design calculation from the e2 rotamer library and the positive design sequence smMLCK-conf. Both L18I and M51Y are predicted to make van der Waals clashes with the CaMKI peptide, but can be accommodated in the smMLCK_p structure (Figure A-2). In the case of L18I, the clash between residue 18 and Gln 305 of the peptide cannot be resolved through alternate conformations due to the constraints on the peptide side chain. In the smMLCK_p-CaM structure, however, the lysine residue at the equivalent position in the peptide is predicted to sample a different conformation that allows space for the isoleucine. In addition, the isoleucine rotamer

chosen in the smMLCK structure is similar to the crystallographic orientation of the wild-type leucine. M51Y is located in the protein-peptide interface near the C-terminus of the peptide (Figures A-2C and A-2D). In the smMLCK_p design, the tyrosine side chain extends toward the solvent and is not sterically occluded by the serine residue at the end of the peptide. The tyrosine in the CaMKI_p structure is predicted to be blocked from extending toward the solvent by the long arginine side chain near the end of the peptide and therefore must take an alternate conformation in which it sterically clashes with Met 316 of the CaMKI peptide. These computational models suggest that both L18I and M51Y might be good candidates for incorporating specificity into CaM-peptide binding. We therefore decided to construct variants containing either single or double mutations to investigate the effects of these positions: L18I, L18I/M51F, L18I/M51Y, M51F, and M51Y.

Native purification. Native purification of CaM has been reported using hydrophobic interaction chromatography.^{54,58-60} In the presence of Ca²⁺, a large hydrophobic surface on CaM is exposed and interacts with the column. After non-hydrophobic proteins are washed off the column, CaM is eluted with a buffer containing EDTA. The EDTA chelates the Ca²⁺, causing the CaM to undergo a conformational change that buries the hydrophobic patch. This method has numerous advantages over the non-native purification method previously used in the lab.¹⁶ First, the protein does not need to undergo refolding and is exposed only to gentle, native conditions, resulting in protein preparations that are less likely to be damaged. Second, only the CaM that undergoes the necessary Ca²⁺-dependent conformation change is purified; damaged

proteins will either be lost in the washes or remain bound to the column after elution. Finally, only proteins with a Ca^{2+} dependency will elute with the buffer containing EDTA, so the protein is greater than 95% pure after a single purification step. We performed an additional gel filtration step to exchange the buffer and ensure that the protein was monodispersive and monomeric.

SDS-PAGE on the purified protein showed that it was more than 98% pure. A double band was often seen when excess EDTA or Ca^{2+} was not added before denaturing the samples. The double band was shown to correspond to Ca^{2+} -bound and Ca^{2+} -free versions of the protein,⁶¹ which has been verified in our laboratory (data not shown).

Tryptophan fluorescence assays. Tryptophan (Trp) fluorescence is a convenient, commonly used technique to determine the affinity of CaM-peptide interactions.^{16,17,62-64} The analysis is simplified because CaM contains no Trps. Fluorescence from the single Trp in the peptide can be monitored to indicate movement from solvent-exposure to a hydrophobic environment, giving a direct measurement of peptide binding to CaM.⁶⁵ After excitation at 295 nm, the emission spectrum was monitored for samples containing a fixed amount of Trp-containing peptide and varying amounts of CaM (Figure A-3). Water emits a significant amount of fluorescence around 330 nm. When the background was subtracted, the emission spectra clearly showed the expected blue-shift and increase in intensity as CaM was titrated in, which is characteristic of the burial of the Trp (Figure A-3B). To eliminate error due to sampling around the water peak, we determined that monitoring emission at 318 nm gave results similar to those at 326 nm, but with smaller error associated with the dissociation constant (data not shown). We therefore plotted

fluorescence at 318 nm as a function of the CaM/peptide ratio in order to calculate K_D (Figure A-3C).

Although the fluorescence-based assay is simple and straightforward, the low intensity of Trp fluorescence emission makes it impossible to use low concentrations of peptide. This is a problem because the K_D s we are analyzing are around 1 nM and our peptide concentration is 1 μ M. This three orders-of-magnitude difference between the expected K_D and the peptide concentration creates a great deal of error in the K_D extrapolated from the data. We experimented with decreasing the concentration of peptide and found that 0.3 μ M peptide gave reasonable signal to noise, but that 0.1 μ M peptide gave poor data (data not shown). However, 0.3 μ M peptide did not get us significantly closer to the “10-fold above the K_D to 10-fold below the K_D ” general rule for concentrations in binding assays, so we continued the experiments with 1 μ M peptide to maximize the signal to noise.

CaM variants-peptide binding: Fluorescence assays. Eight computationally designed variants and wild-type CaM were assayed for binding to both smMLCK_p and CaMKI_p as described in the methods. The results are shown in Table A-3. We successfully reproduced the results from Shifman *et al.* for wild-type binding to both smMLCK_p and CaMKI_p.¹⁶ The K_D resulting from this experiment was two-fold weaker than that reported by Shifman *et al.*, but well within the limitations of the assay. Unfortunately, most of the variants did not affect the specificity of binding as had been anticipated from the designs. Four out of the five variants representing the negative design calculations (L18I, M51F, L18I/M51F, L18I/M51Y) did not show any specificity

for smMLCK_p over CaMKI_p. In addition, the positive designs on the smMLCK_p-CaM structure did not alter the specificity significantly. These assays did result in some interesting variants, however. First, the M51Y single mutant from the negative design calculations showed a 10-fold preference for the smMLCK peptide over the CaMKI peptide. This preference seems to result from an increase in affinity to smMLCK_p as opposed to a decrease in affinity to CaMKI_p, contrary to what was expected from the negative design calculations. The variant corresponding to the positive design calculation performed on the CaMKI_p-CaM structure using the e2 library (CaMKI-e2) also showed an approximately 10-fold preference for binding smMLCK_p. This result was unexpected, as the calculation should have predicted mutations that stabilize the CaMKI_p-CaM interface. Finally, CaMKI-conf appeared to show a moderate specificity switch from a two-fold preference for smMLCK_p for wild-type to a two-fold preference for CaMKI_p. Although these results were far from definitive, we decided to investigate the results from the two CaMKI_p-CaM structure positive designs further.

Five single mutation variants were created to investigate their effects on peptide binding. We created a variant with each mutation found in either CaMKI-conf or CaMKI-e2 that could have an affect on activity. We chose not to make L39I because it was also seen in the negative design result; similarly, V108I was not made because it was seen in both smMLCK_p-CaM designs and both negative design results and had no effect in those proteins. We therefore made V55F, E84I, V91F, V91Y, and V108F constructs and tested them in the Trp fluorescence assay. V55F was chosen during the optimization for both CaMKI_p-CaM positive designs. E84I, V91Y, and V108F were found only in the CaMKI-conf positive design, which has higher affinity for CaMKI_p than smMLCK_p. The final

mutation, V91F, is derived from the CaMKI-e2 positive design sequence that has a 10-fold preference for smMLCK_p. The results from the fluorescence assay show that V55F, E84I, and V108F mutations may create a slight preference for smMLCK_p (Table A-4). These results, however, are inconsistent with their presence in the CaMKI-conf design, which shows a switch toward better binding of CaMKI_p. The V55F mutation may provide some of the increased specificity in CaMKI-e2 binding to smMLCK_p. Mutations at position 91 had no effect on the binding to either smMLCK_p or CaMKI_p, indicating that this mutation is not the cause of the differences seen in the positive designs.

Because these results seemed counter in many ways to our computational designs and had a great deal of error associated with them due to the large difference between the K_{DS} of interest and the concentrations of peptide used, we decided to further study these variants using surface plasmon resonance (SPR, Biacore).

Biacore experiments to determine binding constants. Biacore has been used previously to provide both kinetics and steady-state dissociation constants for CaM binding to various peptides.⁶⁶⁻⁶⁸ We chose to immobilize our two biotinylated peptides through streptavidin interaction. This allowed us to analyze multiple CaM variants at a time without making new chips and oriented the peptides in a consistent way to minimize the error associated with random orientation by amine coupling. Initial experiments indicated that the kinetics for the wild-type CaM interaction with smMLCK_p and CaMKI_p could not be fit to a 1:1 binding model (Figure A-4A). A better fit was obtained using a multivalent model (Figure A-4B). However, these complexes have been shown previously to be 1:1 interactions, and fitting to a multivalent model is not

appropriate.^{16,49,50} We therefore decided to obtain dissociation constants from equilibrium experiments instead of analyzing the kinetics of the binding.

Equilibrium Biacore experiments measure the maximal response for a given analyte concentration. No kinetics information is obtained, but when the equilibrium binding response is plotted as a function of the concentration, a sigmoidal curve fit can give a K_D of binding (Figure A-5). When wild-type data was plotted, however, the shape of the curves was not as expected (Figure A-5B). At lower concentrations, the system acted as predicted, but at high concentrations, the response did not level out to reach a maximal value. Even at 50 μ M CaM, maximal binding was not observed. This indicates that the system is not an ideal 1:1 binding system and instead shows some self-binding or aggregation at high concentrations.

Although the system did not act in an ideal fashion, we proceeded to evaluate the designed variants via equilibrium binding studies. None of the data fit well to the 1:1 model, and therefore K_{DS} could not be determined. Instead, we ranked the variants relative to WT CaM (Figure A-6). The two smMLCK_p-CaM structure positive design variants (smMLCK-conf and smMLCK-e2) bound to both smMLCK_p and CaMKI_p with approximately the same affinity as WT. The mutations in these two variants did not affect the binding to either of these peptides. In addition, CaMKI-conf and CaM8, which had previously shown improved specificity to other peptides, bound significantly worse to both peptides as compared to WT. One or more of the mutations in these variants significantly affected either the stability of the protein or the binding in a negative way that was not peptide specific. One of the variants, CaMKI-e2, showed some specificity between the peptides. This variant binds significantly worse than WT to smMLCK_p but

binds similar to WT to CaMKI_p. This result is consistent with the purpose of the design: to stabilize the CaMKI-CaM structure. Many of the single point mutants derived from CaMKI-conf and CaMKI-e2 (see *CaM variants-peptide binding: Fluorescence assays*) also bound with higher relative affinity to CaMKI_p than smMLCK_p (Figures A-6C, A-6D). The V91F, V91Y, and V108F mutations appear to convey some specificity toward CaMKI_p. The E84I mutation interferes with binding to both peptides, whereas V55F affords some specificity toward smMLCK_p.

Finally, because of the difficulty in getting K_D values from the equilibrium binding experiments, we also attempted competition Biacore assays. These assays were a proof-of-concept experiment and were not completed in full. Instead, only the binding between variants and smMLCK_p was investigated and compared to the fluorescence assays and the equilibrium Biacore assays. In the competition Biacore experiments, each sample contains a constant amount of CaM or CaM variant and a variable amount of competing peptide, in this case non-biotinylated smMLCK_p. A set of sensorgrams for each variant was collected and the initial rate of binding was plotted as a function of the concentration of inhibiting peptide in the sample (Figure A-7).

The competition assays produced reasonable data with small error. The rankings generally agree well with the rankings for the variants binding to smMLCK_p in the equilibrium experiment (compare Figure A-6E and Figure A-7C). The competition experiment shows that, as expected, both positive designs modeled after the smMLCK_p-CaM structure have high affinity for smMLCK_p. In contrast to previous experiments and the expectation of the design, CaMKI-e2 binds smMLCK_p as well as WT. These results also confirm the equilibrium experiment conclusion that CaMKI-conf and CaM8 both

bind with significantly lower affinity than WT to smMLCK_p. The single variant results indicate that V108F is an unfavorable mutation for binding to smMLCK_p, and mutation at position 91 slightly disrupts the interaction.

This proof-of-concept Biacore experiment could be expanded to test the variants against CaMKI to determine whether any specificity changes have occurred and also to test the negative design variants: L18I, M51F, M51Y, L18I/M51F, and L18I/M51Y.

Conclusions

Assaying CaM-peptide complexes. It was shown during the course of this work that assaying CaM-peptide complexes by tryptophan fluorescence is not ideal. The concentration of peptide that must be used to obtain a good signal to noise is too high to evaluate affinities near 1 nM. Instead, we investigated assays using SPR technology (Biacore). The CaM-smMLCK peptide complex does not exhibit 1:1 binding as expected in the kinetics experiment, so we were forced to study the equilibrium binding to obtain relative affinities between the variants. Even with the equilibrium studies, the CaM complexes did not reach saturation even at 50 μ M CaM, indicating that aggregation or self-binding was occurring at high concentrations. We were, however, able to rank the variants tested for their affinities to both the smMLCK and CaMKI peptides to test our computational designs. Additionally, we provided proof-of-concept work that a competition-based assay using the Biacore can provide more accurate data including K_{DS} for the interactions. Future work for CaM-peptide interactions of high affinity should be investigated using the competition-based method.

Evaluating computational designs. Because of the difficulty in assaying CaM-peptide affinities, it was difficult to accurately assess our computational designs. We were, however, able to predict variants that would destabilize binding to smMLCK_p by using computational design to predict mutations that would stabilize a competing structure, including V108F, which on its own as well as in CaMKI-conf acts to decrease the affinity for smMLCK_p. Also, one of our negative design variants, M51Y, shows a 10-fold preference for smMLCK_p over CaMKI_p. Further work to characterize the negative and positive design variants may provide more insight into the computational designs. Additional work is also required to optimize the negative design scoring function, cutoff parameters, and energy thresholds to obtain the best negative design results.

CaM-peptide complexes as a model system for specificity. In addition to the difficulty in determining the binding affinities for CaM-peptide complexes, it is unclear whether the CaM system is ideal for investigating protein-protein specificity. While CaM binds many targets with high affinity, making it a seemingly good candidate, this broad promiscuity may incorporate too much flexibility into the protein-protein interface. A previous study from the Mayo lab showed that single mutations to glutamate residues near the interface can significantly affect the binding to smMLCK_p.¹⁷ It is likely, however, that these mutations not only affect binding but also protein stability in general. It would therefore be useful to probe the interface residues with a site-saturation mutagenesis library to determine which residues are consistent with binding at each position. This type of experiment would lead to a great deal of data that could be used to inform the computational designs and perhaps greatly increase our understanding of

protein-protein specificity. In addition, by incorporating mutations that destroy the binding, we can gain insight into how much mutation is tolerable to the system. Ideally, a model system for specificity would have some flexibility but also be very sensitive to mutation in order to affect binding significantly. Calmodulin, because it binds so many non-homologous native substrates, may have too much flexibility in the binding to be a good model.

Although perhaps not ideal for protein-protein specificity, this system can be very valuable for studying protein-protein interactions. Recently, Roger Tsien in collaboration with David Baker designed a specific peptide-CaM interface using computational design.⁶ In this case, the peptide is a designed variation of smMLCKp that incorporates bulky groups that clash with wild-type CaM, employing a “knobs-and-holes” approach. CaM was then designed around this new peptide, giving a specific interaction. The CaM-peptide system seems ideal for creating new interactions between protein and peptide, taking advantage of the inherently high specificity and large flexibility in the system. Future work to design other non-native peptide interactions with CaM may prove to be quite successful.

Acknowledgements

I want to thank Jost Vielmetter in the Protein Expression Center at Caltech for his invaluable help and advice with the Biacore experiments. I would also like to acknowledge Ben Allen, Heidi Privett, and the rest of the Mayo Lab, without whom the setup and execution of the computational designs would have been much more difficult.

References

1. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. & Baker, D. (2005). Progress in modeling of protein structures and interactions. *Science* **310**, 638-42.
2. Gomez, S. M., Choi, K. & Wu, Y. (2008). Prediction of protein-protein interaction networks. *Curr Protoc Bioinformatics* **Chapter 8**, Unit 8.2.
3. Fernandez-Ballester, G. & Serrano, L. (2006). Prediction of protein-protein interaction based on structure. *Methods Mol Biol* **340**, 207-34.
4. Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J. & Stoddard, B. L. (2002). Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell* **10**, 895-905.
5. Fajardo-Sanchez, E., Stricher, F., Paques, F., Isalan, M. & Serrano, L. (2008). Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. *Nucleic Acids Res* **36**, 2163-73.
6. Palmer, A. E., Giacomello, M., Kortemme, T., Hires, S. A., Lev-Ram, V., Baker, D. & Tsien, R. Y. (2006). Ca²⁺ indicators based on computationally redesigned calmodulin-peptide pairs. *Chem Biol* **13**, 521-30.
7. Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S., Serrano, L. & Gonzalez, C. (2002). Computer-aided design of a PDZ domain to recognize new target sequences. *Nat Struct Biol* **9**, 621-7.
8. Huang, J., Koide, A., Makabe, K. & Koide, S. (2008). Design of protein function leaps by directed domain interface evolution. *Proc Natl Acad Sci U S A* **105**, 6578-83.
9. Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., Stacker, S. A., Baechi, T., Moelling, K. & Hovens, C. M. (1999). Mutagenesis and selection of PDZ domains that bind new protein targets. *Nat Biotechnol* **17**, 170-5.
10. Sidhu, S. S. & Koide, S. (2007). Phage display for engineering and analyzing protein interaction interfaces. *Curr Opin Struct Biol* **17**, 481-7.
11. Potapov, V., Reichmann, D., Abramovich, R., Filchtinski, D., Zohar, N., Ben Halevy, D., Edelman, M., Sobolev, V. & Schreiber, G. (2008). Computational redesign of a protein-protein interface for high affinity and binding specificity using modular architecture and naturally occurring template fragments. *J Mol Biol* **384**, 109-19.
12. Joachimiak, L. A., Kortemme, T., Stoddard, B. L. & Baker, D. (2006). Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J Mol Biol* **361**, 195-208.

13. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* **11**, 371-9.
14. Reynolds, K. A., Hanes, M. S., Thomson, J. M., Antczak, A. J., Berger, J. M., Bonomo, R. A., Kirsch, J. F. & Handel, T. M. (2008). Computational redesign of the SHV-1 beta-lactamase/beta-lactamase inhibitor protein interface. *J Mol Biol* **382**, 1265-75.
15. Huang, P. S., Love, J. J. & Mayo, S. L. (2007). A de novo designed protein protein interface. *Protein Sci* **16**, 2770-4.
16. Shifman, J. M. & Mayo, S. L. (2002). Modulating calmodulin binding specificity through computational protein design. *J Mol Biol* **323**, 417-23.
17. Shifman, J. M. & Mayo, S. L. (2003). Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc Natl Acad Sci U S A* **100**, 13274-9.
18. DeGrado, W. F., Prendergast, F. G., Wolfe, H. R., Jr. & Cox, J. A. (1985). The design, synthesis, and characterization of tight-binding inhibitors of calmodulin. *J Cell Biochem* **29**, 83-93.
19. Green, D. F., Dennis, A. T., Fam, P. S., Tidor, B. & Jasanoff, A. (2006). Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide. *Biochemistry* **45**, 12547-59.
20. Sammond, D. W., Eletr, Z. M., Purbeck, C., Kimple, R. J., Siderovski, D. P. & Kuhlman, B. (2007). Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. *J Mol Biol* **371**, 1392-404.
21. Sood, V. D. & Baker, D. (2006). Recapitulation and design of protein binding peptide structures and sequences. *J Mol Biol* **357**, 917-27.
22. Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nat Struct Biol* **10**, 45-52.
23. Bolon, D. N., Grant, R. A., Baker, T. A. & Sauer, R. T. (2005). Specificity versus stability in computational protein design. *Proc Natl Acad Sci U S A* **102**, 12724-9.
24. Dahiyat, B. I. & Mayo, S. L. (1997). De novo protein design: fully automated sequence selection. *Science* **278**, 82-7.
25. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol* **5**, 470-5.
26. Marshall, S. A. & Mayo, S. L. (2001). Achieving stability and conformational specificity in designed proteins via binary patterning. *J Mol Biol* **305**, 619-31.

27. Datta, D., Wang, P., Carrico, I. S., Mayo, S. L. & Tirrell, D. A. (2002). A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. *J Am Chem Soc* **124**, 5652-3.
28. Bolon, D. N. & Mayo, S. L. (2001). Enzyme-like proteins by computational design. *Proc Natl Acad Sci U S A* **98**, 14274-9.
29. Lassila, J. K., Keeffe, J. R., Oelschlaeger, P. & Mayo, S. L. (2005). Computationally designed variants of Escherichia coli chorismate mutase show altered catalytic activity. *Protein Eng Des Sel* **18**, 161-3.
30. Dunbrack, R. L., Jr. & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* **6**, 1661-81.
31. Lassila, J. K., Privett, H. K., Allen, B. D. & Mayo, S. L. (2006). Combinatorial methods for small-molecule placement in computational enzyme design. *Proc Natl Acad Sci U S A* **103**, 16710-5.
32. Mayo, S. L., Olafson, B. D. & Goddard, W. A. (1990). Dreiding - a generic force-field for molecular simulations. *J Phys Chem* **94**, 8897-8909.
33. Dahiyat, B. I. & Mayo, S. L. (1997). Probing the role of packing specificity in protein design. *Proc Natl Acad Sci U S A* **94**, 10172-7.
34. Street, A. G. & Mayo, S. L. (1998). Pairwise calculation of protein solvent-accessible surface areas. *Fold Des* **3**, 253-8.
35. Gordon, D. B., Hom, G. K., Mayo, S. L. & Pierce, N. A. (2003). Exact rotamer optimization for protein design. *J Comput Chem* **24**, 232-43.
36. Gordon, D. B. & Mayo, S. L. (1998). Radical performance enhancements for combinatorial optimization algorithms based on the dead-end elimination theorem. *J Comp Chem* **19**, 1505-14.
37. Pierce, N. A., Spriet, J. A., Desmet, J. & Mayo, S. L. (2000). Conformational splitting: A more powerful criterion for dead-end elimination. *J Comp Chem* **21**, 999-1009.
38. Allen, B. D. & Mayo, S. L. (2006). Dramatic performance enhancements for the FASTER optimization algorithm. *J Comput Chem* **27**, 1071-5.
39. Desmet, J., Spriet, J. & Lasters, I. (2002). Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Proteins* **48**, 31-43.
40. Kirkpatrick, S., Gelatt, C. D., Jr. & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science* **220**, 671-80.

41. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equation of state calculations by fast computing machines. *J Chem Phys* **21**, 1087-92.
42. Stevens, F. C. (1983). Calmodulin: an introduction. *Can J Biochem Cell Biol* **61**, 906-10.
43. Jurado, L. A., Chockalingam, P. S. & Jarrett, H. W. (1999). Apocalmodulin. *Physiol Rev* **79**, 661-82.
44. Hoeflich, K. P. & Ikura, M. (2002). Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell* **108**, 739-42.
45. DeGrado, W. F. (1988). Design of peptides and proteins. *Adv Protein Chem* **39**, 51-124.
46. O'Neil, K. T. & DeGrado, W. F. (1990). How calmodulin binds its targets: sequence independent recognition of amphiphilic alpha-helices. *Trends Biochem Sci* **15**, 59-64.
47. Cox, J. A., Comte, M., Fitton, J. E. & DeGrado, W. F. (1985). The interaction of calmodulin with amphiphilic peptides. *J Biol Chem* **260**, 2527-34.
48. Kranz, J. K., Lee, E. K., Nairn, A. C. & Wand, A. J. (2002). A direct test of the reductionist approach to structural studies of calmodulin activity: relevance of peptide models of target proteins. *J Biol Chem* **277**, 16351-4.
49. Meador, W. E., Means, A. R. & Quijcho, F. A. (1992). Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex. *Science* **257**, 1251-5.
50. Clapperton, J. A., Martin, S. R., Smerdon, S. J., Gamblin, S. J. & Bayley, P. M. (2002). Structure of the complex of calmodulin with the target sequence of calmodulin-dependent protein kinase I: studies of the kinase activation mechanism. *Biochemistry* **41**, 14669-79.
51. Kurokawa, H., Osawa, M., Kurihara, H., Katayama, N., Tokumitsu, H., Swindells, M. B., Kainosho, M. & Ikura, M. (2001). Target-induced conformational adaptation of calmodulin revealed by the crystal structure of a complex with nematode Ca(2+)/calmodulin-dependent kinase kinase peptide. *J Mol Biol* **312**, 59-68.
52. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83.

53. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. & Galas, D. J. (1989). A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res* **17**, 6545-51.
54. Persechini, A., Blumenthal, D. K., Jarrett, H. W., Klee, C. B., Hardy, D. O. & Kretsinger, R. H. (1989). The effects of deletions in the central helix of calmodulin on enzyme activation and peptide binding. *J Biol Chem* **264**, 8052-8.
55. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319-26.
56. Maune, J. F., Klee, C. B. & Beckingham, K. (1992). Ca²⁺ binding and conformational change in two series of point mutations to the individual Ca(2+)-binding sites of calmodulin. *J Biol Chem* **267**, 5286-95.
57. Lazar, G. A., Dang, W., Karki, S., Vafa, O., Peng, J. S., Hyun, L., Chan, C., Chung, H. S., Eivazi, A., Yoder, S. C., Vielmetter, J., Carmichael, D. F., Hayes, R. J. & Dahiyat, B. I. (2006). Engineered antibody Fc variants with enhanced effector function. *Proc Natl Acad Sci U S A* **103**, 4005-10.
58. Gopalakrishna, R. & Anderson, W. B. (1982). Ca²⁺-induced hydrophobic site on calmodulin: application for purification of calmodulin by phenyl-Sepharose affinity chromatography. *Biochem Biophys Res Commun* **104**, 830-6.
59. Putkey, J. A., Slaughter, G. R. & Means, A. R. (1985). Bacterial expression and characterization of proteins derived from the chicken calmodulin cDNA and a calmodulin processed gene. *J Biol Chem* **260**, 4704-12.
60. Brockerhoff, S. E., Edmonds, C. G. & Davis, T. N. (1992). Structural analysis of wild-type and mutant yeast calmodulins by limited proteolysis and electrospray ionization mass spectrometry. *Protein Sci* **1**, 504-16.
61. Rhyner, J. A., Koller, M., Durussel-Gerber, I., Cox, J. A. & Strehler, E. E. (1992). Characterization of the human calmodulin-like protein expressed in *Escherichia coli*. *Biochemistry* **31**, 12826-32.
62. Findlay, W. A., Martin, S. R., Beckingham, K. & Bayley, P. M. (1995). Recovery of native structure by calcium binding site mutants of calmodulin upon binding of sk-MLCK target peptides. *Biochemistry* **34**, 2087-94.
63. Mirzoeva, S., Weigand, S., Lukas, T. J., Shuvalova, L., Anderson, W. F. & Watterson, D. M. (1999). Analysis of the functional coupling between calmodulin's calcium binding and peptide recognition properties. *Biochemistry* **38**, 3936-47.
64. Martin, S. R. & Bayley, P. M. (2002). Regulatory implications of a novel mode of interaction of calmodulin with a double IQ-motif target sequence from murine dilute myosin V. *Protein Sci* **11**, 2909-23.

65. O'Neil, K. T., Wolfe, H. R., Jr., Erickson-Viitanen, S. & DeGrado, W. F. (1987). Fluorescence properties of calmodulin-binding peptides reflect alpha-helical periodicity. *Science* **236**, 1454-6.
66. Montigiani, S., Neri, G., Neri, P. & Neri, D. (1996). Alanine substitutions in calmodulin-binding peptides result in unexpected affinity enhancement. *J Mol Biol* **258**, 6-13.
67. Sacks, D. B., Mazus, B. & Joyal, J. L. (1995). The activity of calmodulin is altered by phosphorylation: modulation of calmodulin function by the site of phosphate incorporation. *Biochem J* **312** (Pt 1), 197-204.
68. Takano, E., Hatanaka, M. & Maki, M. (1994). Real-time-analysis of the calcium-dependent interaction between calmodulin and a synthetic oligopeptide of calcineurin by a surface plasmon resonance biosensor. *FEBS Lett* **352**, 247-50.

Table 5-1. CaM-binding peptides.

Peptide	Sequence^a	MW (g/mol)	ϵ_{280} ($M^{-1}cm^{-1}$)^d
smMLCK	ARRKWQKTGHA VRAIGRLSS ^b	2278.6	5690
CaMKI	IKKNFAKSKWKQAFNATAVVRHMRK ^c	2988.6	5690
biotin-smMLCK	biotin-LC-LC-GGSGG-ARRKWQKTGHA VRAIGRLSS	3047.6	5690
biotin-CaMKI	biotin-LC-LC-GGSGG- IKKNFAKSKWKQAFNATAVVRHMRK	3757.6	5690

^a LC: 6-aminohexanoic acid linker (“long-chain”).

^b smMLCK peptide sequence corresponds to residues 796-815 of smMLCK.

^c CaMKI peptide sequence corresponds to residues 294-318 of CaMKI.

^d Extinction coefficients were theoretically calculated based on amino acid composition.⁵⁵

Table A-3. K_{DS} for CaM variants as determined by fluorescence assays.

Variant	K_D (nM)	
	smMLCK_p	CaMKI_p
WT¹⁶	1.8 ± 1.3	1.7 ± 0.7
WT	1.6 ± 0.9	3.5 ± 1.2
L18I	5.0 ± 3.1	6.2 ± 2.2
M51F	1.2 ± 0.9	1.9 ± 1.4
M51Y	0.4 ± 0.8	4.4 ± 1.5
L18I/M51F	5.8 ± 2.4	3.2 ± 1.3
L18I/M51Y	0.4 ± 0.7	0.9 ± 1.4
smMLCK-conf	1.7 ± 0.6	3.6 ± 0.8
smMLCK-e2	0.6 ± 0.8	0.2 ± 0.5
CaMKI-conf	4.4 ± 1.5	2.0 ± 1.1
CaMKI-e2	0.6 ± 0.6	7.0 ± 1.9

Table A-4. K_{DS} for CaMKI_p positive design-derived single mutant variants as determined by fluorescence assays.

Variant	K_D (nM)	
	smMLCK _p	CaMKI _p
V55F	1.4 ± 1.3	4.8 ± 1.3
E84I	0.4 ± 0.7	2.6 ± 1.7
V91F	2.1 ± 1.1	1.6 ± 0.9
V91Y	3.0 ± 1.1	2.3 ± 1.6
V108F	4.4 ± 1.9	7.6 ± 2.6

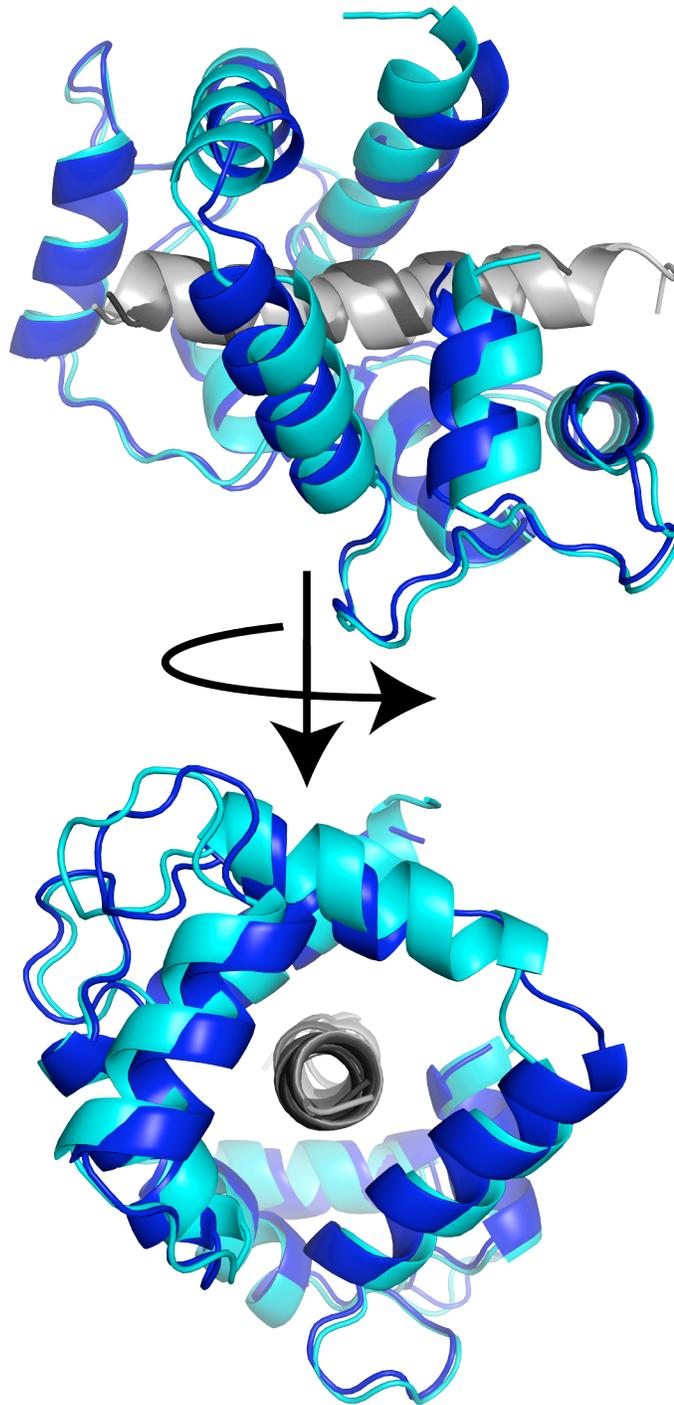


Figure A-1. CaM-peptide structures. The CaM-smMLCK_p structure is shown in dark blue (CaM) and dark gray (peptide), and the CaM-CaMKI_p structure is shown in cyan and light gray. The C α RMSD between the two structures is 1.4 Å. The peptide is bound between the N- and C-terminal domains of CaM such that the peptide is almost completely buried.

