Applications of Computational Protein Design

Thesis by

Jessica Mao

In Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2006

(Defended January 24, 2006)

Acknowledgements

Reflecting back on my graduate school experiences, I realize how many people have contributed to my growth, both on a professional level and on a personal level. These past five years have taught me the rigor of academic research, but also allowed me the freedom to explore areas beyond science.

I would like to thank, first and foremost, Dr. Stephen L. Mayo for allowing me to become a part of his group. I felt welcomed from the very first day. His hands-off approach was a little difficult to get used to at first, but it has given me the freedom to develop independently. While I have not always found the quickest way, he has always been patient and understanding, ready with guidance when I need it. I greatly admire his skill to see to the core of the problems, and his inexhaustible attention to details.

Joining the Mayo lab meant I had to learn a lot of new subjects. Thanks to Shannon Marshall for showing me the basics of molecular biology, PCR, circular dichroism, and ORBIT. Her photographic memory and ability to recall what seemed like every paper she read was uncanny. As my mentor, she and I worked on the cation-π interaction project together, and I learned from her not only proper sterile techniques, but also how to plan out a research project.

Daniel Bolon was a great mentor as well. He taught me everything I know about enzyme design and gave me lots of advice on choosing projects, which have turned out to be quite accurate.

I would also like to thank Premal Shah, my first neighbor and friend in lab. He was fun to talk to, and answered many of my questions about ORBIT and molecular biology. He and Possu Huang were superb biochemists, and could always trouble shoot my PCRs. Possu was also responsible for my becoming a Mac convert. Thanks, Possu, for showing me the way out of frustrating software. Geofferey Hom is perhaps the most social, purest, and most principled person I know, even though he may not think so. I would also like to thank Oscar Alvizo and Heidi Privett, for sharing a lab bay with me. They were always willing to listen to my experimental woes and offer suggestions.

I would like to thank my collaborators. Eun Jung Choi and Amanda L. Cashin. Not only were they great friends to me, they were wonderful collaborators. They motivated me to try again and again. I enjoyed working with them very much. I am also grateful for the ORBIT journal club where I learned the intricacies of protein design. The Mayo lab has a steep learning curve in the beginning and the journal club discussions with Eric Zollars, Kyle Lassila, Oscar Alvizo, Eun Jung Choi etc. made the learning much less painful.

Deepshikha Datta, Shira Jacobson, Chris Voigt, Pavel Strop, Cathy Sarisky, J. J. Plecs, Julia Shifman, John Love (aka Dr. Love), and Scott Ross were in the lab when I joined, and they have all taught me valuable things about my projects, the lab and Caltech in general. Christina Vizcarra, Ben Allan, Heidi Privett, Jennifer Keeffe, Mary Devlin, Peter Oelschlaeger, Karin Crowhurst, Tom Treynor, and Alex Perryman were all valuable additions to the lab and I am very

glad to have overlapped with some of the most intelligent people I know and probably will ever meet.

Of course, I could not discuss the lab without mentioning the three guardian angels: Cynthia Carlson, Rhonda Digiusto, and Marie Ary. Cynthia Carlson is the most efficient person I know. Her cheerfulness and spirit are an inspiration to me, and I hope to one day have as many interesting life stories to tell as she has. Rhonda makes the lab run smoothly, and I can not even begin to count how many hours she has saved me by being so good at her job. Cynthia and Rhonda always remember our birthdays and make the lab a welcoming place to be. Marie has helped me tremendously with my scientific writing, going over very rough first drafts with no complaints. I hope one day to write as well as she does.

I would also like to thank my undergraduate advisor, Daniel Raleigh, for teaching me about proteins and alerting me to the interesting research in the Mayo lab.

Besides people who have contributed scientifically, I would also like to thank those who have helped me deal with the difficulties of research and making graduate life enjoyable. I would like to thank Anand Vadehra, who has always believed in my abilities and was my biggest supporter. No matter what I needed, he was always there to help. He has taught me many things, including charge transfer with DNA, and more importantly, to enjoy the moment. Amanda Cashin's optimism is infectious. I could not imagine going through graduate

school without her. Thanks for those long talks and shopping trips, and we will always have Costa Rica. Other friends who have helped me get through Caltech with fond memories are Pete Choi, Xin Qi, Christie Morrill, the ‘dancing girls”, Angie Mah, Lisa Welp, and all those friends on the east coast who prompted me to action every so often with “did you graduate yet?”

Caltech has allowed me to explore many areas beyond science. I would like to thank the Caltech Biotech Club and everyone I have worked with on the committee for teaching me new skills in organization. Deepshikha Datta had the brilliant idea of starting it, and I am grateful to have been a part of it from the beginning. It has allowed me to experience Caltech in a whole new way. Other campus organizations that have enriched my life are: Caltech Y, Alpine Club, Women’s Center, Surfing and Windsurfing Club, GSC intramural volleyball and softball, and Women’s Ultimate Frisbee Team. Thank you for making my life more multidimensional.

Lastly, I would like to thank my parents, for none of this would have been possible had they not instilled in me the importance of learning and pushed me to do better all the time. They planned very early on to move to the United States so that my sister and I could get a good education, and I am very grateful for their sacrifices. Thank you for your constant love and support.

Abstract

Computational protein design determines the amino acid sequence(s) that will adopt a desired fold. It allows the sampling of a large sequence space in a short amount of time compared to experimental methods. Computational protein design tests our understanding of the physical basis of a protein's structure and function, and over the past decade, has proven to be an effective tool.

We report the diverse applications of computational protein design with ORBIT (Optimization of Rotamers by Iterative Techniques). We successfully utilized ORBIT to construct a reagentless biosensor for nonpolar ligands on the maize non-specific lipid transfer protein, by first removing native disulfide bridges. We identified an important residue position capable of modulating the agonist specificity of the mouse muscle nicotinic acetylcholine receptor (nAChR) for its agonists: acetylcholine, nicotine, and epibatidine. Our efforts on enzyme design produced a lysozyme mutant with ester hydrolysis activity, while progress was made toward the design of a novel aldolase.

Computational protein design has proven to be a powerful tool for the development of novel and improved proteins. As we gain a better understanding of proteins and their functions, protein design will find many more exciting applications.

Table of Contents

Acknowledgements iii

Abstract vii

Table of Contents viii

List of Figures xiii

List of Tables xvi

Abbreviations xvii

Chapter 1. Introduction

Protein Design 2

Computational Protein Design with ORBIT 2

Applications of Computational Protein Design 4

References 7

Chapter 2. Removal of Disulfide Bridges by Computational Protein Design

Introduction 11

Materials and Methods 12

Computational Protein Design 12

Protein Expression and Purification 14

Circular Dichroism Spectroscopy 15

Results and Discussion 15

mLTP Designs 15

Experimental Validation..... 16

Future Direction 18

References 19

Chapter 3. Engineering a Reagentless Biosensor for Nonpolar Ligands

Introduction..... 28

Materials and Methods..... 29

Protein Expression, Purification and Acrylodan Labeling 29

Circular Dichroism..... 31

Fluorescence Emission Scan and Ligand Binding Assay..... 31

Curve Fitting..... 32

Results..... 32

Protein-Acrylodan Conjugates..... 32

Fluorescence of Protein-Acrylodan Conjugates..... 33

Ligand Binding Assays..... 34

Discussion 34

References 36

Chapter 4. Designed Enzymes for Ester Hydrolysis

Introduction..... 46

Materials and Methods..... 48

Protein Design with ORBIT 48

Protein Expression and Purification 49

Circular Dichroism 50

Protein Activity Assay 50

Results 50

Thioredoxin Mutants 50

T4 Lysozyme Designs 51

Discussion 52

References 54

Chapter 5. Enzyme Design: Toward the Computational Design of a Novel Aldolase

Enzyme Design 63

"Compute and Build" 64

Aldolases 65

Target Reaction 67

Protein Scaffold 68

Testing of Active Site Scan on 33F12 69

Hapten-like Rotamer 70

HESR 72

Enzyme Design on TIM 75

Active Site Scan on "Open" Conformation 76

Active Site Scan on “Almost-Closed” Conformation 77

pK_a Calculations 78

Design on Active Site of TIM 79

GBIAS 81

Enzyme Design on Ribose Binding Protein 82

Experimental Results 84

Discussion 86

Reactive Lysines 87

Buried Lysines in Literature 87

Tenth Fibronectin Type III Domain 88

mLTP (Non-specific Lipid-Transfer Protein from Maize) 89

Future Directions 90

References 91

Chapter 6. Double Mutant Cycle Study of Cation-π Interaction

Introduction 126

Materials and Methods 128

Computational Modeling 128

Protein Expression and Purification 130

Circular Dichroism (CD) 131

Double Mutant Cycle Analysis 132

Results and Discussion 132

References 135

Chapter 7. Modulating nAChR Agonist Specificity by Computational Protein Design

Introduction 144

Material and Methods 146

Computational Protein Design with ORBIT 146

Mutagenesis and Channel Expression 148

Electrophysiology 148

Results and Discussion 149

Computational Design 149

Mutagenesis 150

Nicotine Specificity Enhanced by 57R Mutation 151

Conclusions and Future Directions 153

References 155

List of Figures

Figure 2-1. Ribbon diagram of mLTP and the designed variants of each disulfide 23

Figure 2-2. Wavelength scans of mLTP and designed variants 24

Figure 2-3. Thermal denaturations of mLTP and designed variants 25

Figure 3-1. Ribbon representation of non-specific lipid-transfer protein from maize (mLTP) 38

Figure 3-2. Acrylodan and its conjugation site on mLTP C52A 39

Figure 3-3. Circular dichroism wavelength scans of the four protein-acrylodan conjugates 40

Figure 3-4. Fluoresence emission scans of mLTP-acrylodan conjugates 41

Figure 3-5. Titration of C52A/C4-Acrylodan with palmitate monitored by fluorescence emission 42

Figure 3-6. Thermal denaturations of C52A/4C-A monitored by CD 43

Figure 3-7. Space-filling representation of mLTP C52A 44

Figure 4-1. Ribbon model of PZD2 and structure of His-substrate high energy state rotamer 56

Figure 4-2. Sequence comparison of wild-type T4 lysozyme with 134, Rbias1.0 and Rbias2.5 58

Figure 4-3. Lysozyme 134 highlighting the essential residues for catalysis 59

Figure 4-4. Circular dichroism characterization of lysozyme 134 60

Figure 5-1. A generalized aldol reaction 96

Figure 5-2. The enamine mechanism of catalytic antibody aldolases and natural class I aldolases 97

Figure 5-3. Fab' 33F12 binding site 98

Figure 5-4. The target aldol addition between acetone and benzaldehyde 99

Figure 5-5. Structure of Fab' 33F12 101

Figure 5-6. Hapten-like rotamers for active site scan on 33F12 102

Figure 5-7. High-energy state rotamer with varied dihedral angles labeled 104

Figure 5-8. Superposition of 1AXT with the modeled protein 106

Figure 5-9. Ribbon diagram and Cα trace of triosephosphate isomerase 107

Figure 5-10. Superposition of backbone atoms of "open" and "almost-closed" conformations of TIM 110

Figure 5-11. KPY rotamer and the HESR benzal rotamer 114

Figure 5-12. Using GBIAS to retain crystallographic hydrogen bonds in KDPG aldolase 115

Figure 5-13. Ribbon diagram of ribose binding protein in open and closed conformations 116

Figure 5-14. HESR in the binding pocket of RBP 117

Figure 5-15. Modeled active site on RBP for aldol reaction 118

Figure 5-16. CD wavelength scan of RBP and Mutants..... 119

Figure 5-17. Catalytic assay of 38C2 120

Figure 5-18. Catalytic assay of RBP and R141K..... 121

Figure 5-19. Ribbon diagram of tenth fibronectin type III domain..... 122

Figure 5-20. Ribbon diagram of mLTP 123

Figure 5-21. Circular dichroism spectroscopy of mLTP and mutants 124

Figure 6-1 Schematic of the cation-π interaction..... 138

Figure 6-2 Ribbon diagram of engrailed homeodomain 139

Figure 6-3 Modelled Arg9-Trp13 in engrailed homeodomain 140

Figure 6-4 Urea denaturation of homeodomain variants..... 141

Figure 7-1. Sequence alignment of AChBP with nAChR subunits from mouse muscle 158

Figure 7-2. Structures of nAChR agonists, acetylcholine, nicotine, and epibatidine..... 159

Figure 7-3. Predicted mutations from computational design of AChBP..... 160

Figure 7-4. Electrophysiology data..... 161

List of Tables

Table 2-1. Apparent T_ms of mLTP and designed variants 26

Table 4-1. Kinetic parameters of PZD2 and variants for PNPA hydrolysis .. 57

Table 4-2. Kinetic parameters of lysozyme 134 compared to PZD2 for PNPA hydrolysis..... 61

Table 5-1. Catalytic parameters of proline and catalytic antibodies..... 100

Table 5-2. Top 10 results from active site scan of the Fab' antigen-binding region of 33F12 with hapten-like rotamer 103

Table 5-3. Top 10 results from active site scan of the Fab' antigen-binding region of 33F12 with HESR 105

Table 5-4. Top 10 results from active site scan of the open conformation of TIM with hapten-like rotamers 108

Table 5-5. Top 10 results from active site scan of the open conformation of TIM with HESR..... 109

Table 5-6. Top 10 results from active site scan of the almost-closed conformation of TIM with HESR 111

Table 5-7. Results of MCCE pK calculations on test proteins 112

Table 5-8. Results of modeling the HESR at Lys 13, the natural catalytic residue 113

Table 6-1. Thermodynamic parameters of engrailed homeodomain variants from urea denaturation 142

Table 7-1. Mutation enhancing nicotine specificity..... 162

Abbreviations

ORBIT	optimization of rotamers by iterative techniques
GMEC	global minimum energy conformation
DEE	dead-end elimination
LB	Luria broth
HPLC	high performance liquid chromatography
CD	circular dichroism
HES	high energy state
HESR	high energy state rotamer
PNPA	*p*-nitrophenyl acetate
PNP	*p*-nitrophenol
TIM	triosephosphate isomerase
RBP	ribose binding protein
mLTP	non-specific lipid-transfer protein from maize
Ac	acrylodan
PDB	protein data bank
K_d	dissociation constant
K_m	Michaelis constant
UV	ultra-violet
NMR	nuclear magnetic resonance
E. coli	*Escherichia coli*

nAChR	nicotinic acetylcholine receptor
ACh	acetylcholine
Nic	nicotine
Epi	epibatidine

Chapter 1

Introduction

Protein Design

While it remains nontrivial to predict the three-dimensional structure a linear sequence of amino acids will adopt in its native state, much progress has been made in the field of protein folding due to major enhancements in computing power and the development of new algorithms. The inverse of the protein folding problem, the protein design problem, has benefited from the same advances. Protein design determines the amino acid sequence(s) that will adopt a desired fold. Historically, proteins have been designed by applying rules observed from natural proteins, or by employing selection and evolution experiments in which a particular function is used to separate the desired sequences from the pool of largely undesirable sequences. Computational methods have also been used to model proteins and obtain an optimal sequence, the figurative “needle in the haystack.” Computational protein design has the advantage of sampling much larger sequence space in a shorter amount of time compared to experimental methods. Lastly, the computational approach tests our understanding of the physical basis of a protein’s structure and function, and over the past decade, has proven to be an effective tool in protein design.

Computational Protein Design with ORBIT

Computational protein design has three basic requirements: knowledge of the forces that stabilize the folded state of a protein relative to the unfolded state, a forcefield that accurately captures these interactions, and an efficient

optimization algorithm. ORBIT (Optimization of Rotamers by Iterative Techniques) is a protein design software package developed by the Mayo lab. It takes as input a high-resolution structure of the desired fold, and outputs the amino acid sequence(s) that are predicted to adopt the fold. If available, high-resolution crystal structures of proteins are often used for design calculations, although NMR structures, homology models, and even novel folds can be used. A design calculation is then defined to specify the residue positions and residue types to be sampled. A library of discrete amino acid conformations, or rotamers, are then modeled at each position and pair-wise interaction energies are calculated using an energy function based on the atom-based DREIDING forcefield.[1] The forcefield includes terms for van der Waals interactions, hydrogen bonds, electrostatics, and the interaction of the amino acids with water.[2-4] Combinatorial optimization algorithms such as Monte Carlo and algorithms based on the dead-end elimination theorem are then used to determine the global minimum energy conformation (GMEC) or sequences near the GMEC.[5-8] The sequences can be experimentally tested to determine the accuracy of the design calculation. Protein stability and function require a delicate balance of contributing interactions; the closer the energy function gets toward achieving the proper balance, the higher the probability the sequence will adopt the desired fold and function. By utilizing the “design cycle” that iterates from theory to computation to experiment, improvements in the energy function can be continually made, leading to better designed proteins.

The Mayo lab has successfully utilized the design cycle to improve the energy function, and developments in combinatorial optimization algorithms allowed ever-larger design calculations. Consequently, both novel and improved proteins have been designed. The β1 domain of protein G and engrailed homeodomain from *Drosophila* have been designed with greatly increased thermostability compared to their wild-type sequences.[9, 10] Full sequence designs have generated a 28-residue zinc finger that does not require zinc to maintain its three-dimensional fold[3] and an engrailed homeodomain variant that is 80% different from the wild-type sequence yet still retains its fold.[11]

Applications of Computational Protein Design

Generating proteins with increased stability is one application of protein design. Other potential applications include improving the catalysis of existing enzymes, modifying or generating binding specificity for ligands, substrates, peptides, and other proteins, and generating novel proteins and enzymes. New methods continue to be created for protein design to support an ever-wider range of applications. My work has been on the application of computational protein design by ORBIT.

In chapters 2 and 3, we used protein design to remove disulfide bridges from maize non-specific lipid-transfer protein (mLTP). By coupling the resulting conformational flexibility with an environment sensitive fluorescent probe, we generated a reagentless biosensor for nonpolar ligands.

Chapter 4 is an extension of previous work by Bolon and Mayo[12] that generated the first computationally designed enzyme, PZD2, an ester hydrolase. We first probed the effect of four anionic residues (near the catalytic site) on the catalytic rate of PZD2. Separately, we engineered ester hydrolysis activity into T4 lysozyme, demonstrating the general applicability of the "compute and build" method utilized for PZD2.

The same method was applied to generate an enzyme to catalyze the aldol reaction, a carbon-carbon bond-making reaction that is more difficult to catalyze than ester hydrolysis. Chapter 5 details the efforts toward the design of a novel aldolase.

Chapter 6 describes the double mutant cycle study of a cation-π interaction to ascertain its interaction energy. We used protein design to determine the optimal sites for incorporation of the amino acid pair.

In chapter 7, we utilized computational protein design to identify a mutation that modulated the agonist specificity of the nicotinic acetylcholine receptor (nAchR) for its agonists, acetylcholine, nicotine, and epibatidine.

We have shown diverse applications of computational protein design. From the first notable success in 1997, the field has advanced quickly. Other recent advances in protein design include the full sequence design of a protein with a novel fold[13] and dramatic increases in binding specificity of proteins.[14, 15] Hellinga and co-workers achieved nanomolar binding affinity of a designed protein for its non-biological ligands[16] and built a family of biosensors for small

polar ligands from the same family of proteins.[17-19] They also used a combination of protein design and directed evolution experiments to generate triosephosphate isomerase (TIM) activity in ribose binding protein.[20]

Computational protein design has proven to be a powerful tool. It has demonstrated its effectiveness in generating novel and improved proteins. As we gain a better understanding of proteins and their functions, protein design will find many more exciting applications.