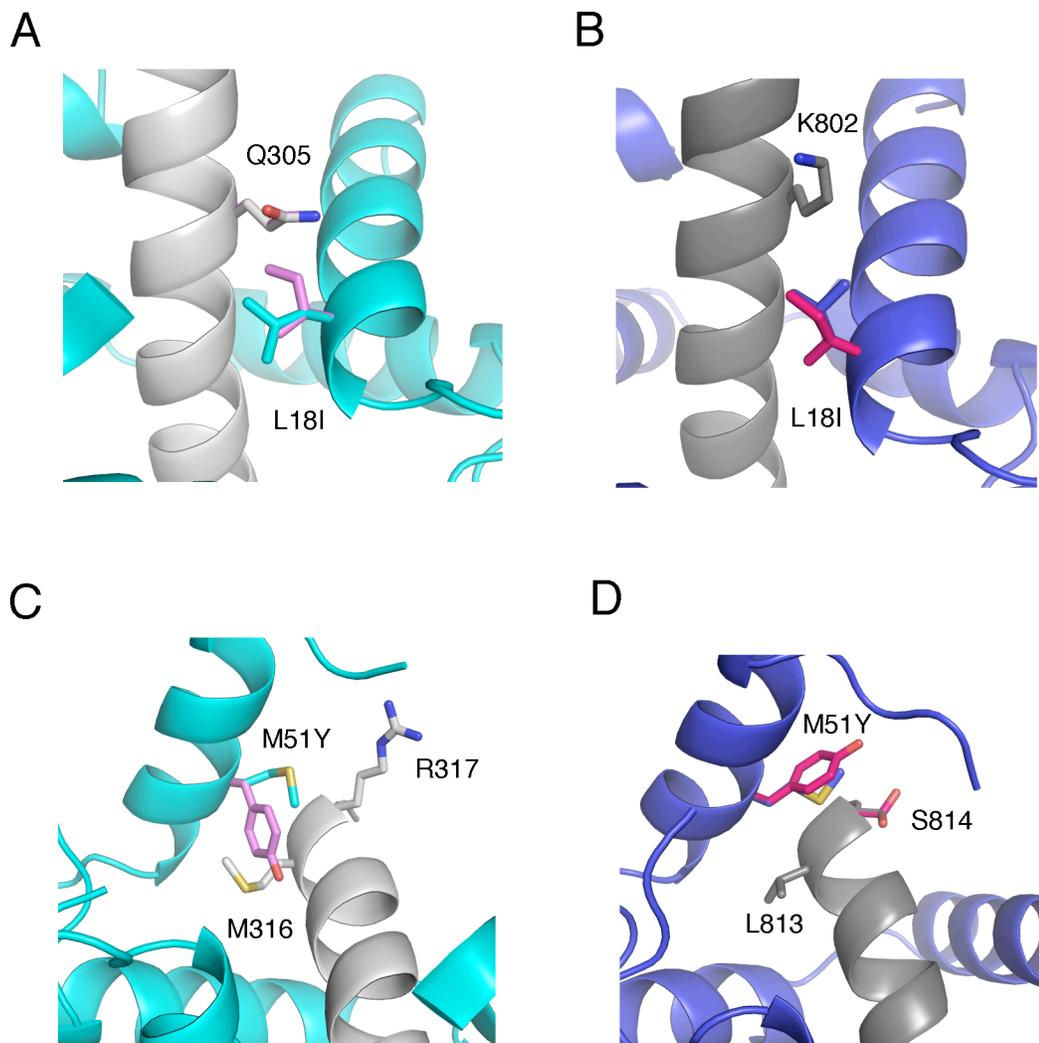


Figure A-2. Negative design result structures. The CaM-CaMKI_p structure (1MXE) is shown in light colors (panels A and C), and the CaM-smMLCK_p structure (1CDL) is shown in dark colors (panels B and D). The peptides are shown in gray, CaM is shown in blue, and the designed residues of interest are shown in magenta. (A) An Ile at position 18 cannot be accommodated in the CaMKI_p-bound structure due to Gln 305 in the peptide. This creates an unfavorable van der Waals interaction, which is expected to destabilize the complex. (B) In the smMLCK_p structure, however, the Lys in the equivalent position of smMLCK_p takes on a conformation that does not clash with position 18. (C) At position 51, a Tyr mutation in the negative design result from the e2 library clashes with M316 of CaMKI_p. The Tyr side chain is pointed inward toward the peptide because it is hindered from pointing toward the solvent by the long side chain of Arg 317. (D) The smaller Ser residue at position 814 (equivalent to R317) in the smMLCK peptide allows the Tyr at position 51 to point toward the solvent, eliminating the unfavorable interactions with the peptide that are seen in the CaMKI_p structure.

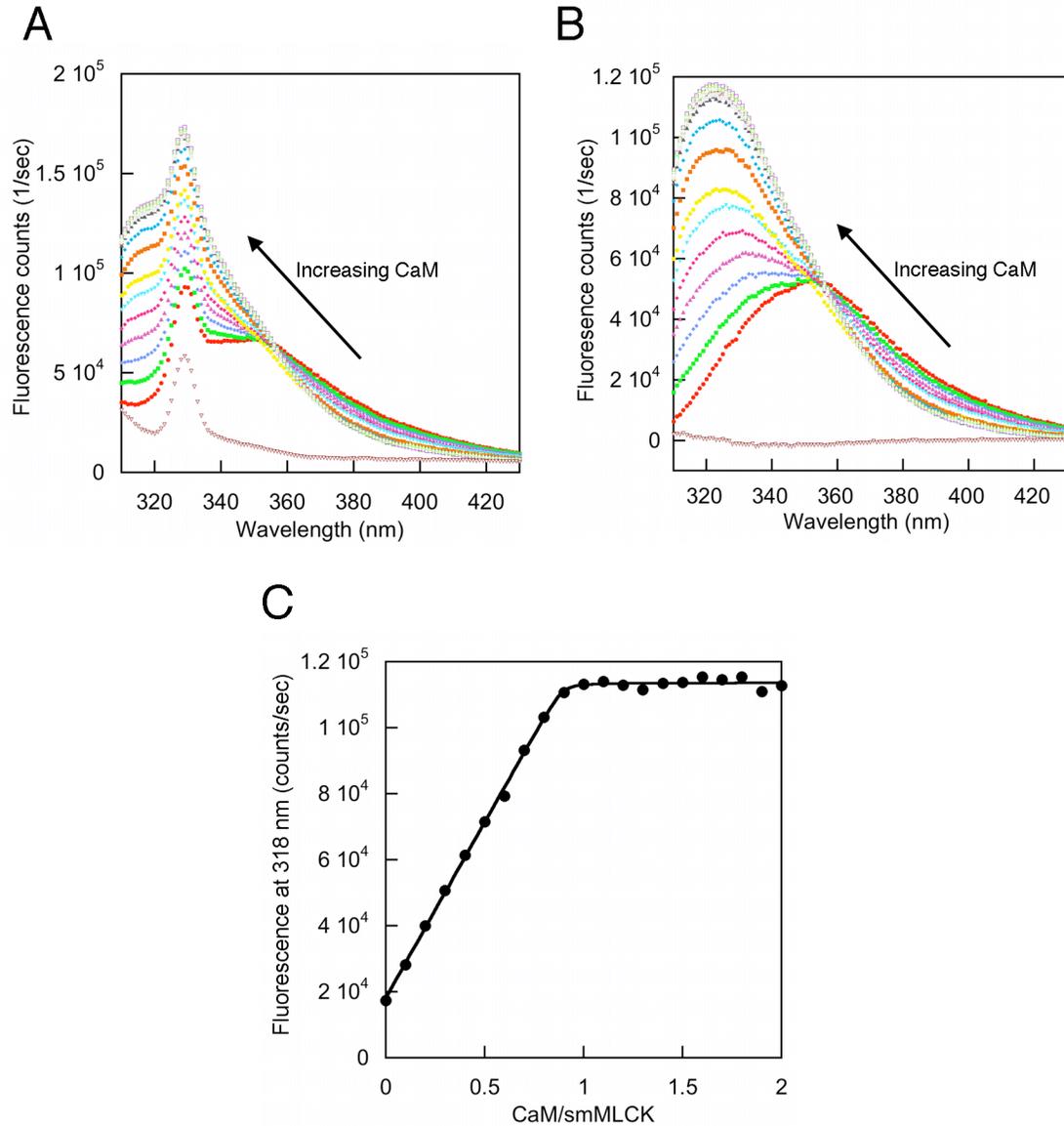


Figure A-3. Tryptophan fluorescence assay data analysis. (A) Samples containing 1 mM peptide and varying amounts of CaM are excited at 295 nm and the emission spectrum from the tryptophan fluorescence is measured. The brown curve at the bottom of both panels (A) and (B) is CaM alone with no peptide present. (B) The buffer is subtracted giving smooth curves that show a blue-shift and increase in intensity, which indicates burial of the tryptophan. (C) The fluorescence intensity at 318 nm is plotted as a function of the ratio of CaM to peptide. This titration curve shows that the binding between CaM and smMLCK_p is 1:1 and the data can be fit to Equation A-2, giving a K_D of interaction.

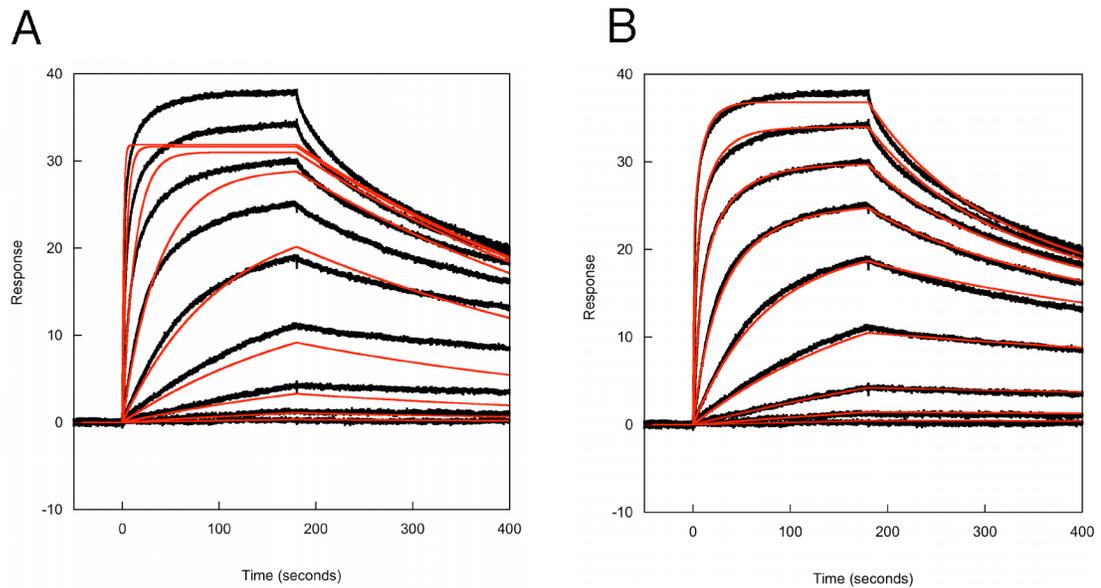


Figure A-4. Biacore kinetics of wild type CaM and smMLCK_p. Biacore data are shown in black and model curve fits are shown in red. Approximately 15 response units of biotinylated smMLCK_p was immobilized to streptavidin on a CM5 chip. Various concentrations of wild-type CaM were flowed over the peptide surface and binding was monitored. BiaEvaluation software was used to fit the data to a (A) 1:1 model or a (B) multivalent model. Neither model was a satisfactory fit due to previous work showing that the complex should be 1:1.

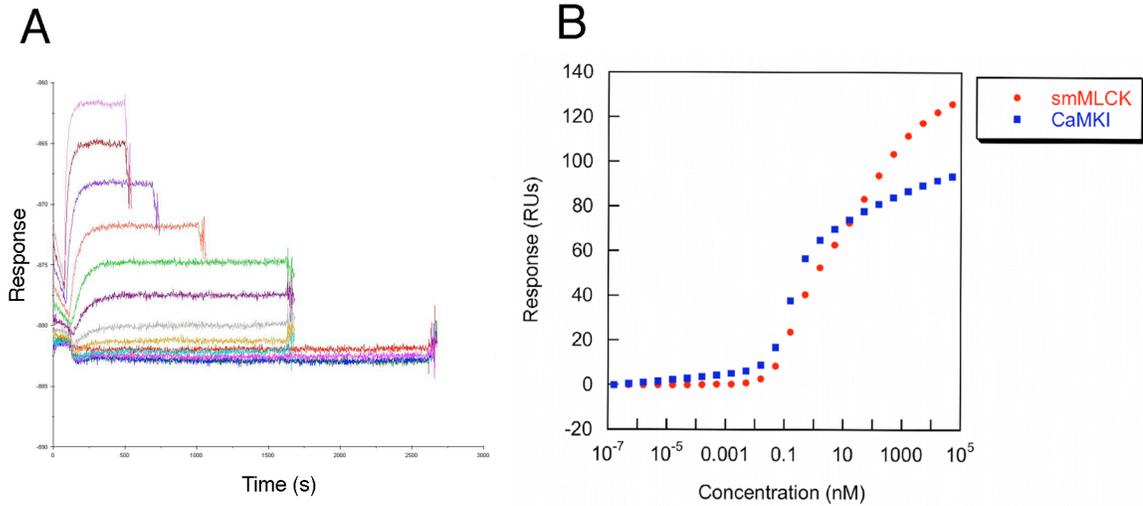
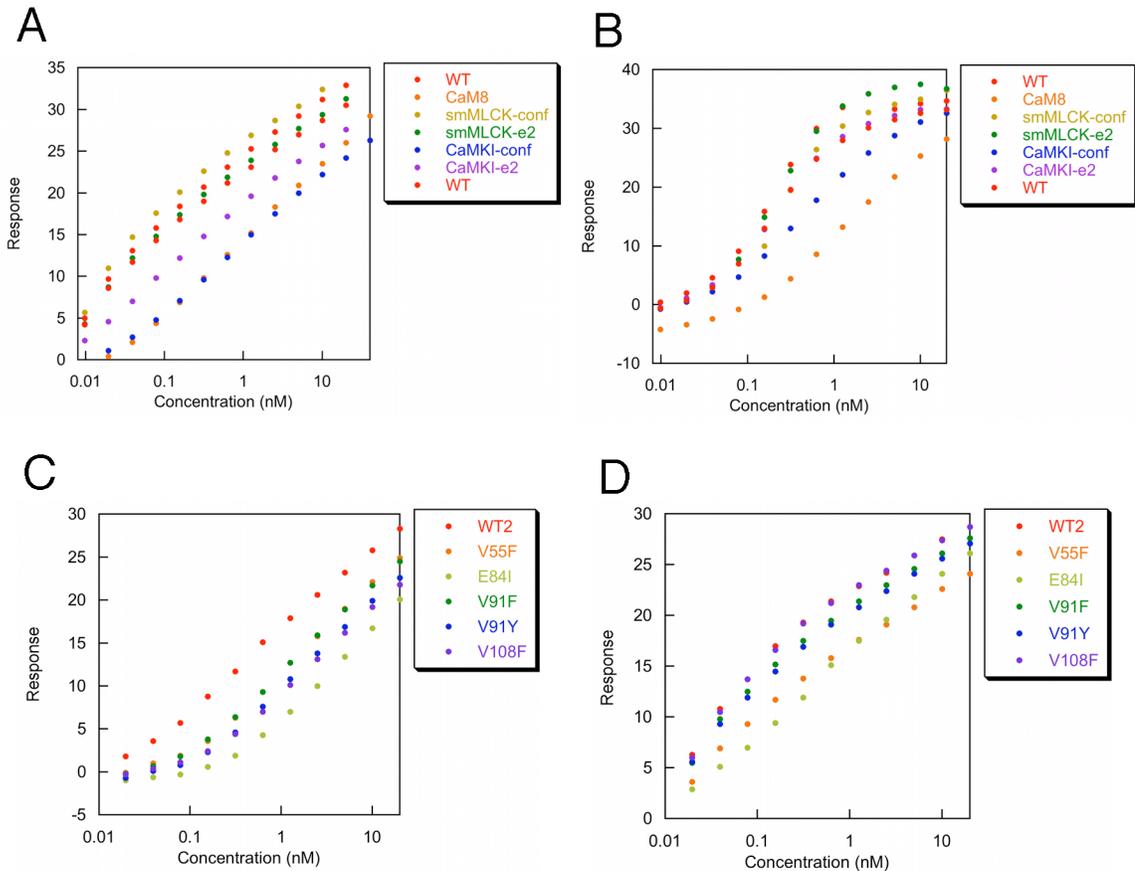


Figure A-5. Biacore equilibrium data. (A) Typical equilibrium data set. Multiple concentrations of CaM are injected over the peptide surface for various amounts of time until equilibrium binding is reached. (B) The equilibrium response value at each concentration is plotted as a function of concentration. For the wild-type interaction with both smMLCK_p and CaMKI_p, the interaction does not reach an R_{max} . This indicates that at high concentrations of CaM we are seeing a type of self-binding or aggregation.



E

For smMLCK_p binding:

WT = smMLCK-conf = smMLCK-e2 < CaMKI-e2 < CaMKI-conf = CaM8

WT < V55F = V91F = V91Y = V108F < E84I

For CaMKI_p binding:

WT = smMLCK-conf = smMLCK-e2 = CaMKI-e2 < CaMKI-conf < CaM8

WT = V91F = V91Y = V108F < V55F = E84I

Figure A-6. CaM variants-peptide equilibrium binding data. (A) Positive design CaM variants binding to smMLCK_p. (B) Positive design CaM variants binding to CaMKI_p. (C) Single mutant CaM variants derived from CaMKI-CaM positive design constructs binding to smMLCK_p. (D) Single mutant CaM variants derived from CaMKI-CaM positive design constructs binding to CaMKI_p. (E) Ranking of variants compared to WT based on Biacore equilibrium experiments.

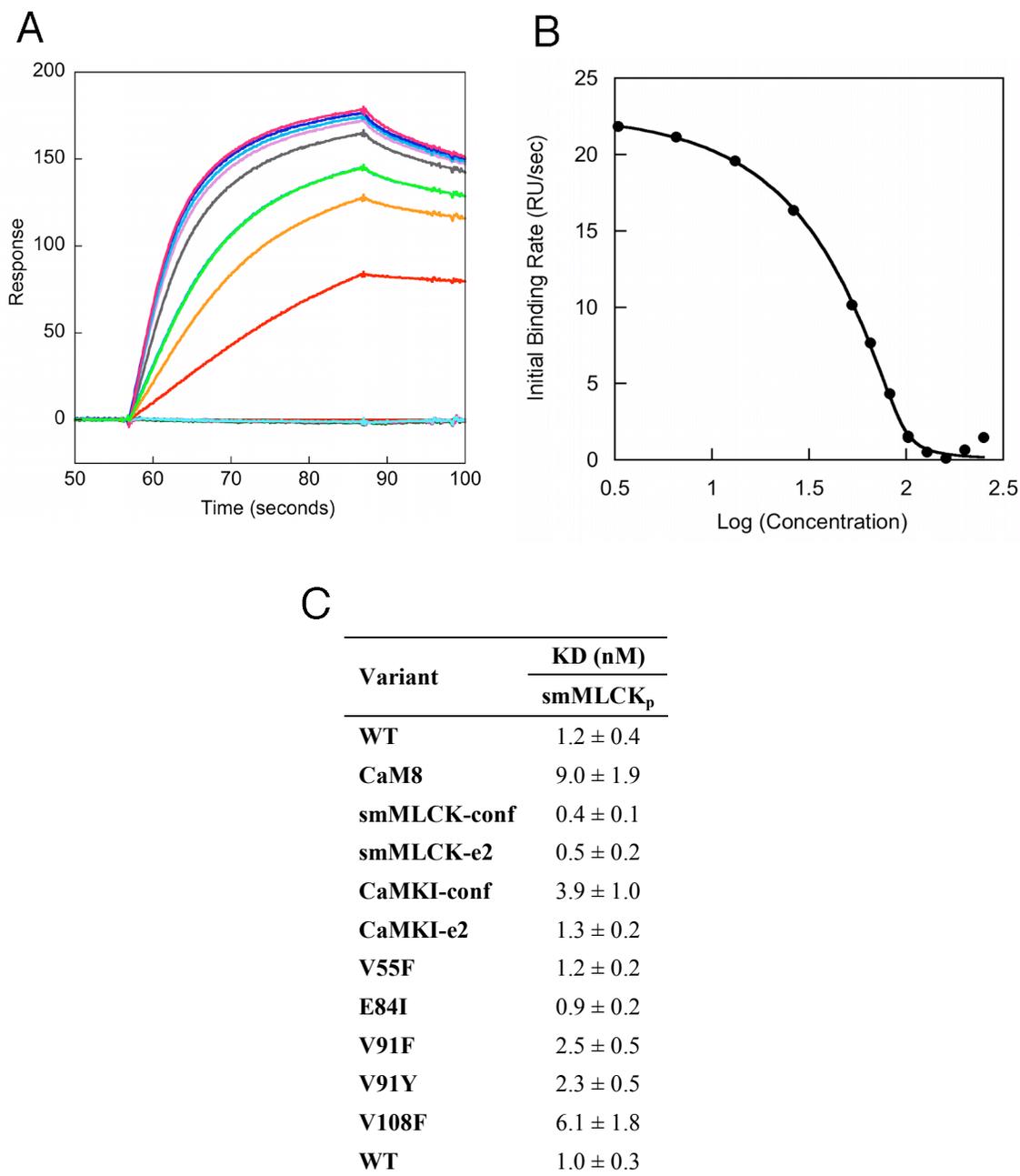


Figure A-7. Competition Biacore assay. (A) Sensorgrams showing the binding of CaM to the smMLCK_p surface and including increasing amounts of inhibiting smMLCK_p. As competitor is added, the maximal response decreases, as expected. (B) A plot of the initial binding rate as a function of the log of the competitor concentration. The initial rates are calculated from the first few seconds of each sensorgram as seen in (A). The data are fit to Equation A-3 to give a K_D . (C) Results of the competition assay. WT CaM was run both at the beginning of the experiment and the end to confirm the stability of the immobilized surface and for reproducibility.

Appendix B

Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity

The text of this chapter has been adapted from a manuscript coauthored with Jonathan Kyle Lassila, Peter Oelschlaeger, and Stephen L. Mayo.

Lassila, J.K., Keefe, J.R., Oelschlaeger, P., and Mayo, S.L. (2005) Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity *Protein Engineering, Design, and Selection* 18, 161-163.

Reproduced with permission from Lassila, J.K. *et al.* (2005)
Protein Engineering, Design, and Selection 18, 161-163.
Copyright 2005 Oxford University Press.

Abstract

Computational protein design methods were used to predict five variants of monofunctional *E. coli* chorismate mutase expected to maintain catalytic activity. The variants were tested experimentally and three active site mutations exhibited catalytic activity similar to or greater than the wild type enzyme. One mutation, Ala32Ser, showed increased catalytic efficiency.

Introduction

The Claisen rearrangement of chorismate to prephenate (Figure B-1) is a rare enzyme-catalyzed pericyclic reaction that proceeds through the same mechanism uncatalyzed in solution. Chorismate mutases from various organisms provide rate enhancements of around 10^6 despite strong dissimilarities in three-dimensional fold.¹⁻³ The metabolic importance of chorismate as the key branch point in the shikimate pathway has prompted extensive experimental investigation of the chorismate-prephenate rearrangement since the 1960s⁴ and has driven complementation experiments to probe the structural determinants of enzyme catalysis.⁵ The concerted, unimolecular nature of the rearrangement and the lack of covalent protein contacts have encouraged numerous theoretical studies of the catalyzed and uncatalyzed reactions. The question of how chorismate mutases achieve rate enhancement has been actively discussed in recent years.⁶⁻¹⁰

We used computational protein design techniques to identify mutations within the active site of the chorismate mutase domain (EcCM) of *Escherichia coli* chorismate mutase-prephenate dehydratase (“P-protein”)¹¹ consistent with catalytic activity. The objective of the study was to evaluate the viability of a rotamer-based approach with an empirical mechanics force field in modeling the active site. We report three active site mutations that permit catalytic activity at or above the level of the wild type enzyme. One of these appears to have enhanced catalytic efficiency relative to wild type.

Materials and Methods

Computational design. An *ab initio* calculated transition state structure¹² was modeled at the position of a transition state analog in the EcCM crystal structure, PDB code 1ECM.³ Translation (± 0.2 Å each axis) and rotation ($\pm 5^\circ$ each axis) of the transition state structure was allowed and all amino acids except glycine, proline, cysteine, and methionine were permitted in the following positions: 28, 32, 35, 39, 46, 47, 48, 51, 52, 55, 81, 84, 85, 88 (chain A); 7, 11, 14, 18 (chain B). Residues from both chains compose each of the two active sites in this symmetric homodimer. A backbone-dependent side chain rotamer library¹³ was used with expansion by one standard deviation about χ_1 and χ_2 for all amino acids except arginine and lysine. The energy function used in the ORBIT protein design software¹⁴ was based on the DREIDING force field¹⁵ and includes a scaled van der Waals term,¹⁶ hydrogen bonding and electrostatic terms,¹⁷ and a surface-area based solvation potential.¹⁸ Transition state partial atomic charges were obtained as previously reported.¹² An additional energy penalty was applied to effectively eliminate from consideration all sequences that could not maintain key contacts between the transition state structure and Arg 11 and Arg 28. These contacts were previously demonstrated to be necessary for catalysis.¹⁹ The HERO rotamer optimization method was used to obtain the minimum energy amino acid sequence and conformations.²⁰ Six mutations were predicted in the optimized structure: Leu7Ile, Ala32Ser, Val35Ile, Asp48Ile, Ile81Leu, and Val85Ile. Visual inspection and subsequent calculations indicated that four of the mutations were independent, so they were treated separately.

Protein expression and kinetic characterization. The gene encoding EcCM residues 1-109 of the bifunctional chorismate mutase-prephenate dehydratase (P protein) was amplified from genomic DNA (ATCC 700926D) and inserted into the pTYB11 vector from the IMPACT-CN intein fusion system (New England Biolabs) between the *SapI* and *XhoI* sites. Inverse PCR mutagenesis²¹ was used to construct five variants: Leu7Ile, Ala32Ser, Val35Ile, Asp48Ile, and Ile81Leu/Val85Ile. Constructs were verified by DNA sequencing. Variant and wild type proteins were expressed in *E. coli* BL21(DE3). Expression was induced with 1 mM IPTG at OD₆₀₀ = 0.6 and cells were grown for 18 hours at 22 °C. Cells were harvested and lysed by French press in 20 mM Tris-Cl pH 8.0, 500 mM NaCl, 1 mM EDTA, 10 mM MgCl₂, 1 mM PMSF, with DNase and RNase. Chitin affinity purification was conducted with a column buffer containing 20 mM Tris-Cl, pH 8.0, 500 mM NaCl, and 1 mM EDTA. On-column cleavage proceeded in 50 mM DTT at 25 °C for 18 hours. Mass spectrometry and SDS-PAGE verified that this protocol results in successful cleavage and yields the expected product. Wild type and variant proteins were further purified by gel filtration on a HiPrep Sephacryl S-100 high-resolution column (Amersham Biosciences) with 20 mM Tris-Cl pH 7.8, 100 mM NaCl as the running buffer and final storage buffer. Protein characterization followed procedures recently reported for the same construct.²² Protein concentration was determined by Bradford assay using BSA as a standard. Chorismate mutase activity was determined by following the disappearance of chorismate with UV absorbance at 275 nm. Activity assays were conducted at 37 °C and contained 20 nM protein in 50 mM Tris pH 7.8, 2.5 mM EDTA, 20 mM β-mercaptoethanol, and 0.01 % BSA. Initial velocities were buffer-corrected and were determined with less than 6 %

depletion of initial substrate concentration. All proteins were initially tested using a substrate concentration range of about 50-500 μM . The wild type and the Ala32Ser mutant were further assayed with substrate ranges of approximately 20-2000 μM and a minimum of five trials including two separate protein preparations each. Kinetic parameters were determined by non-linear fitting to the Michaelis-Menten equation.

Molecular dynamics simulations. Molecular dynamics simulations of EcCM and mutant Asp48Ile as free enzymes and in complex with the oxabicyclic transition state analog²³ were carried out using the *Sander* Program of the *AMBER 7.0* software package.²⁴ For symmetry, both chains were truncated to residues 6 through 95. The transition state analog was minimized and the electrostatic potential surface was calculated using DFT with B3LYP and the lacvp** basis set in the Jaguar 5.5 package (Schrödinger LLC). Atomic charges were obtained with *RESP*. Parameters for the transition state analog were defined according to similar structures in the *AMBER* libraries. The binding sites of the free enzymes were filled with water by solvating the molecules with a TIP3P²⁵ water shell using a closeness parameter of 0.4 in *xLEaP*. All systems were then solvated by a truncated octahedron of TIP3P water 8 Å around the protein and neutralized with counter ions. The ff02EP version of the ff99 force field^{26,27} was used to represent the protein and the ligand. After minimization, all systems were heated to 300 K with constant volume, starting at 10 K with three different initial velocities. Molecular dynamics simulations were prolonged at 300 K and constant pressure until the backbone RMSD reached a constant value of about 1.5 Å compared to

the first frame; this was observed after 500 to 800 ps. Average structures over 10 ps of the equilibrated systems were generated and analyzed visually.

Results and Discussion

As seen in Table B-1, three of the five variants showed similar or greater catalytic efficiency relative to the wild type enzyme. The Ala32Ser mutation results in a slightly more efficient catalyst than wild type due to both a decrease in K_M and an increase in k_{cat} . The fact that substrate binding is enhanced in addition to catalysis is consistent with the suggestion that factors stabilizing the transition state may be likely to contribute to ground state stabilization in the catalyzed rearrangement.²⁸ The rate enhancement corresponds to a change in activation energy of less than 1 kcal/mol, making a detailed structural explanation unwarranted. However, it should be noted that in the predicted structure of this mutant, Ser 32 is capable of hydrogen bonding with Gln 88 (Figure B-2), a residue that makes an essential contact to the ether oxygen of the breaking bond.¹⁹

Val35Ile shows increased k_{cat} but also increased K_M , resulting in k_{cat}/K_M similar to wild type. The Leu7Ile mutation also does not have a significant effect on catalytic efficiency. Both mutations create slightly different hydrophobic packing environments near the essential residue Arg 11 in the predicted structures. The double mutant, Ile81Leu/Val85Ile, was predicted to alter packing against the hydrophobic ring face of the reacting molecule. This rearrangement of hydrophobic amino acids does not result in substantial change in k_{cat} . However, the mutations result in increased K_M and reduced catalytic efficiency. The relative insensitivity of the enzyme to changes in some amino

acids is not surprising given the recent finding that the reaction is efficiently catalyzed by a protein exhibiting all the characteristics of a molten globule.²⁹

The Asp48Ile mutation abolished measurable catalysis under the conditions tested. This position makes backbone contacts to the hydroxyl group of the transition state analog in the EcCM crystal structure. Hydroxyl contacts to a negatively charged residue were suggested to create a favorable electrostatic gradient in the *Bacillus subtilis* enzyme.³⁰ However, in the *E. coli* structure the Asp 48 side chain points away from the active site and is distant from the transition state analog. Molecular dynamics simulations provide some insight into a possible function of Asp 48 that this mutant lacks. Averaged structures from equilibrated systems show Asp 48 hydrogen bonding to Arg 11 in both the unliganded and inhibitor-bound enzyme. In simulations of unliganded Asp48Ile mutants, however, Arg 11 is dramatically displaced into solution, suggesting that one role of Asp 48 may be to stabilize this key active site side chain in a conformation compatible with substrate binding and catalysis.

While the choice of mutations in this experiment was based solely on computational modeling and no sequence alignment information was used in the process, a BLAST search³¹ using the EcCM sequence as the query showed that sequence variations corresponding to the Ala32Ser, Val35Ile, Val85Ile, and Ile81Leu mutations were observed in chorismate mutases from related organisms.