References

1. Mayo, S. L., Olafson, B. D. & Goddard III, W. A. DREIDING: A generic force field for molecular simulations. *Journal of Physical Chemistry* 94, 8897-8909 (1990).
2. Gordon, D. B., Marshall, S. A. & Mayo, S. L. Energy functions for protein design. *Curr Opin Struct Biol* 9, 509-13 (1999).
3. Dahiyat, B. I. & Mayo, S. L. Probing the role of packing specificity in protein design. *Proceedings of the Natational Academy of Sciences of the United States of America* 94, 10172-7. (1997).
4. Street, A. G. & Mayo, S. L. Pairwise calculation of protein solvent - accessible surface areas. *Folding & Design* 3, 253-258 (1998).
5. Gordon, D. B. & Mayo, S. L. Radical performance enhancements for combinatorial optimization algorithms based on the dead-end elimination theorem. *J Comp Chem* 19, 1505-1514 (1998).
6. Gordon, D. B. & Mayo, S. L. Branch-and-Terminate: a combinatorial optimization algorithm for protein design. *Structure Fold Des* 7, 1089-1098 (1999).
7. Pierce, N. A., Spriet, J. A., Desmet, J. & Mayo, S. L. Conformational splitting: a more powerful criterion for dead-end elimination. *J Comp Chem* 21, 999-1009 (2000).

8. Voigt, C. A., Gordon, D. B. & Mayo, S. L. Trading accuracy for speed: a quantitative comparison of search algorithms in protein sequence design. *J Mol Biol* 299, 789-803 (2000).

9. Malakauskas, S. M. & Mayo, S. L. Design, structure and stability of a hyperthermophilic protein variant. *Nature Struct Biol* 5, 470-475 (1998).

10. Marshall, S. A. & Mayo, S. L. Achieving stability and conformational specificity in designed proteins via binary patterning. *J Mol Biol* 305, 619-31. (2001).

11. Shah, P. S. (California Institute of Technology, Pasadena, CA, 2005).

12. Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc Natl Acad Sci U S A* 98, 14274-9. (2001).

13. Kuhlman, B. et al. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* 302, 1364-1368 (2003).

14. Kortemme, T. et al. Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* 11, 371-9 (2004).

15. Shifman, J. M. & Mayo, S. L. Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc Natl Acad Sci U S A* 100, 13274-9 (2003).

16. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. Computational design of receptor and sensor proteins with novel functions. *Nature* 423, 185-90 (2003).

17. Marvin, J. S. & Hellinga, H. W. Engineering Biosensors by Introducing Fluorescent Allosteric Signal Transducers: Construction of a Novel Glucose Sensor. *J. Am. Chem. Soc.* 120, 7-11 (1998).

18. De Lorimier, R. M. et al. Construction of a fluorescent biosensor family. *Protein Sci* 11, 2655-2675 (2002).

19. Marvin, J. S. et al. The rational design of allosteric interactions in a monomeric protein and its applications to the construction†of†biosensors. *PNAS* 94, 4366-4371 (1997).

20. Dwyer, M. A., Looger, L. L. & Hellinga, H. W. Computational design of a biologically active enzyme. *Science* 304, 1967-71 (2004).

Chapter 2

Removal of Disulfide Bridges by Computational Protein Design

Adapted from manuscript in preparation by Jessica Mao, Eun Jung Choi and Stephen L. Mayo. To be submitted.

Introduction

One of the most common posttranslational modifications to extracellular proteins is the disulfide bridge, the covalent bond between two cysteine residues. Disulfide bridges are present in various protein classes and are highly conserved among proteins of related structure and function.[1, 2] They perform multiple functions in proteins. They add stability to the folded protein[3-5] and are important for protein structure and function. Reduction of the disulfide bridges in some enzymes leads to inactivation.[6, 7]

Two general methods have been used to study the effect of disulfide bridges on proteins: the removal of native disulfide bonds and the insertion of novel ones. Protein engineering studies to enhance protein stability by adding disulfide bridges have had mixed results.[8] Addition of individual disulfides in T4 lysozyme resulted in various mutants with raised or lowered T_m, a measure of protein stability.[9, 10] Removal of disulfide bridges led to severely destabilized Conotoxin[11] and produced RNase A mutants with lowered stability and activity.[12, 13]

Typically, mutations to remove disulfide bridges have substituted Cys with Ala, Ser or Thr, depending on the solvent accessibility of the native Cys. However, these mutations do not consider the protein background of the disulfide bridge. For example, Cys to Ala mutations could destabilize the native state by creating cavities. Computational protein design could allow us to compensate for the loss of stability by substituting stabilizing non-covalent interactions. The

protein design software suite ORBIT (Optimization of Rotamers by Iterative Techniques)[14] has been very successful in designing stable proteins,[15, 16] and can predict mutations that would stabilize the native state without the disulfide bridge.

In this paper, we utilized ORBIT to computationally design out disulfide bridges in the non-specific lipid-transfer protein (ns-LTP) from maize (mLTP). mLTP is a 93-residue basic α-helical protein containing four disulfide bridges that are strictly conserved in the plant ns-LTP family.[17-19] The ns-LTPs bind various polar lipids, fatty acids, acyl-coenzyme A,[18] and they are proposed to defend the plant against bacterial and fungal pathogens.[20] The high resolution crystal structure of mLTP[17] makes it a good candidate for computational protein design. Our goal was to computationally remove the disulfide bridges and experimentally determine the effects on mLTP's stability and ligand-binding activity.

Materials and Methods

Computational Protein Design

The crystal structure of mLTP with palmitate (PDB ID 1MZM) was briefly energy minimized and its residues were classified as surface, boundary, or core based on solvent accessibility.[21] Each of the four disulfide bridges were individually reduced by deletion of the S-S bond and addition of hydrogens. The corresponding structures were used in designs for the respective disulfide bridge. The ORBIT protein design suite uses an energy function based on the DREIDING force field,[22] which includes a Lennard-Jones 12-6 potential with all

van der Waals radii scaled by 0.9,[23] hydrogen bonding and electrostatic terms [24] and a solvation potential.

Both solvent-accessible surface area-based solvation[25] and the implicit solvation model developed by Lazaridis and Karplus[26] were tried, but better results were obtained with the Lazaridis-Karplus model, and it was used in all final designs. Polar burial energy was scaled by 0.6, and rotamer probability was scaled by 0.3, as suggested by Oscar Alvizo from fixed composition work with Engrailed homeodomain (unpublished data). Parameters from the Charmm19 force field were used. An algorithm based on the dead-end elimination theorem (DEE) was used to obtain the global minimum energy amino acid sequence and conformation (GMEC).[27]

For each design, non-Pro, non-Gly residues within 4 Å of the two reduced Cys were included as the 1st shell of residues and were designed; that is, their amino acid identities and conformations were optimized by the algorithm. Residues within 4 Å of the designed residues were considered the 2nd shell; these residues were floated, that is, their conformations were allowed to change, but their amino acid identities were held fixed. Finally, the remaining residues were treated as fixed. Based on the results of these design calculations, further restricted designs were carried out where only modeled positions making stabilizing interactions were included.

Protein Expression and Purification

The *Escherichia coli* expression optimized gene encoding the mLTP amino acid sequence was synthesized and ligated into the pET15b vector (Stratagene) by Blue Heron Biotechnology (www.blueheronbio.com). The pET15b vector includes an N-terminal His-tag. Inverse PCR mutagenesis was used to construct five variants, C4H/C52A/N55E, C4Q/C52A/N55S, C14A/C29S, C30A/C75A, and C50A/C89E. The proteins were expressed in BL21(DE3) Gold cells (Stratagene) at 37 °C after induction with IPTG (isopropyl-beta-D-thiogalactopyranoside). The proteins expressed in the soluble fraction. Cells were resuspended in lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole, pH 8.0) and lysed by passing through the Emulsiflex at 15,000 psi, and the soluble fraction was obtained by centrifuge at 20,000*g* for 30 minutes. Protein purification was a two step process. First, the soluble fraction of the cell lysate was loaded onto a Ni-NTA column and eluted with elution buffer (lysis buffer with 400 mM imidazole). The elutions were further purified by gel filtration with phosphate buffer (50 mM sodium phosphate, 150 mM sodium chloride, pH 7.5). Purified proteins were verified by SDS-Page and MALDI-TOF to be of sufficient purity and corresponded to the oxidized form of the proteins. The N-terminal His-tags are present, without the N-terminal Met, as was confirmed by trypsin digests. Protein concentration was determined using the BCA assay (Pierce) with BSA as the standard.

Circular Dichroism

Circular dichroism (CD) data were obtained on an Aviv 62A DS spectropolarimeter equipped with a thermoelectric cell-holder. Wavelength scans and thermal denaturation data were obtained from samples containing 50 µM protein. For wavelength scans, data were collected every 1 nm from 200 to 250 nm with averaging time of 5 seconds. For thermal studies, data were collected every 2 °C from 1 °C to 99 °C using an equilibration time of 120 seconds and an averaging time of 30 seconds. As the thermal denaturations were not reversible, we could not fit the data to a two-state transition. The apparent T_ms were obtained from the inflection point of the data. For thermal denaturations of protein with palmitate, 150 µM palmitate was added to 50 µM protein from stock solution of >30 mM palmitate in ethanol (Sigma Aldrich).

Results and Discussion

mLTP Designs

mLTP contains four disulfide bridges: C4-C52, C14-C29, C30-C75, and C50-C89, and we used the ORBIT protein design suite to design variants with the removal of each disulfide bridge. Calculations were evaluated and five variants were chosen, C4H/C52A/N55E, C4Q/C52A/N55S, C14A/C29S, C30A/C75A, and C50A/C89E (Figure 2-1). For disulfide bridge C4-C52, the disulfide anchors two helices to each other, with C52 more buried than C4. In the final designs C4H/C52A/N55E and C4Q/C52A/N55S, the disulfide bridge is lost but residue 4

and 55 form an interhelical hydrogen bond, 4H-55E and 4Q-55S, with heavy atom distances of 2.8 Å. C14A/C29S gains a hydrogen bond between S29 and S26. For C30-C75, nonpolar residues surround the buried disulfide, and both residues are mutated to Ala. C50-C89 anchors the C-terminal loop to helix 3. The mutation of C89E breaks the disulfide bridge but adds in hydrogen bonds with R47, S90 and K54, and C50 is mutated to Ala.

Experimental Validation

The circular dichroism wavelength scans of mLTP and the variants (Figure 2-2) show three of the five variants (C4H/C52A/N55E, C4Q/C52A/N55S and C50A/C89E) are folded like the wild-type protein, with minimums at 208nm and 222nm, characteristic of helical proteins. C14A/C29S and C30A/C75A are not folded properly, with wavelength scans resembling those of ns-LTP with scrambled disulfides.[28] Interestingly, both C14-C29 and C30-C75 are the more buried of the four disulfides, and are in close proximity to each other.

Of the folded proteins, the gel filtration profile looked similar to that of wild-type mLTP, which we verified to be a monomer by analytical ultracentrifugation (data not shown). We determined the thermal stability of the variants in the absence and presence of palmitate and compared it to wild-type mLTP (Figure 2-3). The removal of the disulfide bridge C4-C52 significantly destabilized the protein relative to wild type, lowering the apparent T_ms by as much as 28 °C (Table 2-1). Disruption of C50-C89 led to only 10 °C lower apparent T_m. The

variants are still able to bind palmitate, as thermal denaturations in the presence of palmitate raised the apparent melting temperatures as it does for the wild-type protein.

For the C4-C52 mutants, C4H/C52A/N55E and C4Q/C52A/N55S behaved similarly, as each variant supplied one potential hydrogen bond to replace the S-S covalent bond. Upon binding palmitate, however, there is a much larger gain in stability than is observed for the wild-type protein; the T_ms vary by as much as 20 °C compared to only 8 °C for wild type. The difference in apparent T_ms for the palmitate bound mutants and wild-type is ~18 °C, 10 °C lower than the 28 °C difference observed for unbound protein. A plausible explanation for the observed difference could be a conformational change between the unbound and bound forms. In the unbound form, the disulfide that anchored the two helices to each other is no longer present, making the N-terminal helix more entropic, causing the protein to be less compact and lose stability. But once palmitate is bound, the helix is brought back to desolvate the palmitate and returns to its compact globular shape.

It is interesting that C50A/C89E is ~20 °C more stable than the C4-C52 variants. The disulfide C50-C89 anchors the long C-terminal loop to helix 3. Disruption of this disulfide only lowered the T_m by 10 °C. This could be due to the three introduced hydrogen bonds that were a direct result of the C89E mutation. The stability gained by palmitate binding only raises the T_m by 6 °C, similar to the 8 °C observed for wild-type mLTP. For wild-type mLTP, the crystal and solution

structures show little change in conformation upon ligand binding,[17, 18] and we suspect this to be the case for C50A/C89E.

We have successfully used computational protein design to remove disulfide bridges in mLTP and experimentally determined its effect on protein stability and ligand binding. Not surprisingly, the removal of the disulfide bridges destabilized mLTP. We determined two of the four disulfide bridges could be removed individually and the designed variants appear to retain their tertiary structure as they are still able to bind palmitate. The C50A/C89E design with three compensating hydrogen bonds was the least destabilized, while C4H/C52A/N55E and C4Q/C52A/N55S appeared to show greater conformational change upon ligand binding.

Future Directions

The C4-C52 variants are promising as the basis for the development of a reagentless biosensor. Fluorescent sensors are extremely sensitive to their environment, by conjugating a sensor molecule to the site of conformational change, the change in sensor signal could be a reporter for ligand binding. Hellinga and co-workers had constructed a family of biosensors for small, polar molecules using the periplasmic binding proteins,[29] but a complementary system for nonpolar molecules has not been developed. Given the nonspecific nature of mLTP ligand binding, mLTP could be engineered to be a reagentless biosensor for small nonpolar molecules.

References

1. van Vlijmen, H. W. T., Gupta, A., Narasimhan, L. S. & Singh, J. A Novel Database of Disulfide Patterns and its Application to the Discovery of Distantly Related Homologs. *Journal of Molecular Biology* 335, 1083-1092 (2004).
2. Gupta, A., Van Vlijmen, H. W. T. & Singh, J. A classification of disulfide patterns and its relationship to protein structure and function. *Protein Sci* 13, 2045-2058 (2004).
3. Betz, S. F. Disulfide bonds and the stability of globular proteins. *Protein Sci* 2, 1551-1558 (1993).
4. Doig, A. J. & Williams, D. H. Is the hydrophobic effect stabilizing or destabilizing in proteins? : The contribution of disulphide bonds to protein stability. *Journal of Molecular Biology* 217, 389-398 (1991).
5. Hinck, A. P., Truckses, D. M. & Markley, J. L. Engineered Disulfide Bonds in Staphylococcal Nuclease: Effects on the Stability and Conformation of the Folded Protein. *Biochemistry* 35, 10328-10338 (1996).
6. Aslund, F. & Beckwith, J. Bridge over Troubled Waters: Sensing Stress by Disulfide Bond Formation. *Cell* 96, 751-753 (1999).
7. Hogg, P. J. Disulfide bonds as switches for protein function. *Trends in Biochemical Sciences* 28, 210-214 (2003).
8. Wetzel, R. Harnessing Disulfide Bonds Using Protein Engineering. *Trends in Biochemical Sciences* 12, 478-482 (1987).

9. Matsumura, M., Becktel, W. J., Levitt, M. & Matthews, B. W. Stabilization of Phage T4 Lysozyme by Engineered Disulfide Bonds. *PNAS* 86, 6562-6566 (1989).

10. Matsumura, M., Signor, G. & Matthews, B. W. Substantial increase of protein stability by multiple disulphide bonds. *Nature* 342, 291-293 (1989).

11. Price-Carter, M., Hull, M. S. & Goldenberg, D. P. Roles of Individual Disulfide Bonds in the Stability and Folding of an ω-Conotoxin. *Biochemistry* 37, 9851-9861 (1998).

12. Klink, T. A., Woycechowsky, K. J., Taylor, K. M. & Raines, R. T. Contribution of disulfide bonds to the conformational stability and catalytic activity of ribonuclease A. *European Journal of Biochemistry* 267, 566-572 (2000).

13. Graziano, G., Catanzano, F. & Notomista, E. Enthalpic and entropic consequences of the removal of disulfide bridges in ribonuclease A. *Thermochimica Acta* 364, 165-172 (2000).

14. Dahiyat, B. I. & Mayo, S. L. Probing the role of packing specificity in protein design. *Proceedings of the Natational Academy of Sciences of the United States of America* 94, 10172-7. (1997).

15. Malakauskas, S. M. & Mayo, S. L. Design, structure and stability of a hyperthermophilic protein variant. *Nature Struct Biol* 5, 470-475 (1998).

16. Marshall, S. A. & Mayo, S. L. Achieving stability and conformational specificity in designed proteins via binary patterning. *J Mol Biol* 305, 619-31. (2001).

17. Shin, D. H., Lee, J. Y., Hwang, K. Y., Kyu Kim, K. & Suh, S. W. High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure* 3, 189-199 (1995).

18. Gomar, J. *et al.* Solution structure and lipid binding of a nonspecific lipid transfer protein extracted from maize seeds. *Protein Sci* 5, 565-577 (1996).

19. Han, G. W. *et al.* Structural basis of non-specific lipid binding in maize lipid-transfer protein complexes revealed by high-resolution X-ray crystallography. *Journal of Molecular Biology* 308, 263-278 (2001).

20. Molina, A., Segura, A. & Garcia-Olmedo, F. Lipid transfer proteins (nsLTPs) from barley and maize leaves are potent inhibitors of bacterial and fungal plant pathogens. *FEBS Letters* 316, 119-122 (1993).

21. Marshall, S. A. & Mayo, S. L. Achieving stability and conformational specificity in designed proteins via binary patterning. *Journal of Molecular Biology* 305, 619-631 (2001).

22. Mayo, S. L., Olafson, B. D. & Goddard, W. A. Dreiding - a Generic Force-Field for Molecular Simulations. *Journal of Physical Chemistry* 94, 8897-8909 (1990).

23. Dahiyat, B. I. & Mayo, S. L. Probing the role of packing specificity in†protein†design. *PNAS* 94, 10172-10177 (1997).

24. Dahiyat, B. I., Gordon, D. B. & Mayo, S. L. Automated design of the surface positions of protein helices. *Protein Sci* 6, 1333-1337 (1997).

25. Street, A. G. & Mayo, S. L. Pairwise calculation of protein solvent-accessible surface areas. *Folding & Design* 3, 253-258 (1998).

26. Lazaridis, T. & Karplus, M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *Journal of Molecular Biology* 288, 477-487 (1999).

27. Pierce, N. A., Spriet, J. A., Desmet, J. & Mayo, S. L. Conformational splitting: a more powerful criterion for dead-end elimination. *J Comp Chem* 21, 999-1009 (2000).

28. Lin, C.-H., Li, L., Lyu, P.-C. & Chang, J.-Y. Distinct Unfolding and Refolding Pathways of Lipid Transfer Proteins LTP1 and LTP2. *The Protein Journal* 23, 553-566 (2004).

29. De Lorimier, R. M. *et al.* Construction of a fluorescent biosensor family. *Protein Science* 11, 2655-2675 (2002).

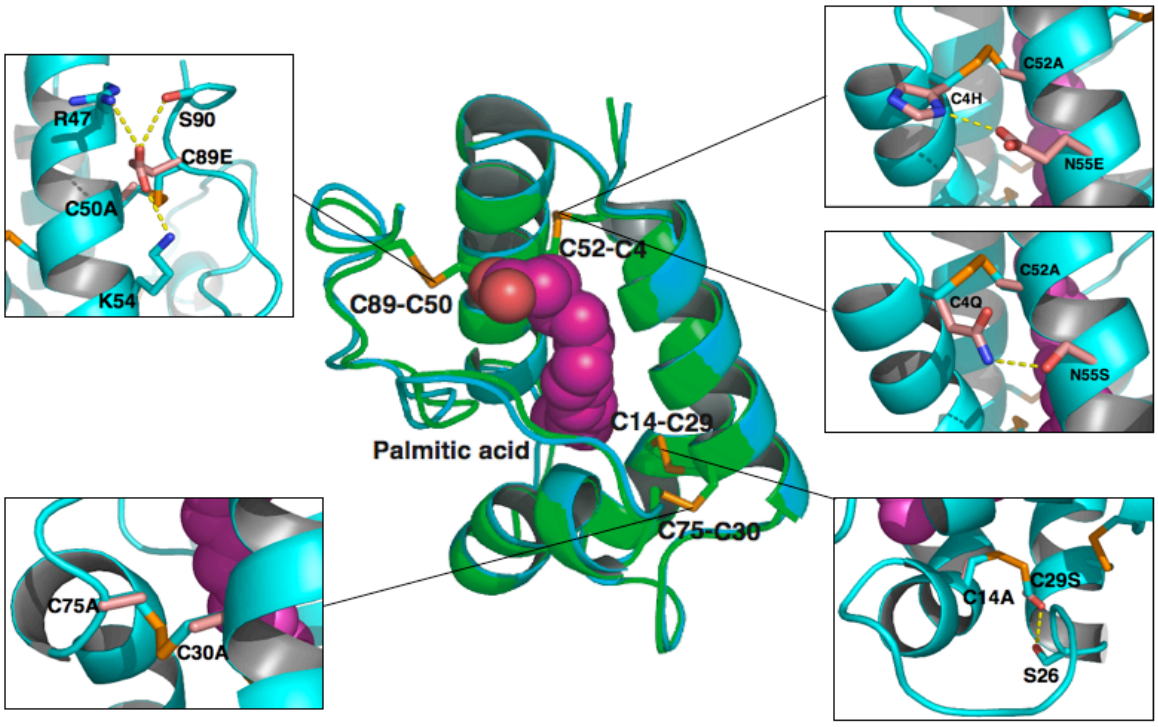


Figure 2-1. Ribbon diagram of mLTP and the designed variants of each disulfide. The palmitate bound mLTP (cyan) is superimposed on the unbound protein (green). Palmitate is shown in spheres with carbon in magenta and oxygen in red. Disulfides are in orange. In panels, mutated residues and the residues they form hydrogen bonds with are shown in stick with CPK-inspired colors, and the modeled hydrogen bonds are shown with yellow dashed lines, with measured heavy atom distances between 2.8 and 3.0 Å.