Our design procedure stabilizes a static active site configuration with a bound transition state structure. Although the substrate of the reaction is not explicitly considered, we expect that modeling interactions using the structure and charges of the transition state should promote some degree of differential stabilization of the transition

state relative to substrate. The present study demonstrates that this approach can be used effectively to represent the active site of a natural enzyme. In this case, the predicted mutations were in residues not directly contacting the reacting molecule. The favorable result from the Ala32Ser mutation suggests that such secondary contacts are important in the enzyme design process. The complete loss of catalytic activity from the Asp48Ile mutation implies that improved treatment of electrostatics and consideration of the unbound enzyme could offer some benefit in future design efforts.

References

1. Chook, Y.M., Hengming, K. and Lipscomb, W.N. (1993) Crystal-structures of the monofunctional chorismate mutase from *Bacillus-subtilis* and its complex with a transition-state analog. *Proc. Natl Acad. Sci. USA*, **90**, 8600-8603.
2. Xue, Y., Lipscomb, W.N., Graf, R., Schnappauf, G. and Braus, G. (1994) The crystal-structure of allosteric chorismate mutase at 2.2-Ångstrom resolution. *Proc. Natl Acad. Sci. USA*, **91**, 10814-10818.
3. Lee, A.Y., Karplus, P.A., Ganem, B. and Clardy, J. (1995) Atomic-structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase. *J. Am. Chem. Soc.*, **117**, 3627-3628.
4. Gibson, M.I. and Gibson, F. (1964) Preliminary studies on isolation + metabolism of intermediate in aromatic biosynthesis - chorismic acid. *Biochem. J.*, **90**, 248-256.
5. Woycechowsky, K.J. and Hilvert, D. (2004) Deciphering enzymes – Genetic selection as a probe of structure and mechanism. *Eur. J. Biochem.*, **271**, 1630-1637.
6. Guimarães, C.R.W., Repasky, M.P., Chandrasekhar, J., Tirado-Rives, J. and Jorgensen, W.L. (2003) Contributions of conformational compression and preferential transition state stabilization to the rate enhancement by chorismate mutase. *J. Am. Chem. Soc.*, **125**, 6892-6899.
7. Strajbl, M., Shurki, A., Kato, M. and Warshel, A. (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J. Am. Chem. Soc.*, **125**, 10228-10237.
8. Hur, S. and Bruice, T.C. (2003) The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proc. Natl Acad. Sci. USA*, **100**, 12015-12020.
9. Crespo, A., Scherlis, D.A., Martí, M.A., Ordejón, P., Roitberg, A.E. and Estrin, D.A. (2003) A DFT-based QM-MM approach designed for the treatment of large molecular systems: Application to chorismate mutase. *J. Phys. Chem. B*, **107**, 13728-13736.
10. Ranaghan, K.E., Ridder, L., Szefczyk, B., Sokalski, W.A., Hermann, J.C. and Mulholland, A.J. (2004) Transition state stabilization and substrate strain in enzyme catalysis: *Ab initio* QM/MM modeling of the chorismate mutase reaction. *Org. Biomol. Chem.*, **2**, 968-980.
11. Stewart, J., Wilson, D.B. and Ganem, B. (1990) A genetically engineered monofunctional chorismate mutase. *J. Am. Chem. Soc.*, **112**, 4582-4584.

12. Wiest, O. and Houk, K.N. (1994) On the transition-state of the chorismate-prephenate rearrangement. *J. Org. Chem.*, **59**, 7582-7584.
13. Dunbrack, R.L. and Cohen, F.E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.*, **6**, 1661-1681.
14. Dahiyat, B.I. and Mayo, S.L. (1997) De novo protein design: Fully automated sequence selection. *Science*, **278**, 82-87.
15. Mayo, S.L., Olafson, B.D. and Goddard, W.A. (1990) DREIDING – A generic force-field for molecular simulations. *J. Phys. Chem.*, **94**, 8897-8909.
16. Dahiyat, B.I. and Mayo, S.L. (1997) Probing the role of packing specificity in protein design. *Proc. Natl Acad. Sci. USA*, **94**, 10172-10177.
17. Dahiyat, B.I., Gordon, D.B. and Mayo, S.L. (1997) Automated design of the surface positions of protein helices. *Protein Sci.*, **6**, 1333-1337.
18. Street, A.G. and Mayo, S.L. (1998) Pairwise calculation of protein solvent-accessible surface areas. *Fold. Des.*, **3**, 253-258.
19. Liu, D.R., Cload, S.T., Pastor, R.M. and Schultz, P.G. (1996) Analysis of active site residues in *Escherichia coli* chorismate mutase by site-directed mutagenesis. *J. Am. Chem. Soc.*, **118**, 1789-1790.
20. Gordon, D.B., Hom, G.K., Mayo, S.L. and Pierce, N.A. (2002) Exact rotamer optimization for protein design. *J. Comput. Chem.*, **24**, 232-243.
21. Hemsley, A., Arnheim, N., Toney, M.D., Cortopassi, G. and Galas, D.J. (1989) A simple method for site-directed mutagenesis using the polymerase chain-reaction. *Nucleic Acids Res.*, **17**, 6545-6551.
22. Zhang, S., Wilson, D.B. and Ganem, B. (2003) An engineered chorismate mutase with allosteric regulation. *Bioorg. Med. Chem.*, **11**, 3109-3114.
23. Bartlett, P.A., Johnson, C.R. (1985) An inhibitor of chorismate mutase resembling the transition-state conformation. *J. Am. Chem. Soc.*, **107**, 7792-7793.
24. Case, C.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E., Wang, J., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K. M. Jr., Stanton, R.V., Cheng, A.L., Vincent, J.J.; Crowley, M., Tsui, V., Gohlke, H., Radmer, R.J., Duan, J., Pitera, J., Massova, I., Seibel, G.L., Singh, U.C., Weiner, P.K., Kollman, P.A. AMBER 7, University of California, San Francisco, 2002.

25. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926-935.
26. Wang, J., Cieplak, P., Kollman, P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21**, 1049-1074.
27. Cornell, W.D., Cieplak, J.W., Bayly, C.I., Gould, I.R., Merz, K.M. Jr., Ferguson, D. M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A. (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.*, **117**, 5179-5197.
28. Strajbl, M., Shurki, A., Kato, M. and Warshel, A. (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J. Am. Chem. Soc.*, **125**, 10228-10237.
29. Vamvaca, K., Vögeli, B., Kast, P., Pervushin, K. and Hilvert, D. (2004) An enzymatic molten globule: Efficient coupling of folding and catalysis. *Proc. Natl Acad. Sci. USA*, **101**, 12860-12864.
30. Kast, P., Hartgerink, J.D., Asif-Ullah, M. and Hilvert, D. (1996) Electrostatic catalysis of the Claisen rearrangement: probing the role of Glu78 in *Bacillus subtilis* chorismate mutase by genetic selection. *J. Am. Chem. Soc.*, **118**, 3069-3070.
31. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389-3402.

Table B-1. Kinetic parameters of wild type and mutant EcCM^a

	k_{cat} min^{-1}	K_{M} μM	$k_{\text{cat}}/K_{\text{M}}$ $\text{min}^{-1}\mu\text{M}^{-1}$	%WT $k_{\text{cat}}/K_{\text{M}}$
WT	2332 ± 306	304 ± 52	7.8 ± 1.0	100
Ala32Ser	2708 ± 364	220 ± 29	12.4 ± 1.0	159
Val35Ile^b	3046 ± 172	365 ± 59	8.5 ± 1.1	109
Leu7Ile^b	2193 ± 291	249 ± 54	9.1 ± 2.4	117
Ile81Leu/Val85Ile^b	2004 ± 241	669 ± 165	3.1 ± 0.6	40
Asp48Ile^c	-	-	-	-

^a Reported errors are standard deviations from at least three trials.

^b Three mutants were assayed with a limited substrate concentration range, see text.

^c Reaction rates with the Asp48Ile mutant were within error of the uncatalyzed solution reaction.

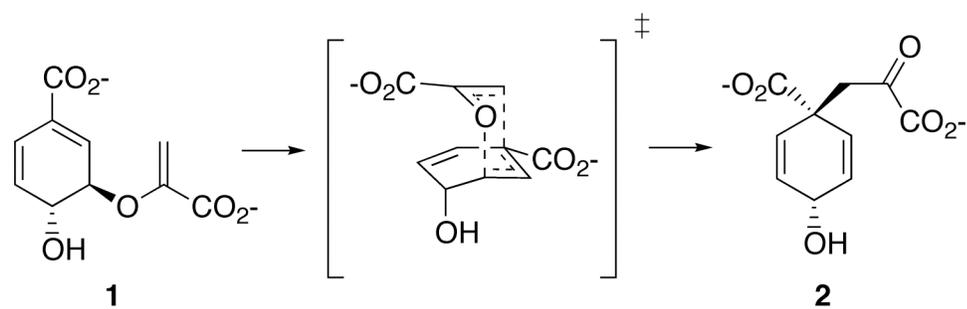


Figure B-1. Claisen rearrangement of chorismate (1) to prephenate (2).

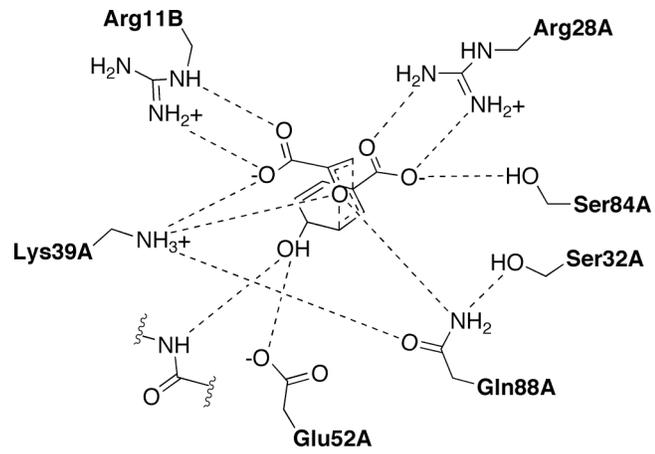


Figure B-2. Predicted hydrogen bonding in the Ala32Ser chorismate mutase variant. Interactions in the wild type crystal structure are the same except for position 32A.

Appendix C

Exhaustive mutagenesis of six secondary active site residues in *E. coli* chorismate mutase shows the importance of hydrophobic side chains and a helix N-capping position for stability and catalysis

The text of this chapter has been adapted from a manuscript coauthored with Jonathan Kyle Lassila, Peter Kast, and Stephen L. Mayo.

Lassila, J.K., Keeffe, J.R., Kast, P., and Mayo, S.L. (2007) Exhaustive mutagenesis of six secondary active site residues in *E. coli* chorismate mutase shows the importance of hydrophobic side chains and a helix N-capping position for stability and catalysis. *Biochemistry* 46, 6883-6891.

Abstract

Secondary active-site residues in enzymes, including hydrophobic amino acids, may contribute to catalysis through critical interactions that position the reacting molecule, organize hydrogen-bonding residues, and define the electrostatic environment of the active site. To ascertain the tolerance of an important model enzyme to mutation of active-site residues that do not directly hydrogen bond with the reacting molecule, all 19 possible amino acid substitutions were investigated in six positions of the engineered chorismate mutase domain of the *Escherichia coli* chorismate mutase-prephenate dehydratase. The six secondary active site residues were selected to clarify results of a previous test of computational enzyme design procedures. Five of the positions encode hydrophobic side chains in the wild-type enzyme and one forms a helix N-capping interaction as well as a salt bridge with a catalytically essential residue. Each mutant was evaluated for its ability to complement an auxotrophic chorismate mutase deletion strain. Kinetic parameters and thermal stabilities were measured for variants with *in vivo* activity. Altogether, we find that the enzyme tolerated 34% of the 114 possible substitutions, with a few mutations leading to increases in the catalytic efficiency of the enzyme. The results show the importance of secondary amino acid residues in determining enzymatic activity and they point to strengths and weaknesses in current computational enzyme design procedures.

Introduction

Chorismate mutases catalyze the Claisen rearrangement of chorismate to prephenate (Scheme C-1). This reaction is an essential step in the biosynthesis of aromatic compounds including tyrosine and phenylalanine in plants, bacteria, and fungi. Isotope studies have established that both solution and enzyme-catalyzed rearrangements proceed through a chairlike transition state^{1,2} in a concerted, asynchronous [3,3]-sigmatropic process.³⁻⁵

As an intramolecular reaction that appears to be catalyzed without intermediate steps, covalent catalysis, or modification of the reaction pathway, the chorismate-prephenate rearrangement has become an important model system for theoretical approaches to the study of enzyme catalysis (recently reviewed in ref. 6). These same features make the reaction a compelling model for evaluating procedures for computational enzyme design. In earlier work,⁷ we tested computational enzyme design methods⁸ through redesign of the active site of a chorismate mutase. Amino acid identities and conformations were optimized for 18 active site residues in the presence of an *ab initio* calculated transition state structure. The calculation returned wild-type residues in 12 of 18 designed positions, and the six mutations predicted by the calculation were to amino acids that do not make hydrogen bond contacts to the reacting molecule. Because the active-site environment is likely to present significant challenges to protein design energy functions that are typically tuned for global stability,⁹⁻¹⁰ we sought to better define the consequences for both stability and catalysis of mutating these six secondary active-site positions of the natural enzyme.

Two structurally distinct chorismate mutase forms have been identified. While the *Bacillus subtilis* enzyme (BsCM) and other AroH mutases adopt a homotrimeric pseudo- α/β -barrel fold,¹¹ the AroQ mutases from *Escherichia coli* and other organisms are all-helical proteins.¹²⁻¹⁵ Chorismate mutases from *E. coli* are fused to their downstream biosynthetic partners, yielding bifunctional chorismate mutase-prephenate dehydratase (P-protein) and chorismate mutase-prephenate dehydrogenase (T-protein) enzymes.¹⁶ Mechanistic investigations of the bifunctional T-protein enzyme were complicated by a rate-limiting step other than the Claisen rearrangement.³ However, truncated, monofunctional constructs of the P-protein (Figure C-1A, referred to as EcCM herein) have permitted structural and mechanistic studies of an isolated AroQ chorismate domain.¹⁷ The rearrangement catalyzed by EcCM shows expected patterns of kinetic isotope effects and no sensitivity to viscosogens, strongly suggesting that the Claisen rearrangement is the rate-determining step.^{5,18} In contrast, BsCM appears to be limited partially by a diffusive process and partially by the chemical step.^{4,18}

Previous mutagenesis and selection studies of both EcCM and BsCM enzymes have demonstrated that efficient catalysis depends on positively charged active site residues poised to interact with substrate carboxylates and ether oxygen as well as negatively charged side chains able to contact the substrate hydroxyl group.¹⁹⁻²⁵

The present work focuses on secondary active site side chains that do not hydrogen bond with the reacting molecule, but nonetheless may participate in catalysis by positioning the substrate, organizing charged and polar active site residues, or defining the electrostatic environment of the active site.²⁶⁻³¹ The impact of secondary active site residues on catalytic activity is an important concern for *de novo* computational enzyme

design efforts, because common methods strongly emphasize placement of primary catalytic residues.^{8,32-35} Secondary active site residues may also have special consequences for enzyme stability. Prior mutagenesis studies have found inverse relationships between changes in protein stability and changes in activity following the mutation of active site residues, consistent with a notion that creation of a cavity with specialized electrostatic properties opposes the normal driving forces of protein stability.³⁶⁻⁴⁰

We have evaluated the effects of all 19 possible single mutations to the six EcCM active site residues that were mutated in our previous computational design experiment: Leu7, Ala32, Val35, Asp48, Ile81, and Val85 (Figure C-1B). Each of the 114 variants was tested for complementation of the chorismate mutase deficiency of the auxotrophic *E. coli* KA12/pKIMP-UAUC system of Kast, Hilvert, and co-workers.²² Biologically active variants were purified and their *in vitro* thermal stabilities and steady-state kinetic parameters were assessed.

Materials and Methods

Materials and Strains. *E. coli* KA12 and plasmid pKIMP-UAUC have been described previously.^{22,41} Chorismic acid (97.5%) was purchased from Sigma-Aldrich.