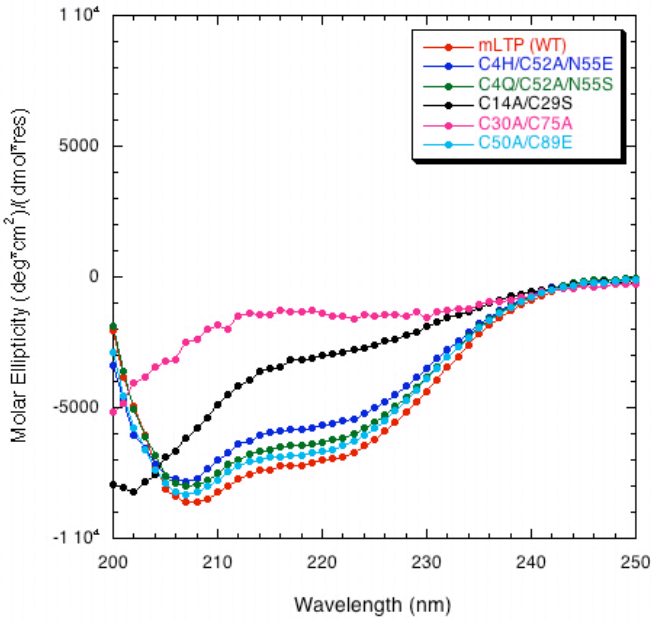


Figure 2-2. Wavelength scans of mLTP and designed variants. Variants C4H/C52A/N55E and C4Q/C52A/N55S and C50A/C89E are folded similar to wild-type mLTP, with minimums at 208nm and 222nm, but C14A/C29S and C30A/C75A are misfolded.

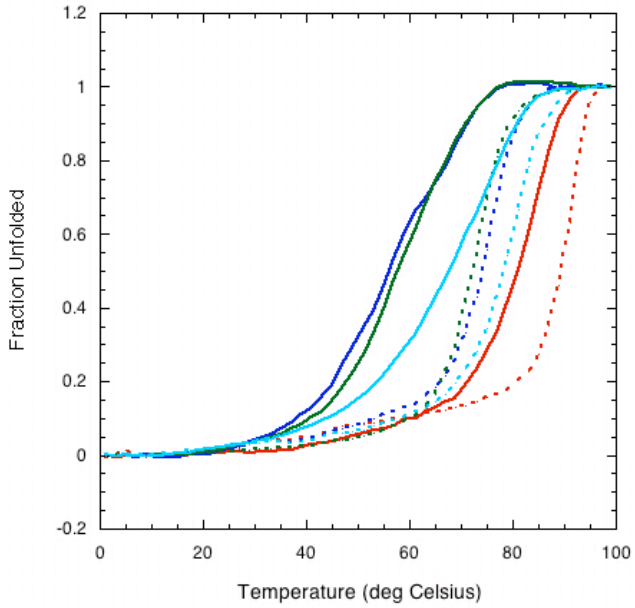


Figure 2-3. Thermal denaturations of mLTP and designed variants: mLTP (red), C4H/C52A/N55E (blue), C4Q/C52A/N55S (green), and C50A/C89E (cyan). Solid lines are protein alone, dashed lines are protein with palmitate added. Removal of disulfide bridges significantly destabilized the protein, but the variants still bound palmitate.

Table 2-1. Apparent T_ms of mLTP and designed variants.

	Apparent T_m		ΔT_m
	Protein alone	Protein + palmitate	
mLTP	84	92	8
C4H/C52A/N55E	56	76	20
C4Q/C52A/N55S	56	74	18
C50A/C89E	74	80	6

Chapter 3

Engineering a Reagentless Biosensor for Nonpolar Ligands

Adapted from manuscript in preparation by Jessica Mao, Eun Jung Choi and Stephen L. Mayo. To be submitted.

Introduction

Recently there has been interest in using proteins as carriers for drugs due to their high affinity and selectivity for their targets.[1] The proteins would not only protect the unstable or harmful molecules from oxidation and degradation, they would also aid in solubilization and ensure a controlled release of the agents. Advances in genetic and chemical modifications on proteins have made it easier to engineer proteins for specific use. Non-specific lipid transfer proteins (ns-LTP) from plants are a family of proteins that are of interest as potential carriers for nonpolar ligands for drug delivery.[2, 3] The two classes of LTPs (LTP1 and LTP2) share eight conserved cysteines that form four disulfide bridges and both have large nonpolar binding pockets.[4-6] The ns-LTP1 bind various polar lipids, fatty acids and acyl-coenzyme A[5] while ns-LTP2 bind bulkier sterol molecules.[7]

In a study to determine the suitability of ns-LTPs as drug carriers, the intrinsic tyrosine fluorescence of wheat ns-LTP1 (wLTP) was monitored, and wLTP was found to bind to BD56, an antitumoral and antileishmania drug, and amphotericin B, an antifungal drug.[3] However, this method is not very sensitive, as there are only two tyrosines in wLTP. Cheng *et al*. virtually screened over 7000 compounds for potential binding to maize ns-LTP1.[2] A reliable, sensitive, high throughput method to screen for binding of the drug compounds to mLTP is still necessary to test the potential of mLTP as drug carriers against known drug molecules.

Gilardi and co-workers engineered the maltose binding protein for reagentless fluorescence sensing of maltose binding;[9] their work was subsequently extended to construct a family of fluorescent biosensors from periplasmic binding proteins. By conjugating various fluorophores to the family of proteins, Hellinga and co-workers were able to construct nanomolar to millimolar sensors for ligands including sugars, amino acids, anions, cations, and dipeptides.[10-12]

Here we extend our previous work on the removal of disulfide bridges on mLTP, and report the engineering of mLTP as a reagentless biosensor for nonpolar ligands by conjugation with acrylodan, a thiol-reactive fluorescent probe.

Materials and Methods

Protein Expression, Purification and Acrylodan Labeling

The *Escherichia coli* expression optimized gene encoding the mLTP amino acid sequence was synthesized and ligated into the pET15b vector (Stratagene) by Blue Heron Biotechnology (www.blueheronbio.com). The pET15b vector includes an N-terminal His-tag. Inverse PCR mutagenesis was used to construct four variants: C52A, C4H/N55E, C50A, and C89E. The proteins were expressed in BL21(DE3) Gold cells (Stratagene) at 37 °C after induction with IPTG (isopropyl-beta-D-thiogalactopyranoside). The proteins expressed in the soluble fraction. Cells were resuspended in lysis buffer (50 mM

sodium phosphate, 300 mM sodium chloride, 10 mM imidazole, pH 8.0) and lysed by passing through the Emulsiflex at 15,000 psi, and the soluble fraction was obtained by centrifuging at 20,000*g* for 30 minutes. Protein purification was a two step process. First, the soluble fraction of the cell lysate was loaded onto a Ni-NTA column, eluted with elution buffer (lysis buffer with 400 mM imidazole), and concentrated to 10-20 μM. 6-acryloyl-2-(dimethylamino)naphthalene (acrylodan) was dissolved in acetonitrile and added to the elutions in 10-fold excess concentration, and the solution was incubated at 4 °C overnight. All solutions containing acrylodan were protected from light. Precipitated acrylodan and protein were removed by centrifugation and filtering through 0.2 μm nylon membrane Acrodisc syringe filters (Gelman Laboratory) and the soluble fraction was concentrated. Unreacted acrylodan and protein impurities were removed by gel filtration with phosphate buffer (50 mM sodium phosphate, 150 mM sodium chloride, pH 7.5), simultaneously monitoring at 280 nm for protein and 391 nm for acrylodan. The peak with both 280 nm and 391 nm absorbance was collected. The conjugation reaction looked to be complete, as both absorbances overlapped. Purified proteins were verified by SDS-Page to be of sufficient purity, and MALDI-TOF showed that they correspond to the oxidized form of the proteins with acrylodan conjugated. Protein concentration was determined with the BCA assay with BSA as the protein standard (Pierce).

Circular Dichroism Spectroscopy

Circular dichroism (CD) data were obtained on an Aviv 62A DS spectropolarimeter equipped with a thermoelectric cell-holder. Wavelength scans and thermal denaturation data were obtained from samples containing 50 µM protein. For wavelength scans, data were collected every 1 nm from 250 to 200 nm with an averaging time of 5 seconds at 25°C. For thermal studies, data were collected every 2 °C, from 1°C to 99°C, using an equilibration time of 120 seconds and an averaging time of 30 seconds. As the thermal denaturations were not reversible, we could not fit the data to a two-state transition. The apparent T_ms were obtained from the inflection point of the data. For thermal denaturations of protein with palmitate, 150 µM palmitate was added to 50 µM protein from stock solution of > 30 mM palmitate in ethanol (Sigma Aldrich).

Fluorescence Emission Scan and Ligand Binding Assay

Ligand binding was monitored by observing the fluorescence emission of protein-acrylodan conjugates with the addition of palmitate. Fluorescence was performed on a Photon Technology International Fluorometer equipped with stirrer at room temperature. Excitation was set to 363 nm and emission was followed from 400 to 600 nm at 2 nm intervals and 0.5 second integration time. The average of three consecutive scans were taken. 2 ml of 500 nM protein-acrylodan conjugate was used, and sodium palmitate (100uM) was titrated in.

Curve Fitting

The dissociation constants (K_d) were determined by fitting the decrease in fluorescence with the addition of palmitate to equation (3-1) assuming one binding site. The concentration of the protein-ligand complex (PL) is expressed in terms of K_d, total protein (P_0) and ligand (L_0) concentrations in equation (3-2).

$$F = F_0(P_0 - [PL]) + F_{max}[PL] \qquad (3\text{-}1)$$

$$[PL] = \frac{(P_0 + K_d + L_0) - \sqrt{(P_0 + K_d + L_0)2 - 4 * P_0 * L_0}}{2} \qquad (3\text{-}2)$$

Results

Protein-Acrylodan Conjugates

Previously, we had successfully expressed mLTP recombinantly in *Escherichia coli*. Our work using computational design to remove disulfide bridges resulted in stable mLTP variants in which the disulfide bridges C4-C52 and C50-C89 were removed individually (Figure 3-1). The variants are less stable than wild-type mLTP, but still bind to palmitate, a natural ligand. The removal of the disulfide bond could make the protein more flexible, and we coupled the conformational change with a detectable probe to develop a reagentless biosensor.

We chose two of the variants, C4H/C52A/N55E and C50A/C89E, and mutated one of the original Cys residues in each variant back. This gave us four new variants; C52A, C4H/N55E, C50A, and C89E. We conjugated acrylodan, an

environment sensitive thiol-reactive fluorophore[13] to the resulting free Cys in each protein. Trypsin digest and tandem mass spectrometry of the C52A-acrylodan complex (C52A/4C-Ac) confirmed the conjugation of acrylodan on Cys4. Figure 3-2 illustrates the site of acrylodan conjugation on C52A. The sulfur atom of Cys4 that forms a covalent bond with acrylodan is ~ 14 Å away from the closest carbon atom on palmitate.

We obtained the circular dichroism wavelength scans of the protein-acrylodan conjugates to ensure they were properly folded (Figure 3-3). While all four conjugates appeared folded, with characteristic helical protein minimums near 208nm and 222nm, only C52A/4C-Ac was most like wild-type mLTP.

Fluorescence of Protein-Acrylodan Conjugates

The fluorescence emission scans of the protein-acrylodan conjugates are varied in intensity and position of λ_{max}. C50A/89C-Ac, with acrylodan on the free Cys at residue 89, is the most shifted, with peak at 444 nm. C89E/50C-Ac, with acrylodan on the more buried C50, has λ_{max} at 464 nm. For the C4-C52 pair, conjugating acrylodan to the more solvent exposed C4 for C52A/4C-Ac results in a peak at 456 nm, while conjugating to the more buried C52 for C4H/N55E/52C-Ac gives a peak at 476 nm. In both C4-C52 and C50-C89, acrylodan in the more buried positions on the protein caused the spectra to be blue shifted compared to its more exposed partners (Figure 3-4).

Ligand Binding Assays

We performed titrations of the protein-acrylodan conjugates with palmitate to test the ability of the engineered mLTPs to act as biosensors. Of the four protein-acrylodan conjugates, C52A/C4-Ac seemed to show the most marked difference in signal when palmitate is added. The fluorescence of C52A/4C-Ac decreased as palmitate is titrated in (Figure 3-5a). The fluorescence emission maximum at 476nm was used to fit a single site binding equation. We determined the K_d to be 70 nM (Figure 3-5b).

To verify the observed fluorescence change was due to palmitate binding, we assayed for binding by comparing the thermal denaturations of C52A/4C-Ac alone, and with palmitate. We observed a change in apparent Tm from 59 ºC to 66 ºC as palmitate is added to the protein-acrlodan conjugate (Figure 3-6). The difference of 7 ºC is similar to the 8 ºC observed in apparent Tm increase for wild-type mLTP.

Discussion

We have successfully engineered mLTP into a fluorescent reagentless biosensor for nonpolar ligands. We believe the change in acrylodan signal is a measure of the local conformational change the protein variants undergo upon ligand binding. The conjugation site for acrylodan is on the surface of the protein, away from the binding pocket (Figure 3-7). It is possible that acrylodan, being a hydrophobic molecule, occupies the binding pocket of mLTP when no ligand is

bound. The removal of the C4-C52 disulfide bridge allows the N-terminal helix more flexibility, and could allow acrylodan to insert into the binding pocket. Upon ligand binding, however, acrylodan is displaced, going from an ordered nonpolar environment to a disordered polar environment. The observed decrease in fluorescence emission as palmitate is added is consistent with this hypothesis.

The engineered mLTP-acrylodan conjugate enables the high-throughput screening of the available drug molecules to determine the suitability of mLTP as a drug-delivery carrier. With the small size of the protein and high-resolution crystal structures available, this protein is a good candidate for computational protein design. The placement of the fluorescent probe away from the binding site allows the binding pocket to be designed for binding to specific ligands. enabling protein design and directed evolution of mLTP for specific binding to drug molecules for use as a carrier.

References

1. De Wolf, F. A. & Brett, G. M. Ligand-Binding Proteins: Their Potential for Application in Systems for Controlled Delivery and Uptake of Ligands. *Pharmacol Rev* 52, 207-236 (2000).

2. Cheng, C.-S. *et al.* Evaluation of plant non-specific lipid-transfer proteins for potential application in drug delivery. *Enzyme and Microbial Technology* 35, 532-539 (2004).

3. Pato, C. *et al.* Potential application of plant lipid transfer proteins for drug delivery. *Biochemical Pharmacology* 62, 555-560 (2001).

4. Shin, D. H., Lee, J. Y., Hwang, K. Y., Kyu Kim, K. & Suh, S. W. High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure* 3, 189-199 (1995).

5. Gomar, J. *et al.* Solution structure and lipid binding of a nonspecific lipid transfer protein extracted from maize seeds. *Protein Sci* 5, 565-577 (1996).

6. Han, G. W. *et al.* Structural basis of non-specific lipid binding in maize lipid-transfer protein complexes revealed by high-resolution X-ray crystallography. *Journal of Molecular Biology* 308, 263-278 (2001).

7. Samuel, D., Liu, Y.-J., Cheng, C.-S. & Lyu, P.-C. Solution Structure of Plant Nonspecific Lipid Transfer Protein-2 from Rice (*Oryza sativa*). *J. Biol. Chem.* 277, 35267-35273 (2002).

8. Gilardi, G., Zhou, L. Q., Hibbert, L. & Cass, A. E. G. Engineering the Maltose-Binding Protein for Reagentless Fluorescence Sensing. *Analytical Chemistry* 66, 3840-3847 (1994).

9. Gilardi, G., Mei, G., Rosato, N., Agro, A. F. & Cass, A. E. Spectroscopic properties of an engineered maltose binding protein. *Protein Eng.* 10, 479-486 (1997).

10. Marvin, J. S. *et al.* The rational design of allosteric interactions in a monomeric protein and its applications to the construction of biosensors. *PNAS* 94, 4366-4371 (1997).

11. Marvin, J. S. & Hellinga, H. W. Engineering Biosensors by Introducing Fluorescent Allosteric Signal Transducers: Construction of a Novel Glucose Sensor. *J. Am. Chem. Soc.* 120, 7-11 (1998).

12. De Lorimier, R. M. *et al.* Construction of a fluorescent biosensor family. *Protein Sci* 11, 2655-2675 (2002).

13. Prendergast, F. G., Meyer, M., Carlson, G. L., Iida, S. & Potter, J. D. Synthesis, spectral properties, and use of 6-acryloyl-2-dimethylaminonaphthalene (Acrylodan). A thiol-selective, polarity-sensitive fluorescent probe. *J. Biol. Chem.* 258, 7541-7544 (1983).

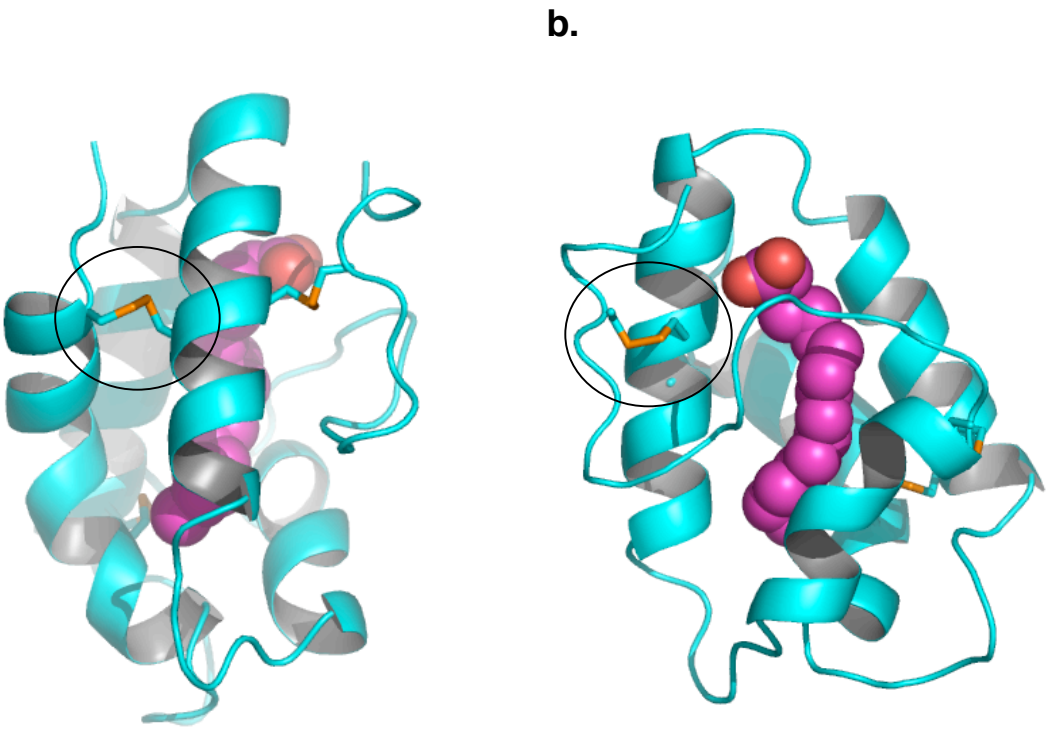
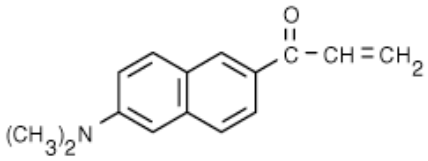


Figure 3-1. Ribbon representation of non-specific lipid-transfer protein from maize (mLTP). mLTP, a ns-LTP1 is shown bound to palmitatic acid, a fatty acid. Like all ns-LTP1s, it has eight conserved Cys, which form four disulfide bridges, shown in stick in orange. Palmitic acid is shown in spheres, with carbons in magenta, and oxygens in red. The disulfide bridge C4-C52 is circled in **a**, and in **b**, the C50-C89 pair is circled. Previous computational design work had created stable mutants of mLTP with the removal of each disulfide bridge.

a.



b.

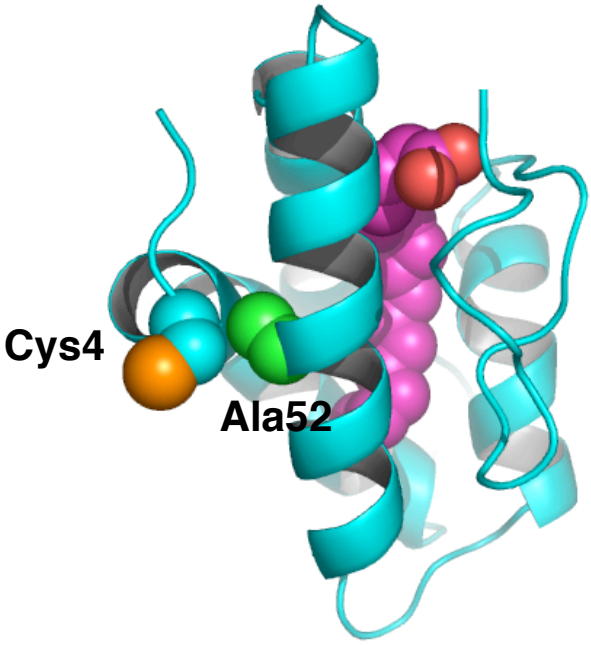


Figure 3-2. Acrylodan and its conjugation site on mLTP C52A. **a.** Structure of acrylodan; **b.** Ribbon representation of mLTP C52A. Palmitate (magenta), Ala52 (green), and Cys4 (cyan) are shown in space-filling models. Acrylodan is conjugated to the sulfur atom shown in orange. The distance between the sulfur atom and the closest carbon atom on palmitate is ~14 Å.

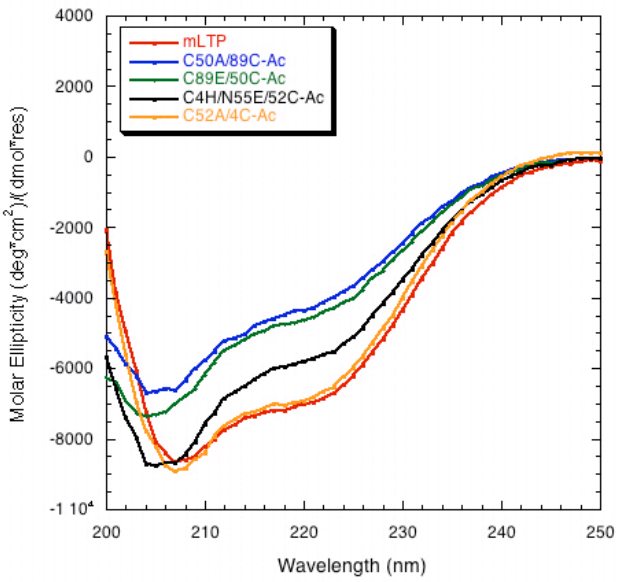


Figure 3-3. Circular dichroism wavelength scans of the four protein-acrylodan conjugates. Each conjugate shows the characteristic minimum near 208nm and 222nm for helical proteins. C52A/4C-Ac is most like wild-type mLTP.

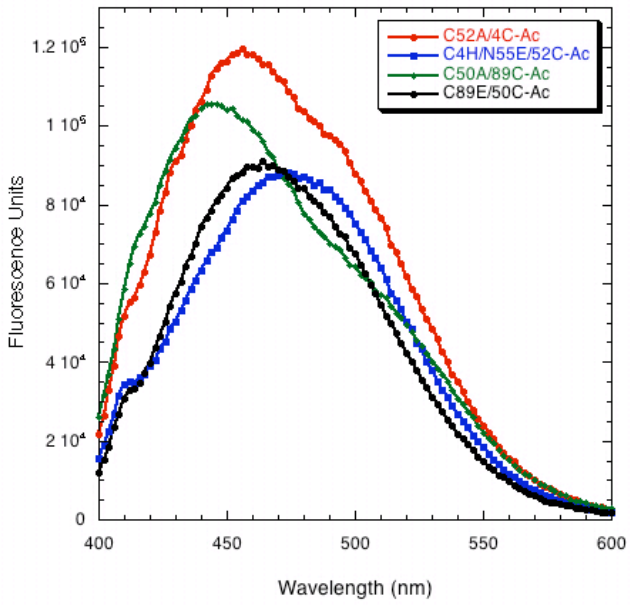


Figure 3-4. Fluoresence emission scans of mLTP-acrylodan conjugates. Excitation at 363 nm. Protein, λ_{max}: C50A/89C-Ac, 444 nm; C89E/50C-Ac, 464 nm; C52A/4C-Ac, 456 nm; and C4H/N55E/52C-Ac, 476 nm. In both C4-C52 and C50-C89, acrylodan in the more buried positions on the protein caused the spectra to be shifted compared to its more exposed partners.

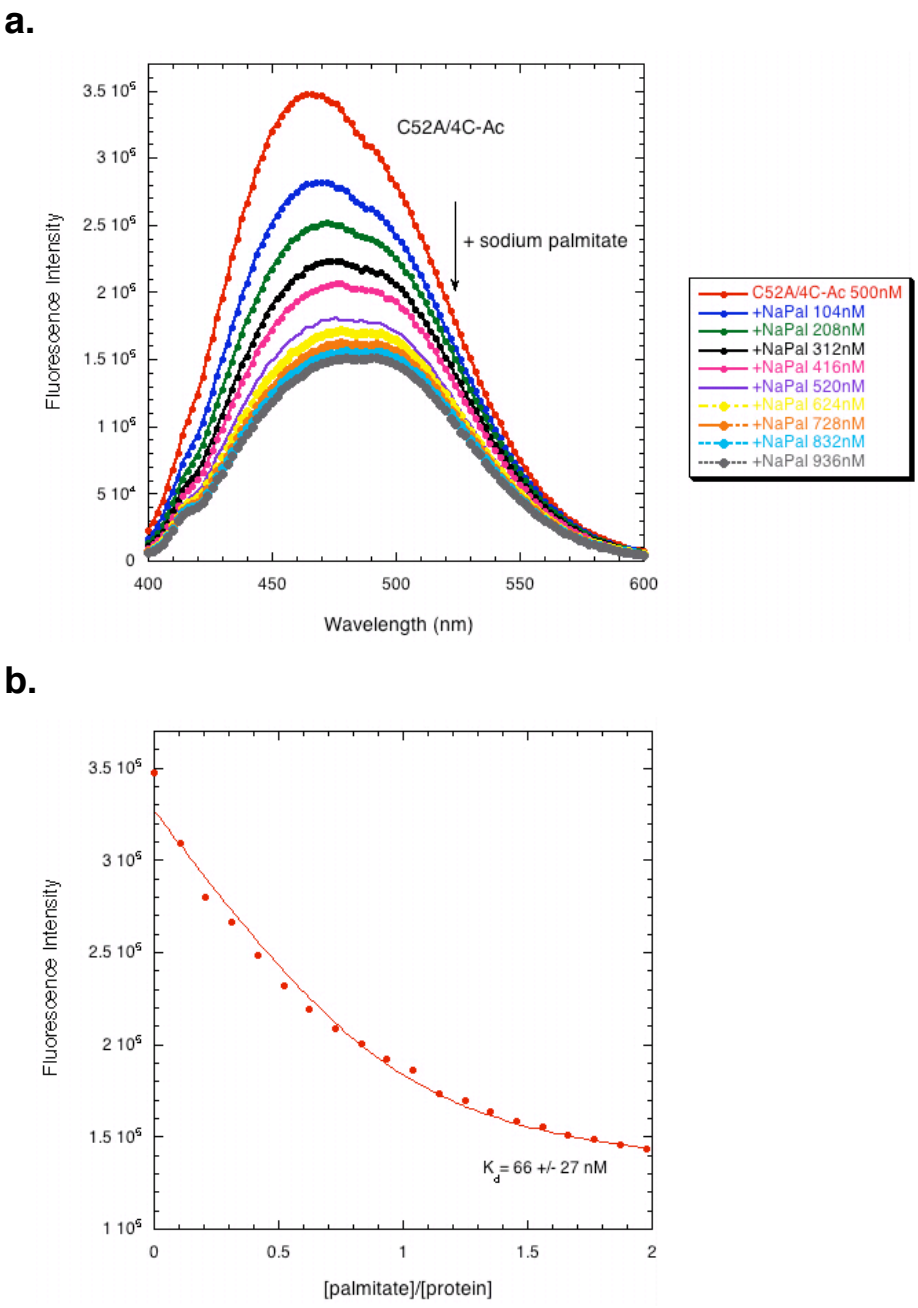


Figure 3-5. Titration of C52A/C4-Acrylodan with palmitate monitored by fluorescence emission. **a.** Fluorescence emission scans of C52A/4C-Ac (red) decreases as increasing concentration of sodium palmitate is added. Only a subset of experimental data is shown. Excitation wavelength is 363nm. **b.** Fluorescence monitored at 466nm was used to fit equation 3-1. K_d is dertermined to be 66 ± 27 nM.

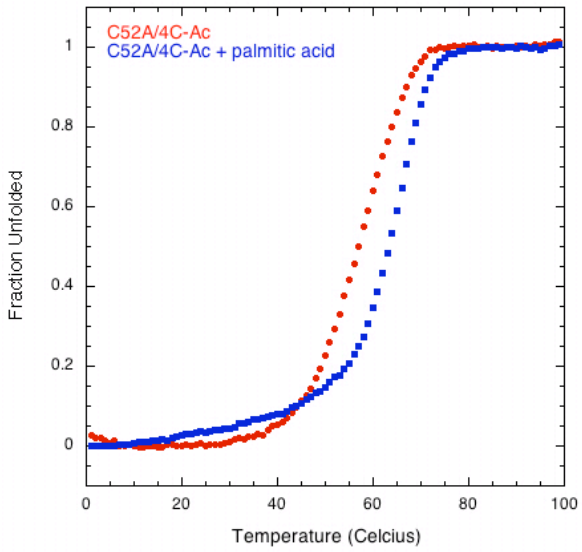


Figure 3-6. Thermal denaturations of C52A/4C-A monitored by CD. The increase in apparent T_m from 59°C for protein alone to 66°C for protein with palmitate indicates binding of palmitate to C52A/4C-Ac. The denaturation was not reversible, therefore the standard two-state model could not be used to fit the curve.

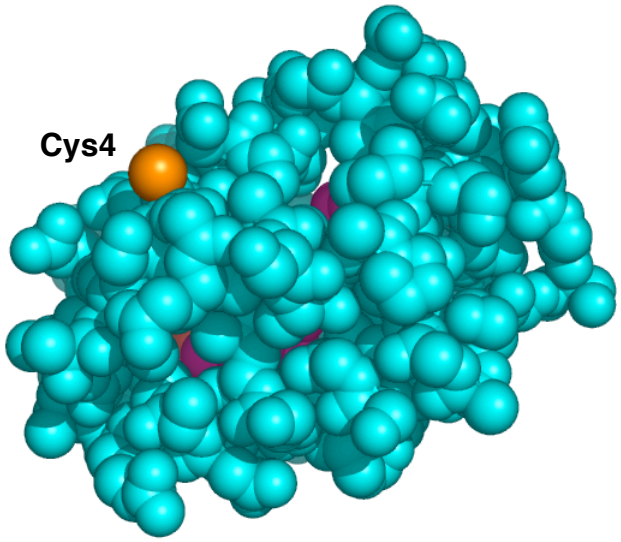


Figure 3-7. Space filling representation of mLTP C52A. Protein is shown in cyan, palmitate in magenta, while the sulfur atom of Cys4, the site of acrylodan conjugation, is shown in orange. Cys4 is on the surface of the protein, away from the binding pocket where palmitate binds.

Chapter 4

Designed Enzymes for Ester Hydrolysis

Introduction

One of the tantalizing promises protein design offers is the ability to design proteins with specified uses. If one could design enzymes with novel functions for the synthesis of industrial chemicals and pharmaceuticals, the processes could become safer and more cost- and environment-friendly. To date, biocatalysts used in industrial settings include natural enzymes, catalytic antibodies, and improved enzymes generated by directed evolution.[1] Great strides have been made via directed evolution, but this approach requires a high-throughput screen and a starting molecule with detectible base activity. Directed evolution is extremely useful in improving enzyme activity, but it cannot introduce novel functions to an inert protein. Selection using phage display or catalytic antibodies can generate proteins with novel function, but the power of these methods is limited by the use of a hapten and the size of the library that is experimentally feasible.[2]

Computational protein design is a method that could introduce novel functions. There are a few cases of computationally designed proteins with novel activities, the first of which is the "protozyme" PZD2 designed to hydrolyze *p*-nitrophenylacetate (PNPA) into *p*-nitrophenol and acetate.[3] This enzyme was built on the scaffold of the oxidation-reduction protein thioredoxin from *E. coli*. Bolon and Mayo utilized the "compute and build" model to create a cavity in thioredoxin that was complementary to the substrate. In the design, they fixed the substrate to the catalytic residue (His) by modeling a covalent bond and built

a rotamer library for the His-PNPA complex (Figure 4-1) by varying its rotatable bonds. The new rotamers, which model the high-energy state, are placed at different residue positions in the protein in a scan to determine the optimal position for the catalytic residue and the necessary mutations for surrounding residues. This method generated a protozyme with rate acceleration on the order of 10^2. In 2003, Looger *et al*. successfully designed an enzyme with triosephosphate isomerase (TIM) activity onto scaffolds of periplasmic binding proteins.[4] They used a method similar to that of Bolon and Mayo after first selecting for a protein that bound to the substrate. The resulting enzyme accelerated the reaction by 10^5, compared to 10^9 for wild-type TIM.

PZD2 was the first experimental validation of the design method, so it is not surprising that its rate acceleration is far less than that of natural enzymes. PZD2 has four anionic side chains located near the catalytic histidine. Since the substrate is negatively charged, we thought that the anionic side chains might be repelling the substrate leading to PZD2s low efficiency. To test this hypothesis, we mutated anionic amino acids near the catalytic site to neutral ones and determined the effect on rate acceleration. We also wanted to validate the design process using a different scaffold. Is the method scaffold independent? Would we get similar rate accelerations on a different scaffold? To answer these questions, we used our design method to confer PNPA hydrolysis activity into T4 lysozyme, a protein that has been well characterized.[5-10]

Materials and Methods

Protein Design with ORBIT

T4 lysozyme (PDB ID 1L63) was minimized briefly and designed using the ORBIT (Optimization of Rotamers by Iterative Techniques) protein design software suite.[11] A new rotamer library for the His-PNPA high energy state rotamer (HESR) was generated using the canonical chi angle values for the rotatable bonds as described.[3] The HESR library rotamers were sequentially placed at each non-glycine, non-proline non-cysteine residue position, and the surrounding residues were allowed to keep their amino acid identity or be mutated to alanine to create a cavity. The design parameters and energy function used were as described.[3] The active site scan resulted in Lysozyme 134, with the HESR placed at position 134.

Two variants, Rbias1.0 and Rbias2.5, (designed by Dan Bolon), focused on the catalytic positions of T4 lysozyme. He placed the HESR at position 26 and repacked the surrounding residues, incorporating ORBIT's RBIAS module.[12] RBIAS provides a way to bias sequence selection to favor interactions with a specified molecule or set of residues. In this case, the interactions between the protein and the HESR were scaled by 1.0 (no bias applied) and 2.5 (interaction energies are multiplied by 2.5) respectively.

Protein Expression and Purification

Thioredoxin mutants generated by site-directed mutagenesis (D10N, D13N, D15N, E85Q, and double mutant D13N_E85Q) were expressed as described.[3] The T4 lysozyme gene and mutants were cloned into pET11a and expressed in BL21-DE3 (Gold) cells from Stratagene. In addition to the designed mutations, D20N was incorporated to decrease the intrinsic activity of lysozyme and help protein expression. The wild-type His at position 31 was mutated to Gln. The cells were induced with IPTG at OD_{600} between 0.7 and1.0, and grown at 37 °C for 3 hours. The cells were lysed by sonication and protein was purified by FPLC and dialyzed into 10 mM sodium phosphate, pH 7.0. Lysozyme 134 was expressed in the soluble fraction, and purified first by ion exchange, followed by size exclusion gel filtration. Rbias1.0 and Rbias2.5 were in inclusion bodies. Induction temperatures of 30°C and 25°C were tried, but the two Rbias mutants were still insoluble. The pellet was washed with 50 mM Tris, 10 mM EDTA, 1 M urea, and 1% Triton-X100 three times and centrifuged. The remaining pellet was solubilized in buffer containing 4 M guanidine hydrochloride, purified by gel filtration in the same buffer and concentrated. The Hampton Research (Aliso Viejo, CA) Fold-It Screen was used to find a suitable buffer condition for protein folding. After CD wavelength scans to verify proper folding, buffer #15 (55 mM MES pH 6.5, 10.56 mM NaCl, 0.44 mM KCl, 1.1 mM EDTA, 440 mM sucrose, 550 mM L-arginine) was chosen and proteins were refolded, and then dialyzed

into 50 mM NaPi (pH 7.0) with 44 mM sucrose. Proteins were verified to be folded after dialysis by circular dichroism.

Circular Dichroism

Circular dichroism (CD) data were obtained on an Aviv 62A DS spectropolarimeter equipped with a thermoelectric cell-holder. Wavelength scans and thermal denaturation data were obtained from samples containing 10 μM protein in 25 mM sodium phosphate, pH 7.05. For wavelength scans, data were collected every 1 nm from 250 to 190 nm with an averaging time of 1 second; values from three scans were averaged. For thermal studies, data were collected every 1°C from 1°C to 99°C using an equilibration time of 120 seconds and an averaging time of 30 seconds. As the thermal denaturations were not reversible, we could not fit the data to a two-state transition. The apparent T_ms were obtained from the inflection point of the data.

Protein Activity Assay

Assays were performed as described in Bolon and Mayo[3] with 4 μM protein. K_m and K_{cat} were determined from nonlinear regression fits using KaleidaGraph.

Results

Thioredoxin Mutants

The computationally designed “protozyme” PZD2 had four anionic amino acids (D10, D13, D15, and E85) within 10 Å of the catalytic His17 (Figure 4-1). One rationale for the low rate acceleration of PZD2 is that the anionic amino acids repelled the negatively charged substrate, *p*-nitrophenylacetate (PNPA). We mutated the anionic amino acids to their neutral counterparts to generate the point mutants D10N, D13N, D15N, and E85Q, and also constructed a double mutant, D13N_E85Q, by mutating the two positions closest to the His17. The rate of PNPA hydrolysis was determined with Briggs-Haldane steady state treatment (Table 4-1). The five mutants all shared the same order of rate acceleration as PZD2. It seems that the anionic side chains near the catalytic His17 are not repelling the negatively charged substrate significantly.