Construction and in vivo evaluation of mutants. The chorismate mutase domain (consisting of the 109 N-terminal residues of the P-protein) described by Ganem and co-workers¹⁷ was previously fused to the C-terminus of a self-cleavable chitin binding domain-intein fusion construct (IMPACT-CN, New England Biolabs)⁷. The Quikchange

multi-site mutagenesis kit (Stratagene) was used to create site-specific variants in this construct by using single oligonucleotides bearing degenerate codons (i.e. NNK) in the position of interest. After transformation of XL10-gold cells (Stratagene), DNA was prepared directly from plated colonies and subsequently used to transform KA12/pKIMP-UAUC. Cells were plated on both selective media and non-selective media. M9c minimal agar plates were prepared as described⁴² except that 20 $\mu\text{g/mL}$ L-Phe was included in the selective plates (non-selective plates contained both L-Phe and L-Tyr). To facilitate identification of active variants, mutagenesis reactions were performed with inactive templates (i.e., stop codons in the position of interest) and DNA was sequenced from viable colonies on selective media. Inactive variants were identified by performing the mutagenesis reaction on a wild-type template and looking for colonies that were nonviable when streaked onto selective media plates. To recover missing variants, mutagenesis reactions with non-degenerate and less-degenerate oligonucleotides were performed. Biological activity or inactivity was confirmed by retransformation of isolated plasmid DNA and plating on selective and nonselective media. All reported constructs were verified by DNA sequencing over the entire EcCM coding region.

Because the N-terminal gene fusion was not sequenced for each variant, we investigated the possibility that disruption of activity might be caused by deleterious mutations to the promoter region or the N-terminally fused chitin-binding/intein region rather than by mutation of the enzyme itself. Evidence against such artifacts was provided for some inactive variants by multiple independent occurrences of the same mutation with identical behavior. For the remaining inactive variants, small-scale expression tests followed by SDS-PAGE verified the presence of appropriately sized, overproduced

protein. Any frameshift within the N-terminal fusion protein would encounter multiple stop codons leading to a visibly truncated protein. Similar expression tests of the N-terminal fusion protein alone showed that the size difference was readily apparent.

In position 48, some variants displayed ambiguous behavior, with only a few colonies growing on selective media. For position 48 variants, an arbitrary threshold was applied, such that a variant was considered biologically active if at least 10% of the number of colonies were observed on selective media as on nonselective media when equal amounts were plated. This resulted in a stringent test for activity, where active variants exhibited at least 5% of wild-type k_{cat}/K_m .

Protein overproduction and purification. Biologically active constructs were expressed in *E. coli* BL21 DE3. Overexpression was induced with 1 mM IPTG followed by growth at room temperature. Cells were lysed using an Emulsiflex-C5 (Avestin) and the fusion proteins were affinity purified with chitin agarose (New England Biolabs). Affinity tag self-cleavage was induced with 50 mM DTT at 25 °C. Mass spectrometry and SDS-PAGE previously indicated that this procedure results in the desired EcCM construct.⁷ Proteins were further purified by gel filtration in 50 mM Tris-HCl pH 7.8, 100 mM NaCl. A well-resolved dimer peak was collected, and variants with significant higher-order aggregation were not further considered. We expected that affinity purification followed by gel filtration would eliminate any significant catalytic contribution of endogenous *E. coli* CM enzymes, since they do not have the affinity tag and are considerably larger in size. In fact, no activity was detected when concentrated extracts from cells harboring empty pTYB-11 fusion vector were subjected to identical

purification steps and assay procedures. For consistency with previously published kinetic studies of the same construct,^{7,23,43} protein concentration was determined by the Bradford assay using BSA as a standard (Bio-Rad).

CD spectrometry and thermal denaturation. Far-UV circular dichroism (CD) spectra were collected at the activity assay temperature of 37 °C. All variants showed nearly superimposable wavelength scans. Thermostability was assessed by thermal denaturation monitored by CD at 222 nm. CD was monitored using an Aviv 62DS spectrometer and a 1 mm pathlength cell. Increments of 1 °C were used with an equilibration time of 90 seconds and a signal averaging time of 30 seconds. Buffer conditions were 50 mM Tris-HCl pH 7.8, 100 mM NaCl. Midpoints of thermal denaturation (T_m) were obtained by fitting the data to a two-state model.⁴⁴ It should be noted, however, that unfolding was not reversible and some samples showed clear evidence of non-two-state behavior, so T_m values should be taken as estimates of relative stabilities rather than thermodynamic parameters. CD data were collected at $22 \pm 3 \mu\text{M}$ protein. Variations in T_m from different concentrations within the range were similar to variations across multiple trials.

Kinetic assays. Activities were measured by following the absorbance of chorismate at 275 nm ($\epsilon = 2630 \text{ M}^{-1} \text{ cm}^{-1}$).³ Activity assays were performed at 37 °C in 50 mM Tris-HCl pH 7.8, 2.5 mM EDTA, 20 mM β -mercaptoethanol, and 0.01 % BSA, conditions used in prior studies of the same construct.^{7,23,43} Given $\text{p}K_{\text{aS}}$ for chorismate and prephenate of < 4.5 ,⁴⁵ these conditions ensure that the reaction of chorismate²⁻ to

prephenate²⁻ was monitored. A protein concentration of 20 nM was generally used; however, less-active variants were assayed with 40 nM or 100 nM protein. For cases where k_{cat} and K_m are reported, at least three trials with 9 or more data points each were performed, and substrate concentration ranges of at least 0.3-3 times K_m were used. Kinetic parameters were determined by non-linear fitting of the buffer-corrected initial rates to the Michaelis-Menten equation.

Results

The effects of all possible single amino acid substitutions were evaluated in six positions in the chorismate mutase domain of the *E. coli* P-protein (Table C-1). Each construct was evaluated for its ability to complement the KA12/pKIMP-UAUC system,^{22,41} which requires chorismate mutase activity for viability in media lacking tyrosine and phenylalanine. Of the 114 mutants, 39 (or 34%) were scored as active.

EcCM variants promoting viability were purified and their *in vitro* stabilities and kinetic parameters were evaluated. Of the mutants active *in vivo*, measured T_m values ranged approximately ± 10 °C relative to wild type, yet the average T_m (61.6 °C) was remarkably similar to that of the wild type (61.4 °C). *In vitro* enzyme activities were reduced relative to wild type, except in a few cases. The greatest increase in activity relative to wild type was found with the Ala32Ser mutation, previously reported as the result of a computational design approach.⁷ This mutation results in an approximately 60% increase in catalytic efficiency with less than 1 °C change in T_m .

Position 7. Based on the wild-type structure with bound transition state analog (1ECM), Leu7 lies near the carboxylate of the enolpyruvyl moiety, and the catalytically essential residue Arg11 is sandwiched between Leu7 and Arg51 from the adjacent helix. Six relatively conservative side chain substitutions were tolerated in this position. Substitution of Phe or Ile for wild-type Leu maintained catalytic efficiency within 10%, and the Leu to Ile mutation increased T_m by 4.4 °C.

Position 32. Ala32 lies closest to the forming C-C bond and the cyclohexadienyl face carboxylate, based on the 1ECM structure. Although six alternative side chains were tolerated in position 32, only the Ala32Ser and Ala32Thr variants could be readily purified without precipitation or aggregation. The Ala32Ser and Ala32Thr mutations increased catalytic efficiency by 62% and 23%, respectively.

Position 35. Val35 is expected to pack against the enolpyruvyl moiety of the reacting molecule. In addition, this residue closely neighbors the catalytically important residues Lys39 and Gln88, both of which contact the ether oxygen of the breaking bond. Six substitutions were tolerated in this position, with four leading to increased thermostability. The Val35Met, Val35Leu, and Val35Ile mutations increased T_m by 7.5 °C, 5.3 °C, and 5.2 °C, respectively. All mutations led to increased K_m , and all led to between 2- and 10-fold decreases in catalytic efficiency except for Val35Ile, which showed a 15% increase.

Position 48. In the wild-type structure, Asp48 hydrogen bonds with both the N-terminus of the second helix in EcCM and Arg11, a critical residue that contacts the transition state analog in the crystal structure (Figure C-2). Four position 48 mutants were

scored as biologically active, and all four showed reductions in T_m of 3-8 °C and large increases in K_m .

Position 81. Ile81 is expected to pack closely against the cyclohexadienyl face carboxylate of the reacting molecule. Six substitutions of this residue were tolerated, although the cysteine, tryptophan, and valine mutants were prone to aggregation *in vitro*. Aside from a 3.2 °C increase in stability with the Ile81Met mutation, the substitutions for which data could be obtained led to reduced stability and catalytic efficiency.

Position 85. Val 85 lies adjacent to Ile81 and close to the cyclohexadienyl face of the reacting molecule at the breaking bond carbon, based on the 1ECM structure. Its C α -C β vector is directed more towards solvent than that of Ile81, however. Position 85 was broadly tolerant of substitution, allowing 12 of 19 possible mutations. Interestingly, 9 of the 11 mutants for which data were recorded resulted in increased T_m relative to wild type, while all of the mutants yielded significant increases in K_m .

Discussion

Tolerance to mutation. In this study, 34% of all single mutants were scored as biologically active. While it is difficult to compare results across systems because of variations in cellular requirements, expression levels, initial protein stability,⁴⁶⁻⁴⁸ and various other factors, this value is consistent with other mutational studies. A sampling from among the many exhaustive and random mutagenesis studies suggests that between 50 and 85% of single mutations typically maintain some level of function, while in catalytic sites, 20-60% on average preserve some function.⁴⁹⁻⁵⁷ Thus, the amino acid

positions studied here have tolerances to single mutation that are similar to active site residues in other enzymes.

Helix N-capping position. The N-termini of alpha helices are preceded by residues with distinct preferences for side chains that can satisfy unpaired helical hydrogen bond donors. These N-capping residues are reflected in protein database surveys, protein stability measurements, and model peptide studies (reviewed in ref. 58). N-cap residues can affect helix stability by as much as 4 kcal/mol⁵⁹ and are found to be important considerations in protein design.⁶⁰ The 1ECM crystal structure suggests that Asp48 occupies an ideal geometry for simultaneous N-capping and salt bridge formation with Arg11, a critical residue that contacts the transition state analog (Figure C-2). Mutational studies have shown the importance of Arg11 to catalysis; the Arg11Lys and Arg11Ala mutations lead to 10³-fold and 10⁴-fold reduction in catalytic efficiency, respectively.²⁰

The reduction in stability of all biologically active position 48 variants coincides with previous studies placing Asp as one of the most favored N-capping residues.⁶¹⁻⁶³ The loss in activity upon mutation may suggest that the Asp48-Arg11 salt bridge is important for positioning of the catalytically important arginine. Alternatively, the interaction may have a role in the substrate binding process. The salt bridge spans the interface between two helices proposed to offer entry into the binding pocket. Structural studies on related enzymes suggest that these two helices move relative to each other, potentially opening an entry pathway.¹⁴ Mutations in position 48 could conceivably interfere with the substrate entry process, which may be reflected in the increased K_m values. The fact that Asp48Asn and Asp48Gln mutants were scored as active while Asp48Glu was not could

suggest that a very precise geometric positioning is required for the catalytic process, which may be compromised in a backbone-Glu48-Arg11 hydrogen bond network.

Comparison to other AroQ family members. Table C-2 shows the structural equivalents of each of the six residues in crystal structures of AroQ chorismate mutases from four other organisms. For each structure, sequence variations among closely related homologs were determined by BLAST searches of the RefSeq database.^{64,65} A number of the sequence variations seen in the homologs were found to retain wild-type-like or have increased activity in the EcCM scaffold: Leu7Ile, Ala32Ser, Ala32Thr, and Val35Ile. Four other mutations were also observed in the sequence alignments but led to precipitated or aggregated protein *in vitro*: Ala32Val, Ala32Gly, Ala32Cys, and Ile81Val. The majority of the alternative residues found in the sequence alignments, however, led in EcCM to reduced *in vitro* activity or loss of function in the *in vivo* assay. Clearly, sequence alignments provide information about possible substitution tolerances in a given enzyme scaffold but activity is ultimately determined within the structural context of the specific protein.

The considerable tolerance to mutation seen with position 85 is also reflected in the sequence alignments. Related proteins feature alternatives including Glu, Lys, and Arg, consistent with the higher solvent exposure of this position. While a previous selection study reported the equivalent of position 85 in *Methanococcus jannaschii* CM to be highly conserved,⁶⁶ that observation may reflect interactions not present in the EcCM structure. The equivalent position is occupied by Lys in the *M. jannaschii* enzyme, as well as in the *Saccharomyces cerevisiae* and *Pyrococcus furiosus* enzymes, where crystal structures show the lysine engaging in salt bridge interactions with a

neighboring helix.¹³ Curiously, the equivalent of position 85 in the *Mycobacterium tuberculosis* enzyme makes a catalytically important contact to the hydroxyl group of the transition state analog that replaces the interaction made in EcCM by Glu52,¹⁴ again suggesting that the role of this position may be strongly scaffold-dependent.

In position 48, the sequence alignments showed much more variation in amino acid identity than was experimentally found to be compatible with EcCM function. Changes in helix register in some of the homologs may be partially responsible for these large variations.

Tradeoff between stability and activity. Various authors have considered the view that evolutionary optimization of active site residues for binding of transition states would lead to selection of residues that are not optimal for protein stability.³⁶⁻⁴⁰ Implicit in the argument is the idea that stabilizing interactions that the enzyme devotes to the reacting molecule are both unsatisfied and likely destabilizing to the protein when the molecule is not bound. A laboratory-evolved chorismate mutase dramatically illustrated this principle by transforming from an apparent molten globule into a more ordered protein in the presence of a transition state analog.⁶⁷⁻⁶⁸

Although a correlation was not seen between changes in stability and activity of the chorismate mutase variants studied here (see supporting information), some of the mutants are consistent with the notion of a stability-activity tradeoff. More mutants (twenty) showed increased stability than decreased stability (twelve). In position 85 of EcCM, all of the eleven biologically active mutants for which data could be obtained showed reduced catalytic efficiency and an increased K_m value and nearly all mutants exhibited increased thermostability. Because Val85 participates in van der Waals

interactions with both the transition state analog and residues from neighboring helices in the 1ECM structure, chorismate binding and stability could be affected by altering the shape of the binding pocket or by interfering with helix-helix interactions that contribute to the structural fortitude of the active site. Specifically, movement of the neighboring helix has been proposed to facilitate substrate entry¹⁴ and modified helix-helix interaction could thus affect substrate affinity. Also, in position 35, all biologically active mutations increased the K_m . While not all of the mutations led to increased T_m values, three showed increases in thermostability of more than 4 °C. Val35 in the 1ECM structure appears to participate in van der Waals interactions with the transition state analog, the catalytically important residues Lys39 and Gln88, and some neighboring helix residues. While the present study cannot be expected to determine the relative contributions of these interactions to EcCM function, it clearly suggests that this residue contributes to the catalytic process and the global stability of the protein.

Relationship to models of EcCM catalytic mechanism. Among theorists, differing computational treatments of chorismate mutase systems have led to varied descriptions of the origin of enzymatic catalysis of the reaction. Despite an observation that the chemical step is not fully rate limiting in the BsCM enzyme,¹⁸ the majority of work has modeled that active site.⁶⁹⁻⁸⁷ The few computational investigations of EcCM have proposed, in opposition to most of the other studies, that both enzymes accelerate the rearrangement exclusively by restricting the conformational space of the substrate, and not by stabilizing the transition state.^{79,88-90} These studies, and an earlier one,⁹¹ explicitly identified Val35 and Ile81 as mediators of substrate conformational restriction. Although we find that both

positions tolerate a number of mutations, we do not expect the data provided here to allow for differentiation between possible modes of EcCM catalysis.

For example, the Val35Ile mutation leads to an increase in k_{cat} and the Val35Ala mutation leads to a decreased k_{cat} value. While these mutations would be expected to reduce and increase the conformational freedom of the bound substrate, respectively, the changes in k_{cat} are relatively modest and could be attributed to any number of factors. In addition to its close proximity to the reacting molecule, Val35 is within van der Waals contact distance of the catalytically essential Lys39 and a number of other residues. Perturbations to any of these interactions could have consequences for the geometry and energetics of critical hydrogen bonds, leading to changes in relative stabilization of the transition state. Simple continuum electrostatics calculations suggest that these mutations could also lead to small changes in electrostatic potential at the reacting molecule. In short, hydrophobic active site residues perform important functions in active sites, but it is difficult to isolate their precise effects on catalysis.

Implications for enzyme design. In the computational design experiment previously performed by our group,⁷ the global minimum energy sequence for 18 designed EcCM active site residues contained mutations to only the six positions of the present study: Leu7Ile, Ala32Ser, Val35Ile, Asp48Ile, Ile81Leu, and Val85Ile. Of these six mutations, two were found to be the most active of all possible single substitutions for the given amino acid position, and one was found to be the most stable variant of the amino acid position.