T4 Lysozyme Designs

The T4 lysozyme variants, Rbias1.0 and Rbias2.5 were designed differently from 134. 134 was designed by an active site scan in which the HESR were placed at all feasible positions on the protein and all other residues were allowed wild type to alanine mutations, the same way PZD2 was designed. 134 ranked high when the modeled energies were sorted. The Rbias mutants were designed by focusing on one active site. The HESR was placed at the natural catalytic residues, 11, 20 and 26 in three separate calculations. Position 26 was chosen for further design in which the neighboring residues were designed to pack against the HESR. The sequences of 134, Rbias1.0 and Rbias2.5 are

compared in Figure 4-2. 134 is a fourfold mutant of lysozyme: D20N was made to reduce the native activity of the enzyme and to aid in protein expression; H31Q was incorporated to get rid of the native histidine and ensure that any observable activity is a result of the designed histidine; the A134H and Y139A mutations resulted directly from the active site scan (Figure 4-3).

The activity assays of the three mutants showed 134 to be active with the same order of rate acceleration as PZD2 (Table 4-2). Circular dichroism studies of 134 show it to be folded, with a wavelength scan and thermal denaturation comparable to wild-type lysozyme;[8] it exhibits irreversible unfolding upon thermal denaturation and has an apparent T_m of 54ºC (Figure 4-4).

Rbias1.0 and Rbias2.5 are both ten-fold mutants of lysozyme, including nonpolar to polar and polar to nonpolar mutations. They were refolded from inclusion bodies and CD wavelength scans had the same characteristics as wild-type lysozyme though signal intensity was only 10% of wild-type lysozyme. Their solubility in buffer was severely compromised and they did not accelerate PNPA hydrolysis above buffer background.

Discussion

The similar rate acceleration obtained by lysozyme 134 compared to PZD2 is reflective of the fact that the same design method was used for both proteins. This result indicates that the design method is scaffold independent. The Rbias mutants were designed to test the method of utilizing the native

catalytic site and additionally stabilizing the HESR in an attempt to stabilize the enzyme-transition state complex. It is unfortunate that the mutations have destabilized the protein scaffold and affected its solubility.

Since this work was carried out, Michael Hecht and co-workers have discovered PNPA-hydrolysis-capable proteins from their library of four-helix bundles.[13] The combinatorial libraries were made by binary patterning of polar and nonpolar amino acids to design sequences that are predisposed to fold. While the reported rate acceleration of 8700 is much higher than that of PZD2 or lysozyme 134, the sequence of S-824 contains 12 histidines and 8 lysines. We do not know if all of them are involved in catalysis, but it is certain that multiple side chains are responsible for the catalysis. For PZD2, it was shown that only the designed histidine is catalytic.

However, what is clear is that the simple reaction mechanism and low activation barrier of the PNPA hydrolysis reaction make it easier to generate *de novo* enzymes to catalyze the reaction. While PZD2 showed the necessity of a cavity for PNPA binding, it seems that the reaction is promiscuous and a nonspecific cavity with a nucleophilic side chain of the proper pK_a is sufficient for PNPA hydrolysis. Our design calculations have not taken side chain pK_a into account; it may be necessary to incorporate this into the design process in order to improve PZD2 and lysozyme 134 activity.

References

1. Valetti, F. & Gilardi, G. Directed evolution of enzymes for product chemistry. *Natural Product Reports* 21, 490-511 (2004).
2. Bolon, D. N., Voigt, C. A. & Mayo, S. L. De novo design of biocatalysts. *Curr Opin Chem Biol* 6, 125-9. (2002).
3. Bolon, D. N. & Mayo, S. L. From the Cover: Enzyme-like proteins by computational design. *PNAS* 98, 14274-14279 (2001).
4. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. Computational design of receptor and sensor proteins with novel functions. *Nature* 423, 185-90 (2003).
5. Bell, J. A. *et al.* Comparison of the crystal structure of bacteriophage T4 lysozyme at low, medium, and high ionic strengths. *Proteins* 10, 10-21 (1991).
6. Matthews, B. W. Studies on protein stability with T4 lysozyme. *Adv Protein Chem* 46, 249-78 (1995).
7. Llinas, M., Gillespie, B., Dahlquist, F. W. & Marqusee, S. The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat Struct Biol* 6, 1072-8 (1999).
8. McHaourab, H. S., Lietzow, M. A., Hideg, K. & Hubbell, W. L. Motion of Spin-Labeled Side Chains in T4 Lysozyme. Correlation with Protein Structure and Dynamics. *Biochemistry* 35, 7692-7704 (1996).

9. McHaourab, H. S., Oh, K. J., Fang, C. J. & Hubbell, W. L. Conformation of T4 lysozyme in solution. Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling. *Biochemistry* 36, 307-16 (1997).

10. Zhang, X. J., Wozniak, J. A. & Matthews, B. W. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *J Mol Biol* 250, 527-52 (1995).

11. Dahiyat, B. I. & Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* 278, 82-7. (1997).

12. Shifman, J. M. & Mayo, S. L. Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc Natl Acad Sci U S A* 100, 13274-9 (2003).

13. Wei, Y. & Hecht, M. H. Enzyme-like proteins from an unselected library of designed amino acid sequences. *Protein Engineering, Design and Selection* 17, 67-75 (2004).

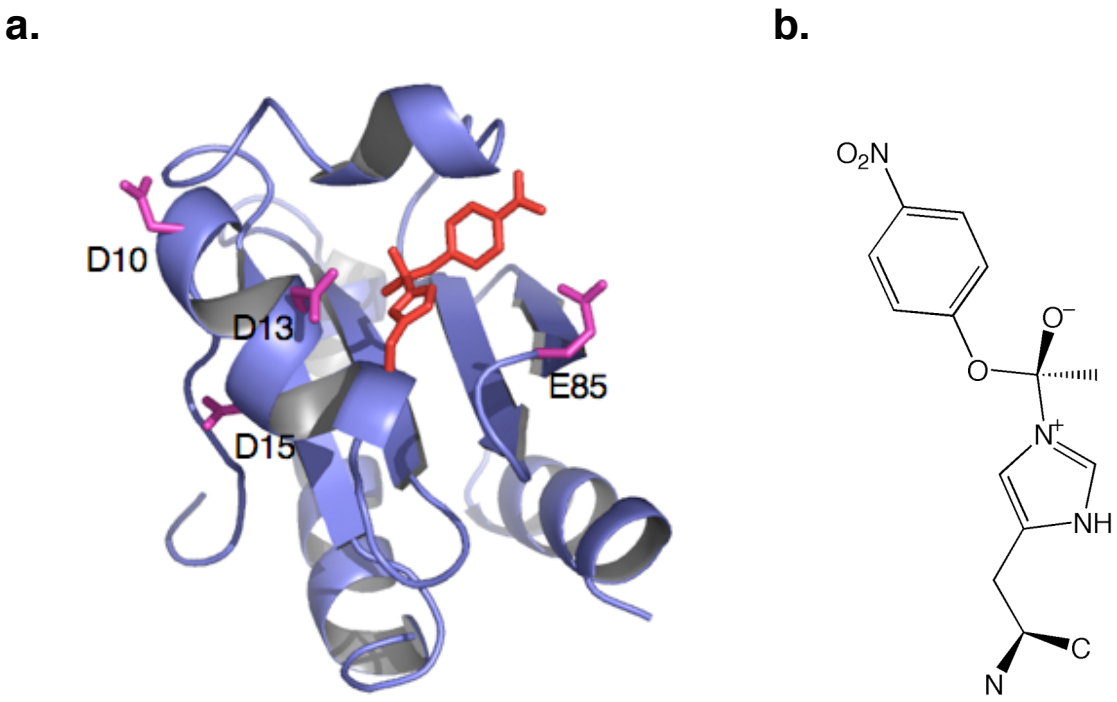


Figure 4-1. Ribbon model of PZD2 and structure of His-substrate high energy state rotamer. a. PZD2: the His-substrate High Energy State Rotamer is shown in red at residue 17. Four anionic residues within 10 Å of the catalytic His17 are shown in magenta (hydrogens not shown). **b.** Structure of the high energy state rotamer. Adapted from Bolon and Mayo.[3]

Table 4-1. Kinetic parameters of PZD2 and variants for PNPA hydrolysis.

	Distance to His17 (Å)	K_m (μM)	K_{cat} (s^{-1})	K_{cat}/K_{uncat}
PZD2	not applicable	170±20	$4.6±0.2*10^{-4}$	180
D13N	3.6	201±58	$7.0±0.6*10^{-4}$	129
E85Q	4.9	289±122	$9.8±1.5*10^{-4}$	131
D15N	6.2	729±801	$10.8±5.5*10^{-4}$	123
D10N	9.6	183±48	$22.2±1.8*10^{-4}$	138
D13N_E85Q	not applicable	197±63	$3.3±0.3*10^{-4}$	131

Residue:	11	20	21	22	26	31	32	70	105	134	139	141	142
WT:	E	D	T	E	E	H	L	D	Q	A	Y	Q	T
134:	-	N	-	-	-	Q	-	-	-	H	A	-	-
Rbias1.0:	L	-	E	L	H	N	I	F	F	-	-	L	I
Rbias2.5:	L	-	E	L	H	N	G	R	W	-	-	L	I

Figure 4-2. Sequence comparison of wild-type T4 lysozyme with 134, Rbias1.0 and Rbias2.5. The catalytic histidines are highlighted by the red boxes. 134 was designed in the same way as PZD2 to generate a cavity for the HESR, while Rbias mutants were designed primarily for stabilization of the neighboring residues with HESR. WT: wild-type T4 lysozyme.

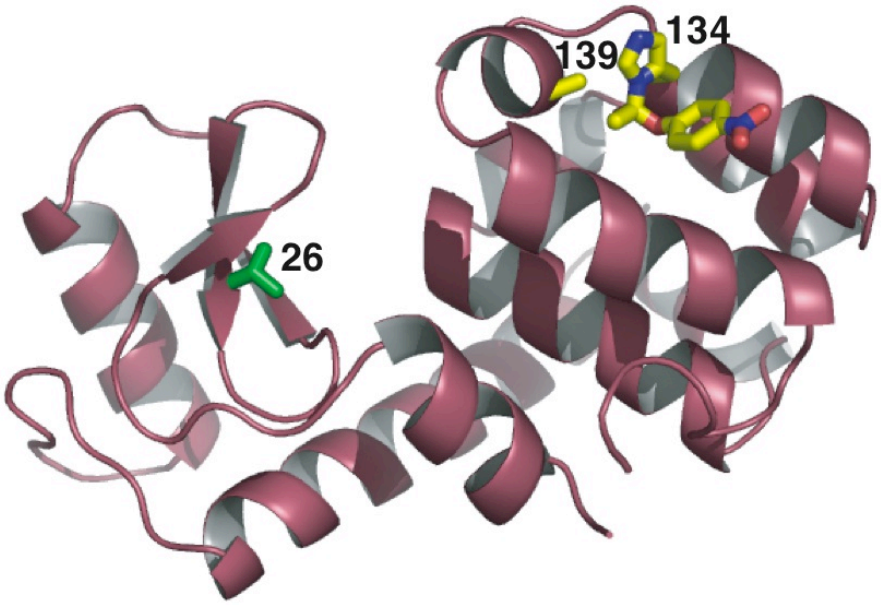
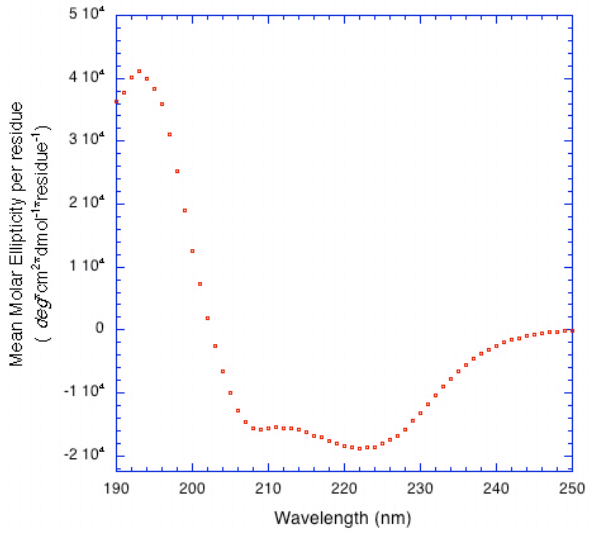


Figure 4-3. Lysozyme 134 highlighting the essential residues for catalysis. A134H and Y139A are the direct results of the active site scan on T4 lysozyme. HESR is placed at 134 and Y139 is mutated to Ala to create the necessary cavity. Residue 26 is shown in green to highlight the proposed active site of Rbias1.0 and Rbias2.5. HESR is shown in CPK-inspired colors.

a.



b.

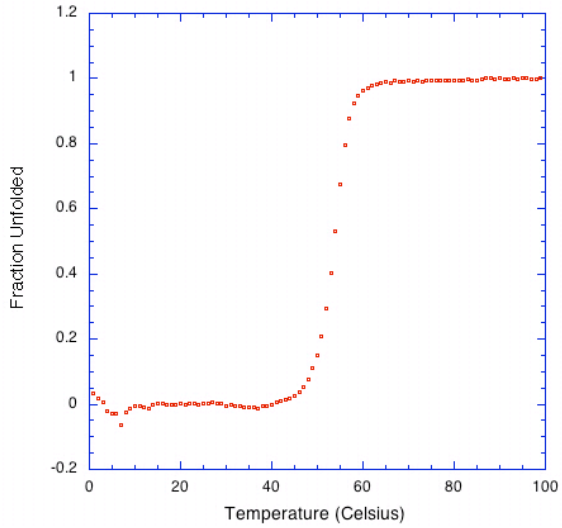


Figure 4-4. Circular dichroism characterization of lysozyme 134. a. Wavelength scan showing characteristic α-helical minimums at 208 and 222 nm. **b.** Thermal denaturation showing apparent T_m of 54°C.

Table 4-2. Kinetic parameters of lysozyme 134 compared to PZD2 for PNPA hydrolysis.

	T4 Lysozyme 134	PZD2
K_{cat}	$6.01*10^{-4}$ ($M*s^{-1}$)	$4.6*10^{-4}$($M*s^{-1}$)
K_{cat}/K_{uncat}	130	180
K_M	196 μM	170 μM

Chapter 5

Enzyme Design:

Toward the Computational Design of a Novel Aldolase

Enzyme Design

Enzymes are efficient protein catalysts. The best enzymes are limited only by the diffusion rate of substrates into the active site of the enzyme. Another major advantage is their substrate specificity and stereoselectivity to generate enantiomeric products. A few enzymes are already used in organic synthesis.[1] Synthesis of enantiomeric compounds is especially important in the pharmaceutical industry.[1, 2] The general goal of enzyme design is to generate designed enzymes that can catalyze a specified reaction. Designed enzymes are attractive industrially for their efficiency, substrate specificity, and stereoselectivity.

To date, directed evolution and catalytic antibodies have been the most proficient methods of obtaining novel proteins capable of catalyzing a desired reaction. However, there are drawbacks to both methods. Directed evolution requires a protein with intrinsic basal activity, while catalytic antibodies are restricted to the antibody fold and have yet to attain the efficiency level of natural enzymes.[3] Rational design of proteins with enzymatic activity does not suffer from the same limitations. Protein design methods allow new enzymes to be developed with any specified fold, regardless of native activity.

The Mayo lab has been successful in designing proteins with greater stability, and now we have turned our attention to designing function into proteins. Bolon and Mayo completed the first *de novo* design of an enzyme, generating a novel esterase, PZD2 on the *E. coli* thioredoxin scaffold.[4] PZD2

catalyzes the ester hydrolysis of *p*-nitrophenyl acetate (PNPA) into *p*-nitrophenol and acetate with histidine as the catalytic nucleophile. PZD2 exhibits "burst" phase kinetics characteristic of enzymes, with kinetic parameters comparable to those of early catalytic antibodies. The "compute and build" method was developed to generate this "protozyme" and can be applied to generate proteins with other functions. In addition to obtaining novel enzymes, we hope to gain insight into the evolution of functions and the sequence/structure/function relationship of proteins.

"Compute and Build"

The "compute and build" method takes advantage of the transition-state stabilization theory of enzyme kinetics. This method generates an active site with sufficient space to fit the substrate(s) and places a catalytic residue in the proper orientation. In generating PZD2 to catalyze the ester hydrolysis of PNPA, a high-energy state of the histidine-catalyzed PNPA hydrolysis reaction pathway was modeled as a series of His-PNPA rotamers.[4] Rotamers are discrete conformations of amino acids (in this case, the substrate (PNPA) was also included).[5] The high-energy state rotamer (HESR) was placed at each residue on the protein to find a proficient site. Neighboring side chains were allowed to mutate to Ala to create the necessary cavity. The protozymes generated by this method do not yet match the catalytic efficiency of natural enzymes. However,

the activity of the protozymes may be enhanced by improving the design scheme.

Aldolases

To demonstrate the applicability of the design scheme we chose a carbon-carbon bond-forming reaction as our target function, the aldol reaction. The aldol reaction is the chemical reaction between two aldehyde/ketone groups, yielding a β-hydroxy-aldehyde/ketone, which can be condensed by acid or base to afford an enone. It is one of the most important and utilized carbon-carbon bond forming reactions in synthetic chemistry (Figure 5-1). While synthetic methods have been successful, they often require multiple steps with protecting groups, preactivation of reactants and various reagents.[6] Therefore, it is desirable to have one-pot syntheses with enzymes that can catalyze specified reactions due to their superiority in efficiency, substrate specificity, stereoselectivity, and ease of reaction. While natural aldolases are efficient, they are limited in their substrate range. Novel aldolases that catalyze reactions between desired substrates would prove a powerful synthetic tool.

There are two classes of natural aldolases. Class I aldolases use the enamine mechanism in which the amino group of a catalytic Lys is covalently linked to the substrate to form a Schiff base intermediate. Class II aldolases are metalloenzymes that use the metal to coordinate the substrate's carboxyl oxygen. Catalytic antibody aldolases have been generated by the reactive

immunization method where a reactive “hapten” is used to elicit antibodies with catalytic residues at the active site.[7-9] The catalytic antibodies 33F12 and 38C2 use the enamine mechanism of class I aldolases (Figure 5-2). This mechanism involves the nucleophilic attack of the carbonyl C of the aldol donor by the unprotonated amino group of the Lys side chain to form Schiff base **1.** The Schiff base isomerizes to form enamine **2**, which undergoes further nucleophilic attack of the carbonyl C of the aldol acceptor. The resulting Schiff base **3** hydrolyzes to form high-energy state **4** which rearranges to release a β-hydroxy ketone without modifying the Lys side chain.[7]

The aldol reaction is an attractive target for enzyme design due to its simplicity and wide use in synthetic chemistry. It requires a single catalytic residue, Lys, with a shifted pK_a such that it is unprotonated. The intrinsic pK_a of Lys is 10.0,[10] yet pH studies of the catalytic Lys in 33F12 and 38C2 suggest that the pK_a of Lys is perturbed to 5.5 and 6.0 respectively.[7] The pK_a of Lys can be perturbed when in proximity to other cationic side chains, or when located in a local hydrophobic environment. The 2.15 Å crystal structure of the Fab’ antigen-binding fragment of 33F12 reveals that the catalytic Lys^{H93} is in a deep hydrophobic pocket (more than 11 Å deep), with mostly hydrophobic side chains within 4 Å (Figure 5-3). Lys^{H93} is in van der Waals contact with residues Leu^{H4}, Met^{H34}, Val^{H37}, Cys^{H92}, Ile^{H94}, Tyr^{H95}, Ser^{H100}, Tyr^{H102}, and Trp^{H103}. This feature is conserved in 38C2, which differs from 33F12 by 9 amino acids each in V_L and

V_H.[7] Clearly, in the absence of nearby cationic side chains a hydrophobic environment is required to keep Lys^{H93} unprotonated in its unliganded form.

Unlike natural aldolases, the catalytic antibody aldolases exhibit broad substrate range. In fact, over 100 aldehyde-aldehyde, aldehyde-ketone, and ketone-ketone aldol addition or condensation reactions have been catalyzed by 33F12 and 38C2.[7] This lack of substrate specificity is an artifact of the *reactive immunization* method used to raise them. Unlike catalytic antibodies raised with unreactive transition-state analogs, this method selects for reactivity instead of molecular complementarity. While these antibodies are useful in synthetic endeavors,[11, 12] their broad substrate range can become a drawback.

Target Reaction

Our goal was to generate a novel aldolase with the substrate specificity that a natural enzyme would exhibit. As a starting point, we chose to catalyze the reaction between benzaldehyde and acetone (Figure 5-4). We chose this reaction for its simplicity. Since this is one of the reactions catalyzed by the antibodies, it would allow us to directly compare our aldolase to the catalytic antibody aldolases. Intermolecular aldol reactions of acetone with aldehydes can be catalyzed by primary and secondary amines, including the amino acid proline.[13-15] Select kinetic parameters are shown in Table 5-1 for the proline- and catalytic antibody-catalyzed asymmetric aldol reaction of benzaldehyde with acetone (other primary and secondary amines have yields similar to that of

proline). Catalytic antibodies are more efficient than proline, with better stereoselectivity and yields.

Protein Scaffold

A protein scaffold that is inert relative to the target reaction is required for our design process. A survey of the PDB database shows that all known class I aldolases are $(\alpha/\beta)_8$, or TIM, barrels. In fact, this fold accounts for ~10% of all known proteins and all but one, Narbonin, are enzymes.[16] The prevalence of the fold and its ability to catalyze a wide variety of reactions make it an interesting system to study. Many $(\alpha/\beta)_8$ proteins have been studied to learn how barrel folds have evolved to have so many chemical functionalities. Debate continues as to whether all $(\alpha/\beta)_8$ proteins evolved from a single ancestor, or if the $(\alpha/\beta)_8$ fold is just a stable structure to which numerous enzymes converged. The IgG fold of antibodies and the $(\alpha/\beta)_8$ barrel represent two general protein folds with multiple functions. By using an $(\alpha/\beta)_8$ scaffold, in addition to catalytic antibodies, we can examine two distinct folds that catalyze the same reaction. These studies will provide insight into the relationship between the backbone structure and the activity of an enzyme.

In 2004, Dwyer *et al.* successfully engineered TIM activity into ribose binding protein (RBP), from the periplasmic binding protein family.[17] RBP is not catalytically active, but through both computational design and selection and 18-20 mutations, the new enzyme accomplishes 10^5-10^6 rate enhancement. The

periplasmic binding proteins have also been engineered into biosensors for a variety of ligands, including sugars, amino acids and dipeptides.[18] The high-energy state of the target aldol reaction is similar in size to the ligands, and the success of Dwyer *et al.* has shown RBP to be tolerant to a large number of mutations. We tried RBP as a scaffold for the target aldol reaction as well.

Testing of Active Site Scan on 33F12

The success of the aldolase design depends on our design method, the parameters we use, and the accuracy of the high energy state rotamer (HESR). Luckily, the crystal structure of the catalytic antibody 33F12 is available. We decided to test whether our design method could return the active site of 33F12.

To test our design scheme, we decided to perform an active site scan on the 2.15 Å crystal structure of the 33F12 Fab' antigen binding fragment (PDB ID: 1AXT), which catalyzes our desired reaction. If the design scheme is valid, then the natural catalytic residue, Lys^{H93}, with lysine on heavy chain position 93, should be within the top results from the scan. The structure of 33F12, which contains the "light" and "heavy" chains (Figure 5-5), was renumbered (Lys^{H93} became Lys^{H99}) and energy minimized for 50 steps. The constant region of the Fab was removed, and the antigen binding region, residues 1-114 of both chains, was scanned for an active site.

Hapten-like Rotamer

First, we generated a set of rotamers that mimicked the hapten used to raise the catalytic antibodies (Figure 5-6). The hapten used was a β-diketone, which serves as a trap for the ε-amino group of a reactive lysine. A reactive lysine has a perturbed pK_a, leaving an unprotonated ε-amino group. The amino group undergoes nucleophilic attack of the carbonyl carbon, causing the hapten to be covalently linked to the lysine and to absorb with λ_{max} at 318 nm. We modeled our hapten-like rotamer after the hapten-linked reactive lysine, with a methyl group in place of the long R group to facilitate the design calculations.

The rotamer was first built in BIOGRAF, with standard charges assigned; the rotatable bonds were allowed to assume the canonical values of 60°, -60°, and 180°, or 90°, -90° and 180° depending on the hybridization states. First, rotamers with all combinations of the different dihedral angles were modeled, and their energies were determined without minimization. The rotamers with severe steric clashes, as evidenced by energies >10,000 kcal/mol, were eliminated from the list. The remainder rotamers were minimized and the minimized energies were compared to further eliminate high energy rotamers to keep the rotamer library a manageable size. In the end, 14766 hapten-like rotamers were kept with minimized energies from 43.8--51.1 kcal/mol. This is a narrow range for ORBIT energies. The set of rotamers were then added to the current rotamer libraries.[5] They were added to the backbone-dependent e0 library where no χ angles were expanded, e2 library where both χ1 and χ2 angles of all amino acids

were expanded ±standard deviation, and the a2h1p0 library where the aromatic side chains were expanded for both χ1 and χ2, other hydrophobic residues were expanded for χ1, and no expansion used for polar residues.

With the new rotamers, we performed the active site scan on 33F12 first with the a2h1p0 library. We scanned residues 1-114 (the antigen binding region) of both the light and heavy chains by modeling the hapten-like rotamer at each qualifying position and allowed surrounding residues to be mutated to Ala to create the necessary space. Standard parameters for ORBIT were used, with 0.9 as the van der Waals radii scale factor, and type II solvation. The results were then sorted by residue energy or total energy (Table 5-2). Residue energy is the interaction energies of the rotamer with other side chains, and total energy is the total modeled energy of the molecule with the rotamer. Surprisingly, the native active site, Lys^{H99}, with Lys on residue 99 of the heavy chain, is not in the top 10 when sorted by residue energy, but is the second best energy when sorted by total energy. When sorted by total energy, we see the hapten-like rotamer is only half buried, as expected. The first one that is mostly buried (%b-T > 90%) is 33H, which is the top hit when sorting by total energy, with the native active site 99H second. Upon closer examination of the scan results, we see that 33H and 99H are lining the same cavity, and they put the hapten-like rotamer in the same cavity, therefore identifying the active site correctly.

HESR

Having correctly identified the active site with the hapten-like rotamer, we had confidence in our active site scan method. We wanted to test the library of high-energy state rotamers for the target aldol reaction. 33F12 is capable of catalyzing over 100 aldol reactions, including the target reaction between acetone and benzaldehyde. An active site scan using the HESR should return the native active site.

The "compute and build" method involves modeling a high-energy state in the reaction mechanism as a series of rotamers. Kinetic studies have indicated that the rate-determining step of the enamine mechanism is the C-C bond-forming step.[13] Of high energy states **3** and **4** shown in Figure 5-2, we chose to model **4** as the HESR. This was chosen instead of Schiff base **3** to allow enough space to be created in the active site for water to hydrolyze the product from the enzyme. The resulting rotamer is shown in Figure 5-7. The nine labeled dihedral angles were varied to generate the whole set of HESR. χ_1 and χ_2 values were taken from the backbone independent library of Dunbrack and Karplus[5] which is based on a survey of the PDB. χ_3 through χ_9 were allowed to be the canonical 60º, 180º, and -60º. Since there are two stereocenters, four new "amino acids" resulted representing all combinations. For each new χ angle, the number of rotamers in the rotamer list was increased 12-fold. To keep the library size manageable, the orientation of the phenyl ring and the second hydroxyl group were not defined specifically.

A rotamer list enumerating all combinations of χ values and stereocenters was generated (78732 total). 59839 rotamers with extremely high energies (>10000 kcal*mol^{-1}) were eliminated. The remaining 18893 rotamers were minimized to allow for small adjustments, and the internal energies were again calculated. An energy cutoff of 50 kcal*mol^{-1} was applied to further reduce the size of the rotamer set to 16111, 20.5% of the original rotamer list.

The set of rotamers were then added to the amino acid rotamer libraries.[5] They were added to the backbone-dependent e0 library where no χ angles were expanded (e0_benzal0), e2 library where both χ1 and χ2 angles of all amino acids were expanded by one standard deviation (e2_benzal0), and the a2h1p0 library where the aromatic side chains were expanded for both χ1 and χ2, other hydrophobic residues were expanded for χ1, and no expansion used for polar residues (a2h1p0_benzal0). Because the HESR set is already so large, no χ angle was expanded. These then served as the new rotamer libraries for our design.

The active site scan was carried out on the Fab binding region of 33F12 like above, and the top 10 results are shown in Table 5-3. The a2h1p0_benzal0 library was used as in scans. Whether we sort the results by residue energy or total energy, the natural catalytic Lys of 33F12 remains one of the 10 best catalytic residues, an encouraging result. A superposition of the modeled vs. natural active site shows the Lys side chain is essentially unchanged (Figure 5-8); χ_1 through χ_3 are approximately the same. Three additional mutations are

suggested by ORBIT after subtracting out mutations without HES present: Tyr^{L36}, Tyr^{H95}, Ser^{H100} are mutated to Ala in the modeled protein. No mutation is necessary to catalyze the desired reaction.

The mutations suggested by ORBIT could be due to the lack of flexibility of HESR. The HESR is not expanded around any χ angle, and χ_3 through χ_9 angles are defined by the canonical 60º, 180º, and -60º. This limits the allowed conformations of HESR. A small variation of ±5º in χ_3 could cause a significant change in the position of the phenyl ring. In addition, the HESRs are minimized individually; thus the HESR used may not represent the minimized conformation in the context of the protein. This is a limitation of the current method.

One way of solving this problem is to generate more HESRs. Once the approximate conformation of HESR is chosen, we can enumerate more rotamers by allowing the χ angles to be expanded by small increments. The new set of HESRs can then be used to see if any suggested mutations using the old HESR set are eliminated.

Both sorting by residue energy and total energy returned the native active site of 33F12 as 99H is in the top two results. While the hapten-like rotamer was able to identify the active site cavity, the HESR is a better predictor of active site residue. This result is very encouraging for aldolase design as it validates our “compute and build” design method for the design of a novel aldolase. We decided to start with TIM as our protein scaffold.

Enzyme Design on TIM

Triosephosphate isomerase (TIM) is the prototypical $(\alpha/\beta)_8$ barrel. TIM from *Trypanosomal brucei brucei* (PDB ID 5TIM) was chosen as our protein scaffold. It exists as a dimer with an estimated $K_D < 10^{-11}$ M.[19] Mutant monomeric versions have been made with decreased activity.[19] The 1.83 Å crystal structure consists of both subunits (residues 2 to 250) of the dimer (Figure 5-9a). Subunit A is crystallized in the "open" conformation, without any ligand bound. Subunit B is in the "almost-closed" conformation: the active site binds a sulfate ion which mimics the phosphate group of the natural substrates, D-glyceraldehyde-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP). The sulfate ion causes a flexible loop (loop 6) to fold over the active site.[20] This provides a convenient system in which two distinct conformations of TIM are available for modeling.

The dimer interface of 5TIM consists of 32 residues, and is defined as any residue within 4 Å of the other subunit. Each subunit inserts a C-terminal loop (loop 3) into the other subunit (Figure 5-9b). A salt bridge network is also present with each subunit donating four charged residues (Figure 5-9c). The natural active site of TIM, as with other TIM barrel proteins, is located on the C-terminal of the barrel. The catalytic residues are K13, H95, and E167. K13 and H95 are part of the interface. To prevent dimer dissociation, the interface residues were left "as is" for most of the modeling studies.

Active Site Scan on "Open" Conformation

The structure of TIM was minimized for 50 steps using ORBIT. For the first round of calculations, subunit A, the "open" conformation was used for the active site scan, while subunit B and the 32 interface residues were kept fixed. The newly generated rotamer libraries e0_benzal0, a2h1p0_benzal0, and e2_benzal0 were each tested. An active site scan involved positioning HESRs at each non-Gly, non-Pro, non-interface residue while finding the optimal sequence of amino acids to interact favorably with a chosen HESR. Since the structure of TIM shows residues 2 to 250, with 32 interface residues,14 Pro, and 31 Gly (3 at interface), each scan generated 175 models with HESR placed at a different catalytic residue position in each. Due to the large size of the protein, it was impractical to allow all the residues to vary. To eliminate residues that are far from the HESR from the design calculations, a preliminary calculation was run with HESR at the specified positions, with all other residues mutated to Ala. The distance of each residue to HESR was calculated, and those that were within 12 Å were selected. In a second calculation, HESR was kept at the specified position and the side chains that were not selected were held fixed. The identity of the selected residues (except Gly, Pro, and Cys) was allowed to be either wild type or Ala. Pairwise calculation of solvent-accessible surface area[21] was calculated for each residue. In this way, an active site scan using the a2h1p0_benzal0 library took about 2 days on 32 processors.

In protein design, there is always a tradeoff between accuracy and speed. In this case, using the e2_benzal0 library would provide us greatest accuracy, but each scan took ~4 days. After testing each library, we decided to use the a2h1p0_benzal0 library, which provided us with results that differed only by a few mutations from the results with the e2_benzal0 library. Even though a calculation using the a2h1p0_benzal0 library is not as fast as the e0_benzal0 library, it provides greater accuracy.

Both the hapten-like rotamer library and the HESR library were used in the active site scan of the open conformation of TIM. The top 10 results sorted by the interaction energy contributed by the HESR or hapten-like rotamer (residue energy), or total energy of the molecule, are shown in Table 5-4 and 5-5. Overall, sorting by residue energy or total energy gave reasonably buried active site rotamers. Residue positions that are highly ranked in both scans are candidates for active site residues.

Active Site Scan on "Almost-Closed" Conformation

The active site scan was also run with subunit B of TIM, the "almost-closed" conformation. This represents an alternate conformation that could be sampled by the protein. There are three regions that are significantly different between the two conformations, loop 5 (residues 129-142), loop 6 (167-180), referred to as the flexible loop, and loop 7 (212-216). The movements of the loops result in a rearrangement of hydrogen-bond interactions. The major

difference is in loop 6, which connects β6 to H6 (Figure 5-10). Gly175 of loop 6 is moved 6.9 Å, while the side chain oxygen atoms of the catalytic residue Glu167 are essentially in the same position.[20] The same minimized structure used in the "open" conformation modeling was used. The interface residues and subunit A were held fixed. The results of the active site scan are listed in Table 5-6.

The loop movements provide significant changes. Since both conformations are accessible states of TIM, we want to find an active site that is amenable to both conformations. The availability of this alternative structure allows us to examine more plausible active sites and in fact is one of the reasons that *Trypanosomal* TIM was chosen.

pK_a Calculations

With the results of the active site scans, we needed an additional method to screen the designs. A requirement of the aldolase is that it has a reactive lysine, which is a lysine with lowered pK_a. A good computational screen would be to calculate the pK_a of the introduced lysines.

While pK_a calculations are difficult to determine accurately, we decided to try the program Multi-Conformation Continuum Electrostatics (MCCE).[21, 22] It combines continuum electrostatics calculated by DelPhi and molecular mechanics force fields in Monte Carlo sampling to simultaneously calculate free energy, net charge, occupancy of side chains, proton positions, and pK_a of

titratable groups.[23] DelPhi implements the finite-difference Poisson-Boltzmann (FDPB) method to calculate electrostatic interactions.[24, 25]

To test the MCCE program, we ran some test cases on ribonuclease T1, phosphatidylinositol-specific phospholipase C, xylanase, and finally 33F12. Of the 17 titratable groups, 9 were within 1 pH unit of the experimentally determined pK_a, 2 were within 2 pH units, and 6 were >2 pH units away (Table 5-7). MCCE is the only pK_a program that allows the side chain conformations to vary, and is thus the most appropriate for our purpose. However, it is not accurate enough to serve as a computational screen for our design results currently.

Design on Active Site of TIM

A visual inspection of the results of the active site scan revealed that, in most cases, the HESR was insufficiently buried. Due to the requirement of the reactive lysine, we needed to insert a Lys into a hydrophobic environment. None of the designs put the Lys in a deep pocket. Also, with the difficulty of generating a new active site, we decided to focus on the native catalytic residue, Lys13. The natural active site already has a cavity to fit its substrates. It would be interesting to see if we can mutate the natural active site of TIM to catalyze our desired reaction. Since Lys13 is part of the interface, it was eliminated from earlier active site scans. In the current modeling studies, we are forcing HESR to be placed at residue 13, in both the “open” and “almost-closed” conformations. Because the protein is a symmetrical dimer, any residue on one subunit must be tolerated by

the other subunit. The results of the calculation are shown in Table 5-8. Interestingly, the “open” conformation led to more HES burial. After subtracting out the mutations that ORBIT predicts with the natural Lys conformation present instead of HESR for subunit A, one mutation (Ile172 to Ala) remains. Ile172 is in van der Waals clash with HESR, so it is mutated to Ala.

The HESR is only ~80% buried as QSURF calculates, and in fact, the rotamer looks accessible to solvent. Additional modeling studies were conducted in which the optimized residues are not limited to their wild type identities or Ala, however, due to the placement of Lys13 on a surface loop, the HESR is not sufficiently buried. The active site of TIM is not suitable for the placement of a reactive lysine.

Next, we turned to the ribose binding protein as the protein scaffold. At the same time, there had been improvements in ORBIT for enzyme design. SUBSTRATE and GBIAS were two new modules added. SUBSTRATE executes user-specified rotational and translational movements on a small molecule against a fixed protein, and GBIAS will add a bias energy to all interactions that satisfy user-specified geometry restraints. GBIAS is a quick way to eliminate rotamers that do not satisfy the restraints prior to calculation of interaction energies and optimization steps, which are the most time consuming steps in the process. Since GBIAS is a new module, we first needed to test its effectiveness in enzyme design.

GBIAS

In order to test GBIAS, we decided to use a natural aldolase. 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase was chosen (PDB ID 1EUA). It is a Class I aldolase whose reaction mechanism involves formation of a Schiff base. It is a trimer of $(\alpha/\beta)_8$ barrel and the 1.95 Å crystal structure has a covalent intermediate trapped.[26] The carbinolamine intermediate between lysine side chain and pyruvate was the basis for a new rotamer library, and, in fact, it is very similar to the HESR library generated for the acetone-benzaldehyde reaction (Figure 5-11). This is a further confirmation of our choice of HESR. The new rotamer library representing the trapped intermediate was named KPY, and all dihedral angles were allowed to be the canonical values of -60º, 60º and 180º.

We tested GBIAS on one subunit of the KDPG aldolase trimer. We put KPY at residue From the crystal structure, we see the contacts the intermediate makes with surrounding residues (Figure 5-12), and except the water-mediated hydrogen bond, we put in our GBIAS geometry definition file all the contacts that are in the crystal structure, allowing hydrogen bonding distances of 2.4--3.4 Å, and donor-hydrogen-acceptor angles between 140º and 180º. GBIAS energy was applied from 0 to 10 kcal/mol, and the results were compared to the crystal structure to determine if we captured the interactions. With no GBIAS energy (bias = 0), we do not retain any of the crystallographic hydrogen bonds. With bias energy of 5, we get 1, and with GBIAS energy of 10kcal/mol for each satisfied interaction, we do retain all the major interactions (Figure 5-12). KPY at

133 superimposes onto the crystallographic trapped intermediate. Arg49 and Thr73 also superimpose with their wild-type orientation. The only sidechain that differs from the wild type is Glu45, but that is probably due to the fact that water-mediated hydrogen bonds were not allowed.

The success of recapturing the active site of KDPG aldolase is a testament to the utility of GBIAS. Without GBIAS, we were not able to retain the hydrogen bonds that are present in the crystal structure. GBIAS was used for the focused design on RBP binding site.