While the five hydrophobic positions were generally modeled well, the strong preference for aspartate in position 48 is not well represented by the molecular mechanics

and rotamer library-based protein design approach. In typical energy models used for design, repulsion of the charged Asp side chain with the carboxylate of the reacting molecule results in disfavoring of the wild-type position 48 residue. Regardless of whether Asp48 contributes to catalysis by positioning of the critical Arg11, solidifying the local structure around the helix N-terminus, or contributing to substrate entry into the binding site, the effect of this residue is likely due to complex and cooperative effects of Asp48 that are difficult to model in a design calculation. Present computational protein design models typically neglect explicit treatment of side-chain entropy, electron polarization, the unbound state of the protein, and changes in backbone structure, any of which could prove critical in defining the energetics that underlie the strong requirement for aspartate in position 48. The present dataset will enable a more detailed evaluation of the strengths and weaknesses of various energy functions for design.

Finally, the ability of EcCM to tolerate a number of mutations without reduction in stability or activity is in accord with previous descriptions of active site plasticity. Secondary active site positions that accept side chain substitutions can allow evolution or design of enzymes with new substrate specificities and chemical mechanisms, sometimes without disruption of the original function.⁹²⁻⁹⁵ These secondary positions offer rich opportunities for testing and refining enzyme design methodologies through active site redesign experiments.

References

1. Sogo, S. G., Widlanski, T. S., Hoare, J. H., Grimshaw, C. E., Berchtold, G. A., and Knowles, J. R. (1984) Stereochemistry of the rearrangement of chorismate to prephenate: Chorismate mutase involves a chair transition state, *J. Am. Chem. Soc.* *106*, 2701-2703.
2. Copley, S. D., and Knowles, J.R. (1985) The uncatalyzed Claisen rearrangement of chorismate to prephenate prefers a transition state of chairlike geometry, *J. Am. Chem. Soc.* *107*, 5306-5308.
3. Addadi, L., Jaffe, E. K., and Knowles, J. R. (1983) Secondary tritium isotope effects as probes of the enzymic and nonenzymic conversion of chorismate to prephenate, *Biochemistry* *22*, 4494-4501.
4. Gustin, D. J., Mattei, P., Kast, P., Wiest, O., Lee, L., Cleland, W. W., and Hilvert, D. (1999) Heavy atom isotope effects reveal a highly polarized transition state for chorismate mutase, *J. Am. Chem. Soc.* *121*, 1756-1757.
5. Wright, S. K., DeClue, M. S., Mandal, A., Lee, L., Wiest, O., Cleland, W. W., and Hilvert, D. (2005) Isotope effects on the enzymatic and nonenzymatic reactions of chorismate, *J. Am. Chem. Soc.* *127*, 12957-12964.
6. Gao, J. L., Ma, S. H., Major, D. T., Nam, K., Pu, J. Z., and Truhlar, D. G. (2006) Mechanisms and free energies of enzymatic reactions, *Chem. Rev.* *106*, 3188-3209.
7. Lassila, J. K., Keeffe, J. R., Oelschlaeger, P., and Mayo, S. L. (2005) Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity, *Prot. Eng. Des. Sel.* *18*, 161-163.
8. Lassila, J. K., Privett, H. K., Allen, B. D., and Mayo, S. L. (2006) Combinatorial methods for small-molecule placement in computational enzyme design, *Proc. Natl. Acad. Sci. USA* *103*, 16710-16715.
9. Mendes, J., Guerois, R., and Serrano, L. (2002) Energy estimation in protein design, *Curr. Opin. Struct. Biol.* *12*, 441-446.
10. Vizcarra, C. L., and Mayo, S. L. (2005) Electrostatics in computational protein design, *Curr. Opin. Chem. Biol.* *9*, 622-626.
11. Chook, Y. M., Ke, H., and Lipscomb, W. N. (1993) Crystal structures of the monofunctional chorismate mutase from *Bacillus subtilis* and its complex with a transition state analog, *Proc. Natl. Acad. Sci. USA* *90*, 8600-8603.

12. Lee, A. Y., Karplus, P. A., Ganem, B., and Clardy, J. (1995) Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase, *J. Am. Chem. Soc.* *117*, 3627-3628.
13. Xue, Y. F., Lipscomb, W. N., Graf, R., Schnappauf, G., and Braus, G. (1994) The crystal structure of allosteric chorismate mutase at 2.2-Ångstrom resolution, *Proc. Natl. Acad. Sci. USA* *91*, 10814-10818.
14. Ökvist, M., Dey, R., Sasso, S., Grahn, E., Kast, P., and Krenzel, U. (2006) 1.6 Å crystal structure of the secreted chorismate mutase from *Mycobacterium tuberculosis*: Novel fold topology revealed, *J. Mol. Biol.* *357*, 1483-1499.
15. Kim, S.-K., Reddy, S. K., Nelson, B. C., Vasquez, G. B., Davis, A., Howard, A. J., Patterson, S., Gilliland, G. L., Ladner, J. E., and Reddy, P. T. (2006) Biochemical and structural characterization of the secreted chorismate mutase (Rv1885c) from *Mycobacterium tuberculosis* H₃₇R_v: an *AroQ enzyme not regulated by the aromatic amino acids, *J. Bacteriol.* *188*, 8638-8648.
16. Cotton, R. G. H., and Gibson, F. (1965) The biosynthesis of phenylalanine and tyrosine: Enzymes converting chorismic acid into prephenic acid and their relationships to prephenate dehydratase and prephenate dehydrogenase, *Biochim. Biophys. Acta* *100*, 76-88.
17. Stewart, J., Wilson, D. B., and Ganem, B. (1990) A genetically engineered monofunctional chorismate mutase, *J. Am. Chem. Soc.* *112*, 4582-4584.
18. Mattei, P., Kast, P., and Hilvert, D. (1999) *Bacillus subtilis* chorismate mutase is partially diffusion-controlled, *Eur. J. Biochem.* *261*, 25-32.
19. Cload, S. T., Liu, D. R., Pastor, R. M., and Schultz, P. G. (1996) Mutagenesis study of active site residues in chorismate mutase from *Bacillus subtilis*, *J. Am. Chem. Soc.* *118*, 1787-1788.
20. Liu, D. R., Cload, S. T., Pastor, R. M., and Schultz, P. G. (1996) Analysis of active site residues in *Escherichia coli* chorismate mutase by site-directed mutagenesis, *J. Am. Chem. Soc.* *118*, 1789-1790.
21. Kast, P., Hartgerink, J. D., Asif-Ullah, M., and Hilvert, D. (1996) Electrostatic catalysis of the Claisen rearrangement: Probing the role of Glu78 in *Bacillus subtilis* chorismate mutase by genetic selection, *J. Am. Chem. Soc.* *118*, 3069-3070.
22. Kast, P., Asif-Ullah, M., Jiang, N., and Hilvert, D. (1996) Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: The importance of electrostatic catalysis, *Proc. Natl. Acad. Sci. USA* *93*, 5043-5048.

23. Zhang, S., Kongsaree, P., Clardy, J., Wilson, D. B., and Ganem, B. (1996) Site-directed mutagenesis of monofunctional chorismate mutase engineered from the *E. coli* P-protein, *Bioorg. Med. Chem.* 4, 1015-1020.
24. Kast, P., Grisostomi, C., Chen, I.A., Li, S. L., Krenzel, U., Xue, Y. F., and Hilvert, D. (2000) A strategically positioned cation is crucial for efficient catalysis by chorismate mutase, *J. Biol. Chem.* 275, 36832-36838.
25. Kienhöfer, A., Kast, P., and Hilvert, D. (2003) Selective stabilization of the chorismate mutase transition state by a positively charged hydrogen bond donor, *J. Am. Chem. Soc.* 125, 3206-3207.
26. Warshel, A. (1978) Energetics of enzyme catalysis, *Proc. Natl. Acad. Sci. USA* 75, 5250-5254.
27. Honig, B., and Nicholls, A. (1995) Classical electrostatics in biology and chemistry, *Science* 268, 1144-1149.
28. Warshel, A. (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites, *J. Biol. Chem.* 273, 27035-27038.
29. Shan, S. O., and Herschlag, D. (1996) The change in hydrogen bond strength accompanying charge rearrangement: Implications for enzymatic catalysis, *Proc. Natl. Acad. Sci. USA* 93, 14474-14479.
30. Kraut, D. A., Carroll, K. S., and Herschlag, D. (2003) Challenges in enzyme mechanism and energetics, *Annu. Rev. Biochem.* 72, 517-571.
31. Benkovic, S. J., and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis, *Science* 301, 1196-1202.
32. Hellinga, H. W., and Richards, F. M. (1991) Construction of new ligand binding sites in proteins of known structure, *J. Mol. Biol.* 222, 763-785.
33. Bolon, D. N., and Mayo, S. L. (2001) Enzyme-like proteins by computational design, *Proc. Natl. Acad. Sci. USA* 98, 14274-14279.
34. Dwyer, M. A., Looger, L. L., and Hellinga, H. W. (2004) Computational design of a biologically active enzyme, *Science* 304, 1967-1971.
35. Zanghellini, A., Jiang, L., Wollacott, A. M., Cheng, G., Althoff, E. A., Rothlisberger, D., and Baker, D. (2006) New algorithms and an *in silico* benchmark for computational enzyme design, *Protein Sci.* 15, 2785-2794.
36. Zhi, W., Srere, P. A., and Evans, C. T. (1991) Conformational stability of pig citrate synthase and some active-site mutants, *Biochemistry* 30, 9281-9286.

37. Meiering, E. M., Serrano, L., and Fersht, A. R. (1992) Effect of active site residues in barnase on activity and stability, *J. Mol. Biol.* 225, 585-589.
38. Shoichet, B. K., Baase, W. A., Kuroki, R., and Matthews, B. W. (1995) A relationship between protein stability and protein function, *Proc. Natl. Acad. Sci. USA* 92, 452-456.
39. Beadle, B. M., and Shoichet, B. K. (2002) Structural bases of stability-function tradeoffs in enzymes, *J. Mol. Biol.* 321, 285-296.
40. Mukaiyama, A., Haruki, M., Ota, M., Koga, Y., Takano, K., and Kanaya, S. (2006) A hyperthermophilic protein acquires function at the cost of stability, *Biochemistry* 45, 12673-12679.
41. Kast, P., Asif-Ullah, M., and Hilvert, D. (1996) Is chorismate mutase a prototypic entropy trap? – Activation parameters for the *Bacillus subtilis* enzyme, *Tet. Lett.* 37, 2691-2694.
42. Gamper, M., Hilvert, D., and Kast, P. (2000) Probing the role of the C-terminus of *Bacillus subtilis* chorismate mutase by a novel random protein-termination strategy, *Biochemistry* 39, 14087-14094.
43. Zhang, S., Wilson, D. B., and Ganem, B. (2003) An engineered chorismate mutase with allosteric regulation, *Bioorg. Med. Chem.* 11, 3109-3114.
44. Becktel, W. J., and Schellman, J. A. (1987) Protein stability curves, *Biopolymers* 26, 1859-1877.
45. Kast, P., Tewari, Y. B., Wiest, O., Hilvert, D., Houk, K. N., and Goldberg, R. N. (1997) Thermodynamics of the conversion of chorismate to prephenate: Experimental results and theoretical predictions, *J. Phys. Chem. B* 101, 10976-10982.
46. Bloom, J. D., Silberg, J. J., Wilke, C. O., Drummond, D. A., Adami, C., and Arnold, F. H. (2005) Thermodynamic prediction of protein neutrality, *Proc. Natl. Acad. Sci. USA* 102, 606-611.
47. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D. S. (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein, *Nature* 444, 929-932.
48. Bloom, J. D., Arnold, F. H., and Wilke, C. O. (2007) Breaking proteins with mutations: Threads and thresholds in evolution, *Mol. Syst. Biol.* 3, 76.

49. Shortle, D., and Lin, B. (1985) Genetic analysis of staphylococcal nuclease: Identification of three intragenic “global” suppressors of nuclease-minus mutations, *Genetics* 110, 539-555.
50. Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E., and Hutchison, C. A. (1989) Complete mutagenesis of the HIV-1 protease, *Nature* 340, 397-400.
51. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., and Sauer, R. T. (1990) Deciphering the message in protein sequences: Tolerance to amino acid substitutions, *Science* 247, 1306-1310.
52. Rennell, D., Bouvier, S. E., Hardy, L. W., and Poteete, A. R. (1991) Systematic mutation of bacteriophage T4 lysozyme, *J. Mol. Biol.* 222, 67-87.
53. Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., and Miller, J. H. (1994) Genetic studies of the *lac* repressor XIV: Analysis of 4000 altered *Escherichia coli lac* repressors reveals essential and non-essential residues, as well as “spacers” which do not require a specific sequence, *J. Mol. Biol.* 240, 421-433.
54. Guo, H. H., Choe, J., and Loeb, L. A. (2004) Protein tolerance to random amino acid change, *Proc. Natl. Acad. Sci. USA* 101, 9205-9210.
55. Liu, L., and Santi, D. V. (1993) Asparagine 229 in thymidylate synthase contributes to, but is not essential for, catalysis, *Proc. Natl. Acad. Sci. USA* 90, 8604-8608.
56. Warren, M. S., Marolewski, A. E., and Benkovic, S. J. (1996) A rapid screen of active site mutants in glycinamide ribonucleotide transformylase, *Biochemistry* 35, 8855-8862.
57. Axe, D. D., Foster, N. W., and Fersht, A. R. (1998) A search for single substitutions that eliminate enzymatic function in a bacterial ribonuclease, *Biochemistry* 37, 7157-7166.
58. Aurora, R., and Rose, G. D. (1998) Helix capping. *Protein Sci.* 7, 21-38.
59. Doig, A. J., and Baldwin, R. L. (1995) N- and C-capping preferences for all 20 amino acids in α -helical peptides. *Protein Sci.* 4, 1325-1336.
60. Marshall, S. A., Morgan, C. S., and Mayo, S. L. (2002) Electrostatics significantly affect the stability of designed homeodomain variants. *J. Mol. Biol.* 316, 189-199.
61. Richardson, J. S., and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of α -helices. *Science* 240, 1648-1652.

62. Serrano, L., and Fersht, A. R. (1989) Capping and α -helix stability. *Nature* 342, 296-299.
63. Doig, A. J., MacArthur, M. W., Stapley, B. J., and Thornton, J. M. (1997) Structures of N-termini of helices in proteins. *Protein Sci.* 6, 147-155.
64. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs, *Nucleic Acids Res.* 25, 3389-3402.
65. Pruitt, K. D., Tatusova, T., Maglott, D. R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, *Nucleic Acids Res.* 33, D501-D504.
66. Taylor, S. V., Walter, K. U., Kast, P., and Hilvert, D. (2001) Searching sequence space for protein catalysts, *Proc. Natl. Acad. Sci. USA* 98, 10596-10601.
67. MacBeath, G., Kast, P., and Hilvert, D. (1998) Redesigning enzyme topology by directed evolution, *Science* 279, 1958-1961.
68. Vamvaca, K., Vögeli, B., Kast, P., Pervushin, K., and Hilvert, D. (2004) An enzymatic molten globule: Efficient coupling of folding and catalysis, *Proc. Natl. Acad. Sci. USA* 101, 12860-12864.
69. Lyne, P. D., Mulholland, A. J., and Richards, W. G. (1995) Insights into chorismate mutase catalysis from a combined QM/MM simulation of the enzyme reaction, *J. Am. Chem. Soc.* 117, 11345-11350.
70. Davidson, M. M., Gould, I. R., and Hillier, I. H. (1996) The mechanism of the catalysis of the Claisen rearrangement of chorismate to prephenate by the chorismate mutase from *Bacillus subtilis*. A molecular mechanics and hybrid quantum mechanical/molecular mechanical study, *J. Chem. Soc., Perkin Trans. 2*, 525-532.
71. Hall, R. J., Hindle, S. A., Burton, N. A., and Hillier, I. H. (2000) Aspects of hybrid QM/MM calculations: The treatment of the QM/MM interface regions and geometry optimization with an application to chorismate mutase, *J. Comp. Chem.* 21, 1433-1441.
72. Worthington, S. E., Roitberg, A. E., and Krauss, M. (2001) An MD/QM study of the chorismate mutase-catalyzed Claisen rearrangement reaction, *J. Phys. Chem. B* 105, 7087-7095.
73. Guo, H., Cui, Q., Lipscomb, W. N., and Karplus, M. (2001) Substrate conformational transitions in the active site of chorismate mutase: Their role in the catalytic mechanism, *Proc. Natl. Acad. Sci. USA* 98, 9032-9037.

74. Marti, S., Andres, J., Moliner, V., Silla, E., Tunon, I., Bertran, J., and Field, M. J. (2001) A hybrid potential reaction path and free energy study of the chorismate mutase reaction, *J. Am. Chem. Soc.* *123*, 1709-1712.
75. Lee, Y. S., Worthington, S. E., Krauss, M., and Brooks, B. R. (2002) Reaction mechanism of chorismate mutase studied by the combined potentials of quantum mechanics and molecular mechanics, *J. Phys. Chem. B* *106*, 12059-12065.
76. Guimarães, C. R. W., Repasky, M. P., Chandrasekhar, J., Tirado-Rives, J., and Jorgensen, W. L. (2003) Contributions of conformational compression and preferential transition state stabilization to the rate enhancement by chorismate mutase, *J. Am. Chem. Soc.* *125*, 6892-6899.
77. Štrajbl, M., Shurki, A., Kato, M., and Warshel, A. (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization, *J. Am. Chem. Soc.* *125*, 10228-10237.
78. Hur, S., and Bruice, T. C. (2003) The near attack conformation approach to the study of the chorismate to prephenate reaction, *Proc. Natl. Acad. Sci. USA* *100*, 12015-12020.
79. Crespo, A., Scherlis, D. A., Marti, M. A., Ordejon, P., Roitberg, A. E., and Estrin, D. A. (2003) A DFT-based QM-MM approach designed for the treatment of large molecular systems: Application to chorismate mutase *J. Phys. Chem. B* *107*, 13728-13736.
80. Guo, H., Cui, Q., Lipscomb, W. N., and Karplus, M. (2003) Understanding the role of active-site residues in chorismate mutase catalysis from molecular-dynamics simulations, *Angew. Chem. Int. Ed.* *42*, 1508-1511.
81. Ranaghan, K. E., and Mulholland, A. J. (2004) Conformational effects in enzyme catalysis: QM/MM free energy calculations of the 'NAC' contribution in chorismate mutase, *Chem. Commun.*, 1238-1239.
82. Ranaghan, K. E., Ridder, L., Szeftczyk, B., Sokalski, W. A., Hermann, J. C., and Mulholland, A. J. (2004) Transition state stabilization and substrate strain in enzyme catalysis: *ab initio* QM/MM modeling of the chorismate mutase reaction, *Org. Biomol. Chem.* *2*, 968-980.
83. Szeftczyk, B., Mulholland, A. J., Ranaghan, K. E., and Sokalski, W. A. (2004) Differential transition-state stabilization in enzyme catalysis: Quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field, *J. Am. Chem. Soc.* *126*, 16148-16159.

84. Guimarães, C. R. W., Udier-Blagovic, M., Tubert-Brohman, I., and Jorgensen, W. L. (2005) Effects of Arg90 neutralization on the enzyme-catalyzed rearrangement of chorismate to prephenate, *J. Chem. Theory Comput.* 1, 617-625.
85. Claeysens, F., Ranaghan, K. E., Manby, F. R., Harvey, J. N., and Mulholland, A. J. (2005) Multiple high-level QM/MM reaction paths demonstrate transition-state stabilization in chorismate mutase: Correlation of barrier height with transition-state stabilization, *Chem. Commun.* 40, 5068-5070.
86. Crespo, A., Marti, M. A., Estrin, D. A., and Roitberg, A. E. (2005) Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase, *J. Am. Chem. Soc.* 127, 6940-6941.
87. Ishida, T., Fedorov, D. G., and Kitaura, K. (2006) All electron quantum chemical calculation of the entire enzyme system confirms a collective catalytic device in the chorismate mutase reaction, *J. Phys. Chem B* 110, 1457-1463.
88. Hur, S., and Bruice, T. C. (2002) The mechanism of catalysis of the chorismate to prephenate reaction by the *Escherichia coli* mutase enzyme, *Proc. Natl. Acad. Sci. USA* 99, 1176-1181.
89. Hur, S., and Bruice, T. C. (2003) Comparison of formation of reactive conformers (NACs) for the Claisen rearrangement of chorismate to prephenate in water and in the *E. coli* mutase: The efficiency of the enzyme catalysis, *J. Am. Chem. Soc.* 125, 5964-5972.
90. Zhang, X. D., Zhang, X. H., and Bruice, T. C. (2005) A definitive mechanism for chorismate mutase, *Biochemistry* 44, 10443-10448.
91. Khanjin, N. A., Snyder, J. P., and Menger, F. M. (1999) Mechanism of chorismate mutase: Contribution of conformational restriction to catalysis in the Claisen rearrangement, *J. Am. Chem. Soc.* 121, 11831-11846.
92. Bone, R., Silen, J. L., and Agard, D. A. (1989) Structural plasticity broadens the specificity of an engineered protease, *Nature* 339, 191-195.
93. Gerlt, J. A., Babbitt, P. C., and Rayment, I. (2005) Divergent evolution in the enolase superfamily: The interplay of mechanism and specificity, *Arch. Biochem. Biophys.* 433, 59-70.
94. Aharoni, A., Gaidukiov, L., Kheronsky, O., Gould, S. M., Roodveldt, C., and Tawfik, D. S. (2005) The 'evolvability' of promiscuous protein functions, *Nature Genet.* 37, 73-76.
95. Yoshikuni, Y., Ferrin, T. E., and Keasling, J. D. (2006) Designed divergent evolution of enzyme function, *Nature* 440, 1078-1082.

Table C-1. Results for complementation tests, kinetic parameters, and thermal stabilities^a

pos 7	C/NC ^b	$k_{\text{cat}}, \text{s}^{-1}$	$K_{\text{m}}, \mu\text{M}$	$k_{\text{cat}}/K_{\text{m}}, \text{M}^{-1}\text{s}^{-1}$	$T_{\text{m}}, ^\circ\text{C}$
Ala	-				
Arg	-				
Asn	-				
Asp	-				
Cys	+		> 750	1.0×10^4	52.0
Gln	-				
Glu	-				
Gly	-				
His	-				
Ile	+	36 ± 2	300 ± 80	1.2×10^5	65.8
Leu*	+	39 ± 5	300 ± 50	1.3×10^5	61.4 ± 0.6
Lys	-				
Met	+		> 750	4.4×10^4	60.3
Phe	+	48 ± 2	350 ± 30	1.4×10^5	62.3
Pro	-				
Ser	-				
Thr	+		> 750	1.3×10^4	59.2
Trp	-				
Tyr	-				
Val	+	36 ± 3	700 ± 100	5×10^4	61.9

pos 32	C/NC ^b	$k_{\text{cat}}, \text{s}^{-1}$	$K_{\text{m}}, \mu\text{M}$	$k_{\text{cat}}/K_{\text{m}}, \text{M}^{-1}\text{s}^{-1}$	$T_{\text{m}}, ^\circ\text{C}$
Ala*	+	39 ± 5	300 ± 50	1.3×10^5	61.4 ± 0.6
Arg	-				
Asn	-				
Asp	-				
Cys	+				aggr
Gln	-				
Glu	-				
Gly	+				aggr
His	-				
Ile	+				aggr
Leu	-				
Lys	-				
Met	-				
Phe	-				
Pro	-				
Ser	+	45 ± 6	220 ± 30	2.1×10^5	61.7
Thr	+	61 ± 4	380 ± 70	1.6×10^5	62.8
Trp	-				
Tyr	-				
Val	+				aggr

pos 35	C/NC ^b	$k_{\text{cat}}, \text{s}^{-1}$	$K_{\text{m}}, \mu\text{M}$	$k_{\text{cat}}/K_{\text{m}}, \text{M}^{-1}\text{s}^{-1}$	$T_{\text{m}}, ^\circ\text{C}$
Ala	+	20 ± 2	520 ± 90	3.9×10^4	61.2
Arg	-				
Asn	-				
Asp	-				
Cys	+	18.6 ± 0.6	360 ± 40	5.2×10^4	61.8
Gln	-				
Glu	-				
Gly	-				
His	-				
Ile	+	65 ± 9	440 ± 80	1.5×10^5	66.6
Leu	+		> 750	1.3×10^4	66.7
Lys	-				
Met	+		> 750	5.2×10^4	68.9
Phe	-				
Pro	-				
Ser	-				
Thr	+		> 750	2.0×10^4	56.5
Trp	-				
Tyr	-				
Val*	+	39 ± 5	300 ± 50	1.3×10^5	61.4 ± 0.6

pos 48	C/NC ^b	$k_{\text{cat}}, \text{s}^{-1}$	$K_{\text{m}}, \mu\text{M}$	$k_{\text{cat}}/K_{\text{m}}, \text{M}^{-1}\text{s}^{-1}$	$T_{\text{m}}, ^\circ\text{C}$
Ala	-				
Arg	-				
Asn	+		> 750	7.9×10^3	56.5
Asp*	+	39 ± 5	300 ± 50	1.3×10^5	61.4 ± 0.6
Cys	+		> 750	1.8×10^4	53.8
Gln	+		> 750	9.8×10^3	55.5
Glu	-				
Gly	-				
His	-				
Ile	-				
Leu	+		> 750	7.1×10^3	57.7
Lys	-				
Met	-				
Phe	-				
Pro	-				
Ser	-				
Thr	-				
Trp	-				
Tyr	-				
Val	-				

pos 81	C/NC ^b	k_{cat} , s ⁻¹	K_m , μM	k_{cat}/K_m , M ⁻¹ s ⁻¹	T_m , °C
Ala	-				
Arg	-				
Asn	-				
Asp	-				
Cys	+				aggr
Gln	-				
Glu	-				
Gly	-				
His	-				
Ile*	+	39 ± 5	300 ± 50	1.3 x 10⁵	61.4 ± 0.6
Leu	+		> 750	5.5 x 10 ⁴	57.7
Lys	-				
Met	+	27 ± 2	270 ± 20	1.0 x 10 ⁵	64.6
Phe	+		> 750	8.7 x 10 ³	59.0
Pro	-				
Ser	-				
Thr	-				
Trp	+				aggr
Tyr	-				
Val	+				aggr

pos 85	C/NC ^b	k_{cat} , s ⁻¹	K_m , μM	k_{cat}/K_m , M ⁻¹ s ⁻¹	T_m , °C
Ala	+		> 750	1.2 x 10 ⁴	62.7
Arg	+		> 750	4.2 x 10 ⁴	66.9
Asn	+		> 750	1.2 x 10 ⁴	66.8
Asp	-				
Cys	+	32 ± 4	570 ± 90	5.4 x 10 ⁴	54.7
Gln	-				
Glu	-				
Gly	-				
His	-				
Ile	+	55 ± 1	550 ± 30	1.0 x 10 ⁵	62.9
Leu	+		> 750	4.3 x 10 ⁴	62.2
Lys	+		> 750	1.3 x 10 ⁴	67.1
Met	+		> 750	1.5 x 10 ⁴	65.6
Phe	+		> 750	2.1 x 10 ⁴	61.7
Pro	-				
Ser	-				
Thr	+		> 750	2.2 x 10 ⁴	62.1
Trp	+				aggr
Tyr	+		> 750	2.6 x 10 ⁴	65.7
Val*	+	39 ± 5	300 ± 50	1.3 x 10⁵	61.4 ± 0.6

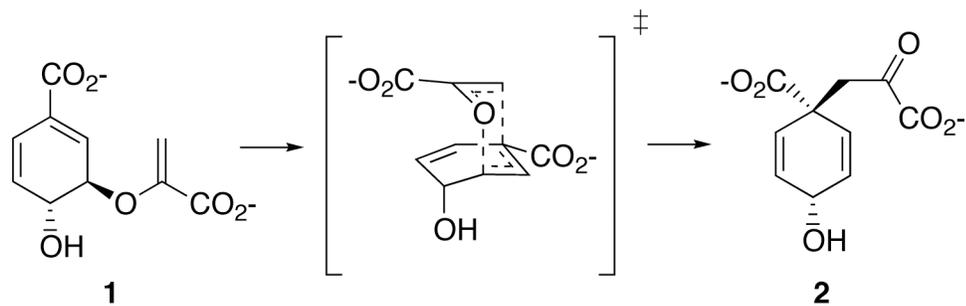
^a Assays were performed at 37 °C and pH 7.8; see materials and methods. Wild-type residues are indicated with bold type and asterisks, and wild-type values differ by less than 5% from published values for the same construct and conditions.^{23,43} *aggr*, protein precipitated or aggregated during purification. Where uncertainties are indicated, the means and standard deviations from at least three trials are reported.

^b Variants complementing the deletion strain are indicated by +, non-complementing by -

Table C-2. Structural alignment of solved AroQ mutase structures and BLAST results^a

<i>Escherichia coli</i> ; 1ECM	Leu 7	Ala 32 [STV]	Val 35 [I]	Asp 48 [AEKQSTV]	Ile 81 [L]	Val 85
<i>Saccharomyces cerevisiae</i> ; 4CSM	Leu 12	Gly 161	Val 164	Asn 194 [DYFKS]	Ile 239 [M]	Lys 243
<i>Mycobacterium tuberculosis</i> ; 2FP2	Leu 130	Ala 53	Val 56 [IM]	Asp 69	Ile 102	Glu 106 [KV]
<i>Pyrococcus furiosus</i> ; 1YBZ	Leu 4	Ala 29 [CS]	Ile 32 [M]	Asp 45 [V]	Leu 67 [V]	Lys 71 [R]
<i>Thermus thermophilus</i> ; 2D8D	Ile 5 [LM]	Val 30 [A]	Ile 33 [V]	Asp 46	Phe 78	Leu 82 [T]

^a Structures from the pdb codes indicated were aligned over the transition state analog if present, and otherwise over all residues. For each pdb protein sequence, protein-protein BLAST⁶⁴ was performed using the RefSeq database.⁶⁵ Variations of the position are reported in brackets for matches with bitcores of 85 or greater (60 for 1YBZ); these scores corresponded to sequences with ~40% sequence identity or greater.

Scheme C-1. Claisen rearrangement of chorismate (1) to prephenate (2).

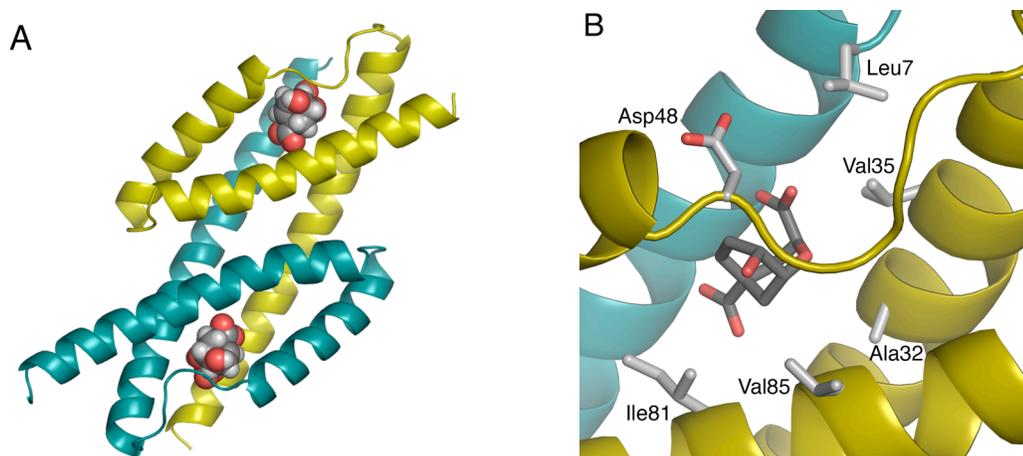


Figure C-1: Truncated, monofunctional chorismate mutase domain from *Escherichia coli* chorismate mutase-prephenate dehydratase in complex with a transition state analog (dark gray carbons), from pdb code 1ECM. (A) Overall fold, showing the two chains of the symmetric homodimer in teal and gold. The bound transition state analog is shown as a CPK model. (B) Active site residues investigated in this study are shown with light gray carbon atoms.

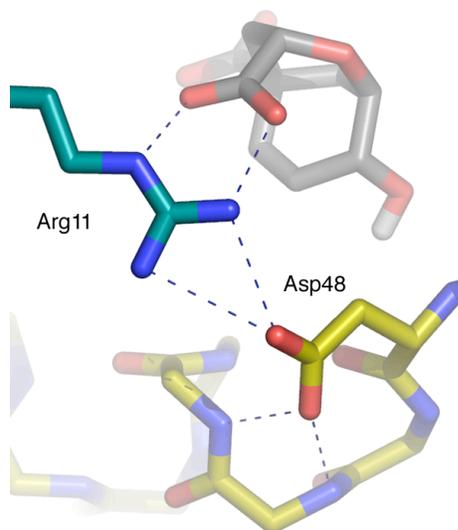


Figure C-2: Interactions of Asp48 with the N-terminus of the second helix and with Arg11. The carbon atoms of the bound transition state analog are in gray, while those of the side chains are colored in teal and gold to indicate their association with the corresponding EcCM chain. Many side chains, protons, and other elements are omitted for clarity.