Enzyme Design on Ribose Binding Protein

The ribose binding protein is a periplasmic transport protein. It is a two domain protein connected by a hinge region, which undergoes conformational change upon association with ribose. It binds ribose in a "clam-shell"-like manner, where the domains "close" on the ligand (Figure 5-13).[27] RBP binds ribose tightly, with K_d of 130nM. In the closed conformation, Asp89, Asp215, Arg91, Arg141, and Asn13 form an extensive hydrogen bonding network with ribose in the binding pocket. Because the binding pocket already has two cationic residues, Arg91 and Arg141, we felt this was a good candidate as a scaffold for the aldol reaction. A quick design calculation to put Lys instead of Arg at those positions yielded high probability rotamers for Lys. The HESR also has two hydroxl groups that could benefit from the hydrogen bond network available.

Due to the improvements in computing and the addition of GBIAS to ORBIT, we could process more rotamers than when we first started this project. We decided to build a new library of HESR to allow us a more accurate design. We added two more dihedral angles to vary. In addition to the 9 dihedral angles in Figure 5-7, the dihedral angle for the second hydroxyl group was allowed to be -60°, 60° and 180° while the phenyl ring could rotate as well. χ1 and χ2 were also expanded by ±15°, like that of a true e2 library. The new rotamer list was generated by varying all 11 angles, and rotamers with the lowest energies (minimum plus 5) were retained for merging with the backbone dependent e2QERK0 library where all residues except Q, E, R, K were expanded around χ1 and χ2. The HESR library contained 37,381 rotamers.

With the new rotamer library, we placed HESR at position 90 and 141 in separate calculations in the closed conformation (PDB ID 2DRI) to determine the better site for HESR. We superimposed the models with HESR at those positions with ribose in its crystallographic coordinates (Figure 5-14). HESR at position 141 better superimposed with ribose, meaning it would use the same binding residues, so further targeted designs focused on HESR at 141. For these designs, type 2 solvation was used penalizing for burial of polar surface area, and HERO obtained the global minimum energy conformation (GMEC). Residues surrounding 141 were allowed to be all residues except Met, and a second shell of residues were allowed to change conformation but not their amino acid identity. The crystallographic conformations of side chains were

allowed as well. Residues 215 and 235 were not allowed to be anionic residues since an anionic residue so close to the catalytic Lys would make it less likely to be unprotonated. Both geometry and energy pruning was used to cut down the number of rotamers allowed so the calculations were manageable. SBIAS was utilized to decrease the number of extraneous mutations by biasing toward the wild-type amino acid sequence. It was determined that 4 mutations were necessary to accommodate HESR at 141, D89V, N105S, D215A, and Q235L. These 4 mutations had the strongest rotamer-rotamer interaction energy with HESR at 141. The final model was minimized briefly and it shows positive contacts for HESR with surrounding residues (Figure 5-15). Both hydroxyl groups have the potential to make hydrogen bonds, and the phenyl ring of HESR is in a cage of phenyl rings as it is stacked in between the phenyl rings of Phe15 and Phe164, and perpendicular to Phe16.

Experiemental Results

Site-directed mutagenesis was used introduce R141K, D89V, N105S, D215V, and Q235L. Previously, Kyle Lassila had added a His-tag to the RBP gene for Ni-NTA column purification. Wild-type RBP and mutants were expressed in BL21(DE3) Gold cells at 37 °C, induction with 1mM IPTG. Cells were harvested and sonicated. The proteins expressed in the soluble fraction, and after centrifugation, were bound to Ni-NTA beads and purified. All single mutants were first made, then different double mutant and triple mutant

combinations containing R141K were expressed along the way. All proteins were verified by SDS-PAGE and MALDI-TOF. Circular dichroism wavelength scans probed the secondary structure of the mutants (Figure 5-16). Unfortunately, D89V/N105S/R141K (VSK) and the 5-fold mutant, D89V/N105S/R141K/D215A/Q235L (VSKAL) were not folded properly. R141K/D215A/Q235L (KAL) and the R141K single mutant both appeared folded, with intense minimums at 208nm and 222nm as is characteristic of helical proteins.

Even though our design was not folded properly, we decided to test the protein mutants we made for activity. The assay we selected was the same one used to screen for the catalytic antibodies 33F12 and 38C2. We incubated the proteins with 1,4-pentadione (acetylacetone) and looked for the vinylogous amide formation by observing UV absorption. Acetylacetone is a diketone, a smaller diketone than the hapten used to raise the antibodies. We chose this smaller diketone to ensure it could fit in the binding pocket of RBP. If a reactive Lys was present in the binding pocket, the Schiff base would have formed and equilibrated to the vinylogous amide which has a λ_{max} of 318nm. To test this method, we first assayed the commercially available 38C2. To 9 μM of antibody in PBS, we added an excess of acetylacetone, and monitored UV absorption from 200 to 400nm. UV absorption increased at 318nm within seconds of adding acetylacetone, in accordance with the formation of the vinylogous amide (Figure 5-17). This method can reliably show vinylogous amide formation, and therefore

is an easy and reliable method to determine whether the reactive Lys is in the binding pocket. We performed the catalytic assay on all the mutants, but did not observe an increase in UV absorbance at 318nm. The mutants behaved the same as wild-type RBP and R141K in the catalytic assay, which are shown in Figure 5-18. Incubation with acetone and benzaldehyde also did not lead to observation of the product by HPLC.

Discussion

As we mentioned above, RBP exists in the open conformation without ligand, and in the closed conformation with ligand. The binding pocket is more exposed to the solvent in the open conformation than in the closed conformation. It is possible that the introduced lysine is protonated in the open conformation, and the energy to deprotonate the side chain is too great. It may also be that the hapten and substrates of the aldol reaction cannot cause the conformational change to the closed conformation. This is a shortcoming of performing design calculations on one conformation when there are multiple conformations available. We can not be certain the designed conformation is the dominant structure. In this case, it is better to design on proteins with only one dominant conformation.

The shifted pK_a (~6.0) of the catalytic lysine in 33F12 is attributed to its burial in a hydrophobic microenvironment without any countercharge.[28] Observations from natural class I adolases show the presence of a second

positively charged residue in close proximity to the reactive lysine can also lower its pK_a.[29] The presence of the reactive lysine is essential to the success of the project, and we decided to introduce a lysine into the hydrophobic core of a protein.

Reactive Lysines

Buried Lysines in Literature

Studies to introduce lysine into the hydrophobic core of *E. coli* thioredoxin led to $\Delta\Delta G$ of -4 kcal*mol^{-1} and $\Delta\Delta C_p$ of approximately -1 kcal*mol^{-1}*K^{-1}.[30] The reduction in ΔC_p is attributed to structural perturbations leading to localized unfolding and the exposure of the hydrophobic core residues to solvent. Mutations of completely buried hydrophobic residues in the core of Staphylococcal nuclease to lysine have led to pK_a of 5.6 and 6.4. ΔG for the burial of the lysine costs 5-6 kcal/mol.[31, 32] The protein unfolds, however, when the lysine is protonated, except in the case of a hyperstable mutant of Staphylococcal nuclease as the background.[33] It is clear the burial of lysine in a hydrophobic environment is energetically unfavorable and costly. A compensation for the inevitable loss of stability is to use a hyperstable protein scaffold as the background for the mutation. Two proteins that fit this criteria were the tenth fibronectin type III domain (^{10}Fn3) and non-specific lipid transfer protein from maize (mLTP). We tested the burial of lysine in the hydrophobic cores of these proteins,.

Tenth Fibronectin Type III Domain

^{10}Fn3 was chosen as a protein scaffold for its exceptional thermostability (T_m = 90 °C) and because it is an antibody-mimic. Its structure is similar to that of the variable region of an antibody.[34] It is a common scaffold for directed evolution and selection studies. It has high expression in *E. coli* and is >15mg/ml soluble in aqueous solutions. We scanned the core of ^{10}Fn3 for optimal sites for the placement of Lys. For each residue that is considered "core" by RESCLASS, we set the residue to Lys, and allowed the remaining protein to retain their wild-type identities. We picked four positions for Lys placement from a visual inspection of each resulting model. They are W22, Y32, I34, and I70 (Figure 5-19). Each of the four sidechains extends into the core of the protein along the length of the protein.

The four mutants were made by site-directed mutagenesis of the ^{10}Fn3 gene and expressed in *E. coli* along with the wild-type protein for comparison. All five proteins were highly expressed, but only the wild-type protein was present in the soluble fraction and properly folded. Attempts were made to refold the four mutants from inclusion bodies by rapid-dilution, step-wise dialysis, and solubilization in buffers with various pH and ionic strength, but the proteins were not soluble. The Lys incorporation in the core had unfolded the protein.

mLTP (Non-specific Lipid-Transfer Protein from Maize)

mLTP is a small protein with four disulfide bridges that does not undergo conformational change upon ligand binding.[35] We had successfully expressed mLTP in *E. coli* previously and determined its apparent T_m to be 82 °C. It binds fatty acids and other nonpolar ligands in its deep hydrophobic binding pocket. The residues involved in ligand contact (11, 18, 33, 36, 40, 49, 53, 60, 71, 79, 83) are all classified as "core" by RESCLASS. We placed a lysine sidechain in the position of each of the ligand-binding residues and allowed the rest of the protein to retain their amino acid identity. From the 11 sidechain placement designs, we chose 5 positions to mutate to lysine, I11, A18, V33, A49 and I79 (Figure 5-20).

Encouragingly, of the five mutations, only I11K was not folded. The remaining four mutants were properly folded and had apparent T_ms above 65 °C (Figure 5-21). The four mutants were tested for reactive lysine by incubating with 1,4-pentadione as performed in the catalytic assay for 33F12, however no vinylogous amide formation was observed. It is possible that the 1,4-pentadione does not conjugate to the lysine due to inaccessibility rather than the lack of lowered pK_a. However, additional experiments such as multidimensional NMR are necessary to determine if the lysine pK_a has shifted.

Future Directions

Though we were unable to generate a protein with a reactive lysine for the aldol condensation reaction, we succeeded in placing lysine in the hydrophobic binding pocket of mLTP without destabilizing the protein irrevocably. The resulting mLTP mutants can be further designed for additional mutations to lower the pK_a of the lysine side chains.

While protein design with ORBIT has been successful in generating highly stable proteins and novel proteins to catalyze simple reactions, it has not been very successful in modeling the more complicated aldolase enzyme function. Enzymes have evolved to maintain a balance between stability and function. The energy functions currently used have been very successful for modeling protein stability, as it is dominated by van der Waal forces; however they do not adequately capture the electrostatic forces that are often the basis of enzyme function. Many enzymes use a general acid or base for catalysis, an accurate method to incorporate pK_a calculation into the design process would be very valuable. Enzyme function is also not a static event as currently modeled in ORBIT. We now know the “lock and key” hypothesis does not adequately describe enzyme-substrate interactions. Multiple side chains often interact with the substrate consecutively as the protein backbone flexes and moves. A small movement in the backbone could have large effects on the active site. Improved electrostatic energy approximations and the incorporation of dynamic backbones will contribute to the success of computational enzyme design.

References

1. Seoane, G. Enzymatic C-C bond-forming reactions in organic synthesis. *Current Organic Chemistry* 4, 283-304 (2000).
2. Nicolaou, K. C., Vourloumis, D., Winssinger, N. & Baran, P. S. The art and science of total synthesis at the dawn of the twenty-first century. *Angewandte Chemie-International Edition* 39, 44-122 (2000).
3. Bolon, D. N., Voigt, C. A. & Mayo, S. L. De novo design of biocatalysts. *Curr Opin Chem Biol* 6, 125-9. (2002).
4. Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc Natl Acad Sci U S A* 98, 14274-9 (2001).
5. Dunbrack, R. L., Jr. & Karplus, M. Backbone-dependent rotamer library for proteins. Application to side- chain prediction. *J Mol Biol* 230, 543-74. (1993).
6. Machajewski, T. D. & Wong, C. H. The catalytic asymmetric aldol reaction. *Angewandte Chemie-International Edition* 39, 1352-1374 (2000).
7. Barbas, C. F., III *et al.* Immune versus natural selection: antibody aldolases with enzymic rates but broader scope. *Science* 278, 2085-92. (1997).
8. Hoffmann, T. *et al.* Aldolase antibodies of remarkable scope. *Journal of the American Chemical Society* 120, 2768-2779 (1998).

9. Wagner, J., Lerner, R. A. & Barbas, C. F., 3rd. Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. *Science* 270, 1797-800. (1995).

10. Mathews, C. K. & Van Holde, K. E. *Biochemistry* (Menlo Park, CA. The Benjamin/Cummings Publishing Company, Inc., 1996).

11. Sinha, S. C., Sun, J., Miller, G., Barbas, C. F., 3rd & Lerner, R. A. Sets of aldolase antibodies with antipodal reactivities. Formal synthesis of epothilone E by large-scale antibody-catalyzed resolution of thiazole aldol. *Org Lett* 1, 1623-6. (1999).

12. List, B., Lerner, R. A. & Barbas, C. F., 3rd. Enantioselective aldol cyclodehydrations catalyzed by antibody 38C2. *Org Lett* 1, 59-61. (1999).

13. Bahmanyar, S. & Houk, K. N. Transition states of amine-catalyzed aldol reactions involving enamine interdemiates: Theoretical studies of mechanism, reactivity, and stereoselectivity. *Journal of the American Chemical Society* 123, 11273-11283 (2001).

14. Sakthivel, K., Notz, W., Bui, T. & Barbas III, C. F. Amino acid catalyzed direct asymmetric aldol reactions: A bioorganic approach to catalytic asymmetric carbon-carbon bond-forming reactions. *Journal of the American Chemical Society* 123, 5260-5267 (2001).

15. List, B., Lerner, R. A. & Barbas III, C. F. Proline-catalyzed direct asymmetric aldol reactions. *Journal of the American Chemical Society* 122, 2395-2396 (2000).

16. Hennig, M. *et al.* A TIM barrel protein without enzymatic activity? Crystal-structure of narbonin at 1.8 A resolution. *FEBS Lett* 306, 80-4 (1992).

17. Dwyer, M. A., Looger, L. L. & Hellinga, H. W. Computational design of a biologically active enzyme. *Science* 304, 1967-71 (2004).

18. De Lorimier, R. M. *et al.* Construction of a fluorescent biosensor family. *Protein Science* 11, 2655-2675 (2002).

19. Borchert, T. V., Abagyan, R., Jaenicke, R. & Wierenga, R. K. Design, creation, and characterization of a stable, monomeric triosephosphate isomerase. *Proc Natl Acad Sci U S A* 91, 1515-8 (1994).

20. Wierenga, R. K., Noble, M. E., Vriend, G., Nauche, S. & Hol, W. G. Refined 1.83 A structure of trypanosomal triosephosphate isomerase crystallized in the presence of 2.4 M-ammonium sulphate. A comparison with the structure of the trypanosomal triosephosphate isomerase-glycerol-3-phosphate complex. *J Mol Biol* 220, 995-1015. (1991).

21. Alexov, E. G. & Gunner, M. R. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys J* 72, 2075-93 (1997).

22. Alexov, E. G. & Gunner, M. R. Calculated protein and proton motions coupled to electron transfer: electron transfer from QA- to QB in bacterial photosynthetic reaction centers. *Biochemistry* 38, 8253-70 (1999).

23. Georgescu, R. E., Alexov, E. G. & Gunner, M. R. Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins. *Biophys J* 83, 1731-48 (2002).

24. Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* 268, 1144-9 (1995).

25. Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K. & Honig, B. On the calculation of pKas in proteins. *Proteins* 15, 252-65 (1993).

26. Allard, J., Grochulski, P. & Sygusch, J. Covalent intermediate trapped in 2-keto-3-deoxy-6- phosphogluconate (KDPG) aldolase structure at 1.95- Å resolution. *Proc Natl Acad Sci U S A* 98, 3679-84 (2001).

27. Bjorkman, A. J. & Mowbray, S. L. Multiple open forms of ribose-binding protein trace the path of its conformational change. *Journal of Molecular Biology* 279, 651-664 (1998).

28. Zhu, X. *et al.* The origin of enantioselectivity in aldolase antibodies: crystal structure, site-directed mutagenesis, and computational analysis. *J Mol Biol* 343, 1269-80 (2004).

29. Heine, A., Luz, J. G., Wong, C. H. & Wilson, I. A. Analysis of the class I aldolase binding site architecture based on the crystal structure of 2-deoxyribose-5-phosphate aldolase at 0.99Å resolution. *J Mol Biol* 343, 1019-34 (2004).

30. Ladbury, J. E., Wynn, R., Thomson, J. A. & Sturtevant, J. M. Substitution of charged residues into the hydrophobic core of Escherichia coli

thioredoxin results in a change in heat capacity of the native protein. *Biochemistry* 34, 2148-52 (1995).

31. Stites, W. E., Gittis, A. G., Lattman, E. E. & Shortle, D. In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core. *J Mol Biol* 221, 7-14 (1991).

32. Nguyen, D. M., Leila Reynald, R., Gittis, A. G. & Lattman, E. E. X-ray and thermodynamic studies of staphylococcal nuclease variants I92E and I92K: insights into polarity of the protein interior. *J Mol Biol* 341, 565-74 (2004).

33. Fitch, C. A. *et al.* Experimental pK(a) values of buried residues: analysis with continuum methods and role of water penetration. *Biophys J* 82, 3289-304 (2002).

34. Xu, L. *et al.* Directed evolution of high-affinity antibody mimics using mRNA display. *Chem Biol* 9, 933-42 (2002).

35. Shin, D. H., Lee, J. Y., Hwang, K. Y., Kyu Kim, K. & Suh, S. W. High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure* 3, 189-199 (1995).

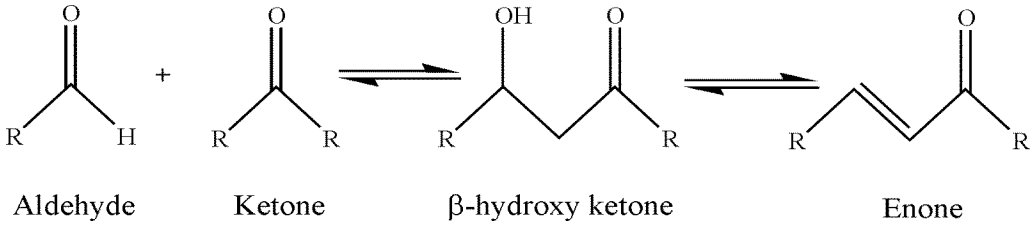


Figure 5-1. A generalized aldol reaction. The aldol condensation reaction of an aldehyde and ketone to form an enone. The hydroxy ketone can be acid or base catalyzed to form the enone.

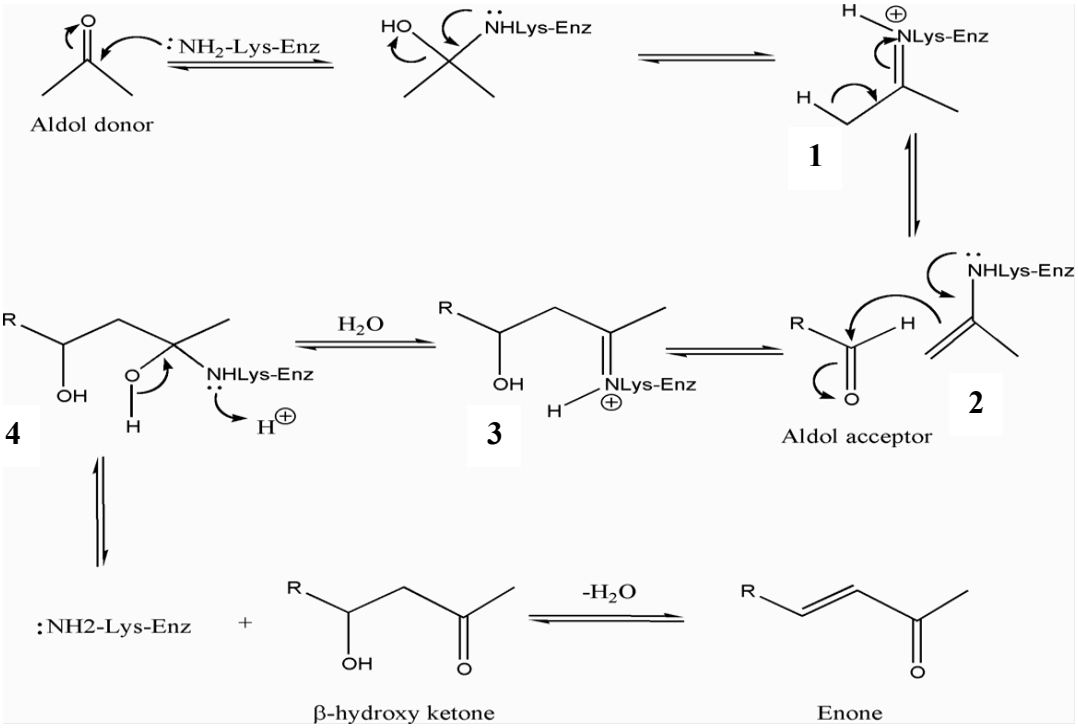


Figure 5-2. The enamine mechanism of catalytic antibody aldolases and natural class I aldolases. Acetone is shown as the aldol donor, though it can be substituted by other ketones or aldehydes. (Figure from Barbas *et al.,* Science, 1997)[7].

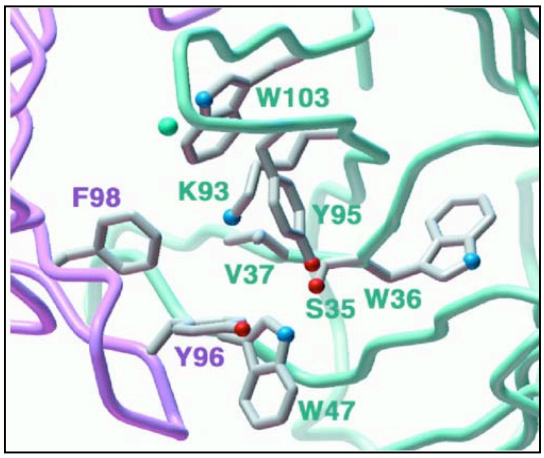


Figure 5-3. Fab' 33F12 binding site. Side chains for residues within 4 Å of Lys^{H93} are shown. The light chain is in purple, and heavy chain in green. (Figure from Barbas *et al.,* Science, 1997)[7].

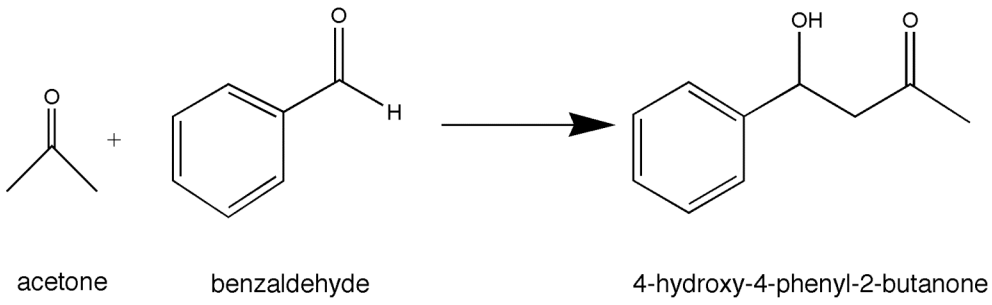


Figure 5-4. The target aldol addition between acetone and benzaldehyde. The product has one stereocenter at the carbon with the hydroxyl group.

Table 5-1 Catalytic parameters of proline and catalytic antibodies.

Parameters for the aldol reaction shown in Figure 5-4.

Catalyst	Yield	ee[1] (%)	Amt. used	K_{cat}/K_{uncat}	Reference
(L)-Proline	62%	60	20-30 mol %	N/A	Sakthivel *et al.* 2001[14]
38C2 and 33F12	67-82%	>99	0.4 mol %	10^5 - 10^7	Hoffmann *et al.* 1998[8]

[1]ee: enantiomeric excess (%) is calculated as: ee = ([A] – [B]) / ([A] + [B]) * 100 where [A] is the concentration of major enantiomer and [B] the concentration of minor enantiomer

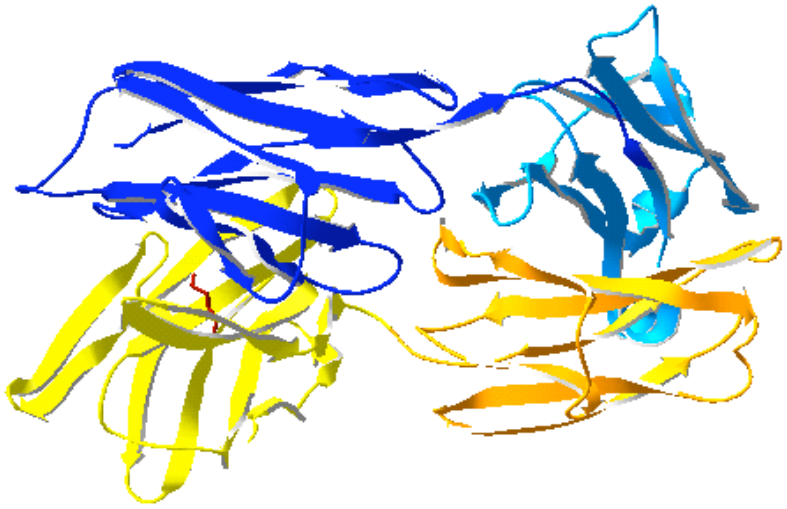
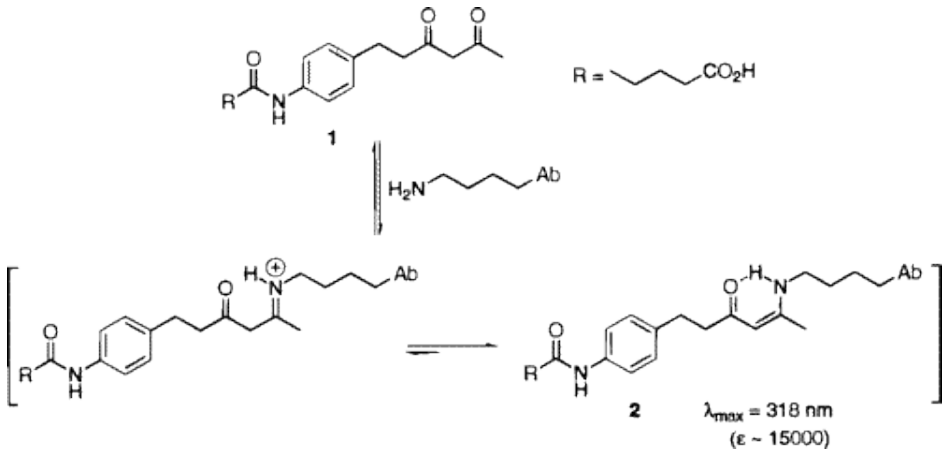


Figure 5-5. Structure of Fab' 33F12. The light chain is in dark and light blue, and heavy chain is in yellow and orange. Residues 1-114 of light chain (dark blue) and heavy chain (yellow) were scanned. Light blue and orange portions were treated as "template,", their conformations were not allowed to change. Side chain of Lys^{H93} is shown in red.

a.



b.

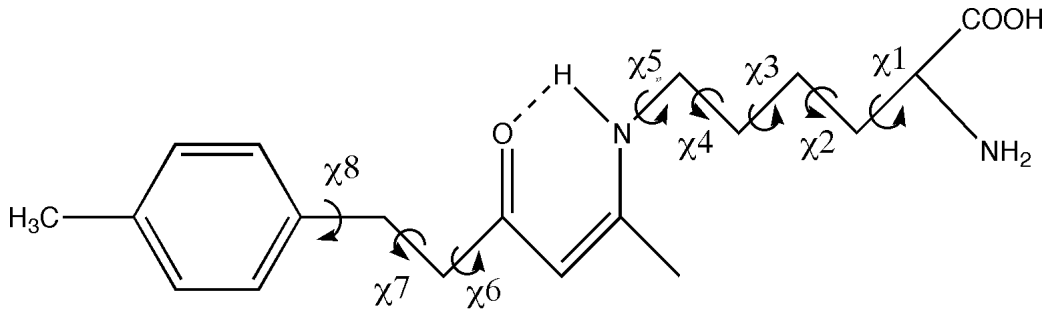


Figure 5-6. Hapten-like rotamers for active site scan on 33F12. a. Suggested mechanism of the β-diketone hapten **1** trapping the reactive lysine of the antibody, to form a β-keto imine that finally tautomerizes into a stable enaminone **2**, which absorbs with λmax at 318nm. (Figure from Hoffmann *et al.*, JACS, 1998)[8] **b.** The hapten-like rotamer used to test the active site scan on 33F12. Labelled dihedral angles were varied. The R group was shorted to methyl group for ease of design calculations.

Sorted by Residue Energy

Rank	ASresidue	residueE	totalE	#mutations	%b-H	%b-P	%b-T
1	70L	-23.51	-998.37	4	56.2	76.9	57.2
2	47L	-19.86	-1006.77	4	60.9	64.7	61.2
3	20L	-19.67	-816.33	4	55.2	94.3	56.7
4	33H	-19.02	-1100.37	9	93.6	85.2	93.1
5	59H	-18.23	-1012.46	12	68.4	89.1	69.3
6	108L	-18.08	-1047.35	4	77.4	99.6	79.4
7	4L	-16.88	-704.29	8	76.4	90.3	77
8	83L	-16.11	-788.06	14	93.2	98.2	93.7
9	50L	-15.00	-1058.12	7	76.7	100	78.4
10	52H	-14.55	-1059.69	9	82	81.1	81.9

Sorted by Total Energy

Rank	ASresidue	residueE	totalE	#mutations	%b-H	%b-P	%b-T
1	33H	-19.02	-1100.37	9	93.6	85.2	93.1
2	99H	-12.39	-1095.02	7	97.3	100	97.6
3	101H	8.27	-1070.34	8	74.5	98.2	76.6
4	52H	-14.55	-1059.69	9	82	81.1	81.9
5	50L	-15.00	-1058.12	7	76.7	100	78.4
6	106H	-8.28	-1057.25	8	97.9	86	97.6
7	35L	-4.30	-1053.43	8	74.2	63.4	73.3
8	37L	-10.99	-1050.56	8	91.2	97.4	91.9
9	108L	-18.08	-1047.35	4	77.4	99.6	79.4
10	37H	-7.19	-1041.70	10	98	98	98

Table 5-2. Top 10 results from active site scan of the Fab' antigen-binding region of 33F12 with hapten-like rotamer. Results are sorted by residue energy of the hapten-like rotamer at the active site, or total energy of the molecule. ASresidue: active site residue; %b-P: fraction polar burial of rotamer; %b-T: fraction total burial of rotamer; #mutations: number of mutations ORBIT predicts. The energies (kcal mol^{-1}) are calculated by ORBIT using the DREIDING force field. They are not absolute energies. The natural active site residue is highlighted in yellow.

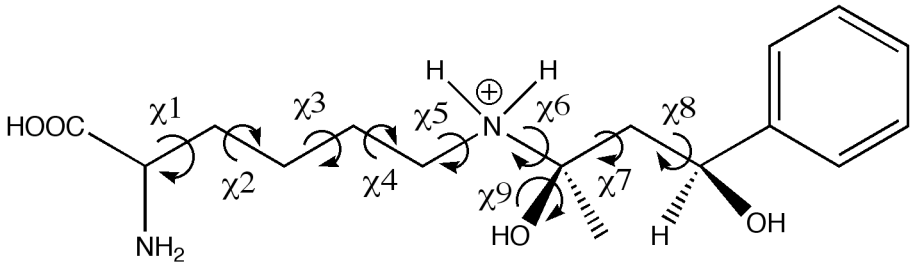


Figure 5-7. High-energy state rotamer with varied dihedral angles labeled.

One of the four high-energy state rotamer used in the design process. Labeled dihedral angles were varied to generate the series of rotamers.

Sorting by Residue Energy

Rank	ASresidue	residueE	totalE	#mutations	%b-H	%b-P	%b-T
1	70H	-30.62	-920.05	7	100	100	100
2	99H	-27.22	-1106.87	9	100	100	100
3	20H	-25.54	-935.15	8	99.5	100	99.6
4	70L	-23.51	-1008.38	3	64.6	8.1	49.8
5	36L	-23.00	-981.14	7	65.3	18.2	58.4
6	85L	-21.55	-941.74	2	47.2	47.2	47.2
7	20L	-19.67	-1023.74	3	47.2	11.9	41.5
8	108H	-19.43	-1047.85	8	100	100	100
9	47L	-19.04	-989.36	4	75.7	21.6	63.5
10	52H	-18.59	-1073.68	7	80.1	14.2	68.3

Sorting by Total Energy

Rank	ASresidue	residueE	totalE	#mutations	%b-H	%b-P	%b-T
1	99H	-27.22	-1106.87	9	100	100	100
2	39L	-11.85	-1102.94	5	98.6	100	98.8
3	35H	-13.31	-1078.62	11	99	81.8	95.8
4	52H	-18.59	-1073.68	7	80.1	14.2	68.3
5	41L	14.12	-1070.09	7	99.9	99.3	99.8
6	33H	-15.93	-1066.01	7	91.5	67.5	87.3
7	94L	-12.74	-1064.84	8	97.6	100	98
8	50L	-15.00	-1058.85	6	77.2	35.6	68.5
9	106H	-18.14	-1050.84	8	98.5	100	98.7
10	108H	-19.43	-1047.85	8	100	100	100

Table 5-3. Top 10 results from active site scan of the Fab' antigen-binding region of 33F12 with HESR. Results are sorted by residue energy of the hapten-like rotamer at the active site, or total energy of the molecule. ASresidue: active site residue; %b-P: fraction polar burial of rotamer; %b-T: fraction total burial of rotamer; #mutations: number of mutations ORBIT predicts. The energies (kcal mol^{-1}) are calculated by ORBIT using the DREIDING force field. They are not absolute energies. The natural active site residue is highlighted in yellow.

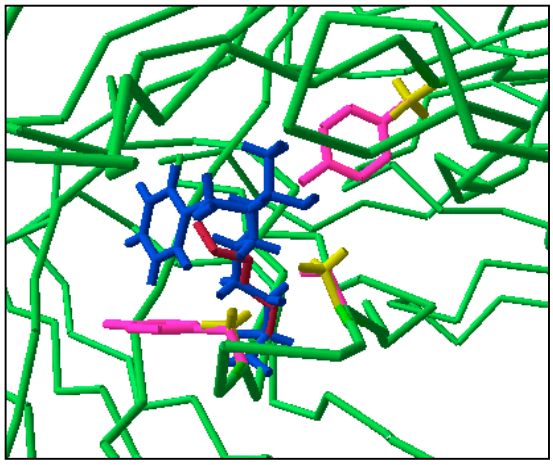


Figure 5-8. Superposition of 1AXT with the modeled protein. The C_α trace is shown in green. Lys^{H93} is in red, HESR (H99 in model) is in blue. χ_1 through χ_3 of the two side chains are approximately the same. The three additional mutations suggested by ORBIT are Tyr^{L36}, Tyr^{H95}, Ser^{H100} to Ala. The wild type side chains are shown in magenta, and Ala mutations in yellow.

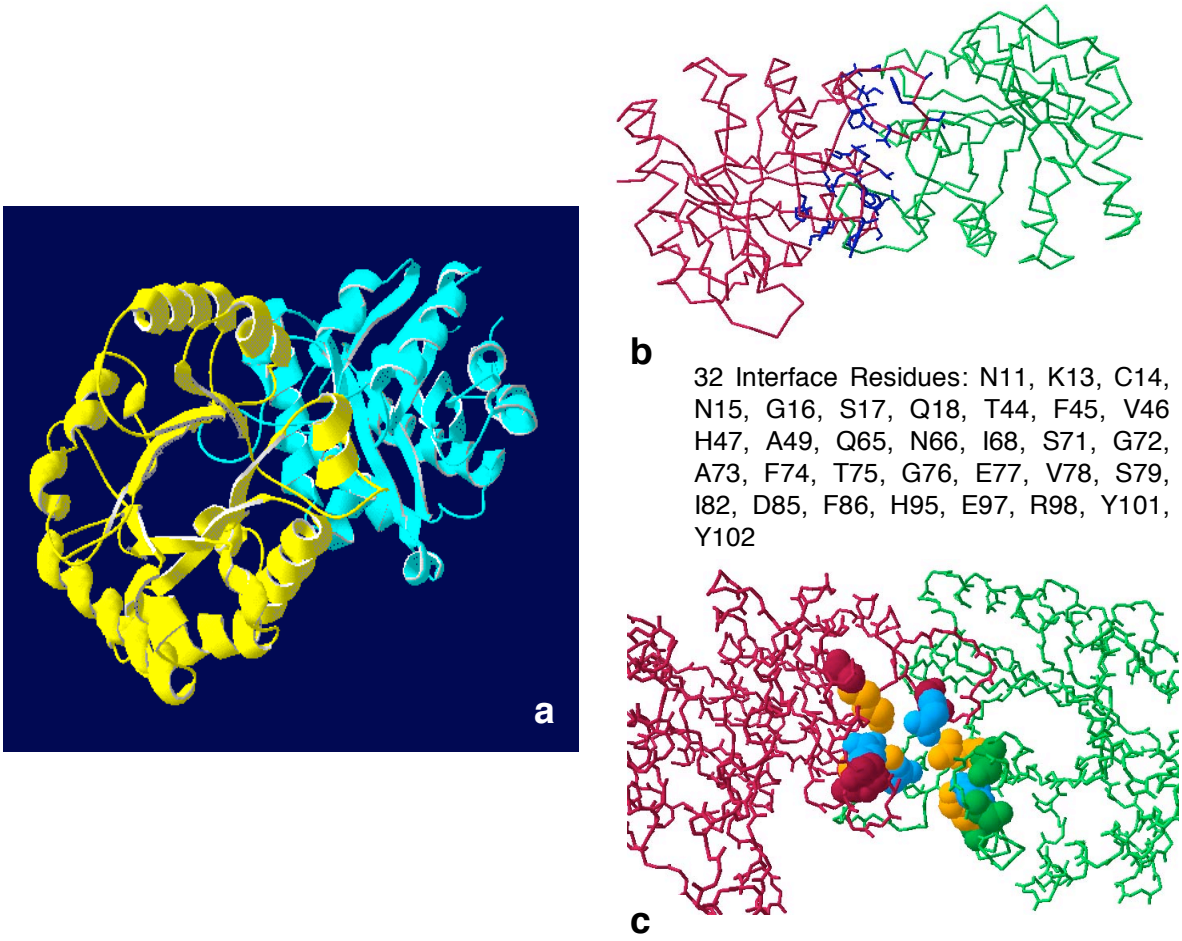


Figure 5-9. Ribbon diagram and Cα trace of triosephosphate isomerase.

Crystal structure of 5TIM showing the prototypical $(\alpha/\beta)_8$ barrel fold. **a**. Subunit A is shown in yellow, subunit B in cyan. **b.** Cα trace of both subunits, with the 32 interface residue sidechains shown in blue. The interweaving loops are easy to distinguish. A red loop inserts into the green subunit, and vice versa. **c**. The interface salt bridge network involving Glu 77, Glu 104, Arg 98, Lys 112. Anionic sidechains are in blue, cationic side chains in orange. Backbone atoms are in red and green.

Hapten-like Rotamer Library

Sorting by Residue Energy

Rank	ASresidue	residueE	totalE	#mutations	%b-H	%b-P	%b-T
1	38	-22.41	-1371.34	6	67.5	34.6	65
2	162	-18.82	-1287.05	10	99.7	94.7	99.3
3	61	-17.84	-1363.4	6	73.7	69.1	73.3
4	104	-16.94	-1336.55	4	85.4	97.7	86.2
5	130	-12.08	-1337.31	6	67.8	99.6	71.1
6	232	-11.1	-1358.49	8	83.9	100	84.8
7	178	-10.87	-1355.94	6	77.1	92.1	78.4
8	176	-9.16	-1284.61	5	65	88.1	66.6
9	122	-8.92	-1335.61	8	69.9	63.9	69.5
10	215	-8.77	-1311.79	3	70.1	79.3	70.8

Sorting by Total Energy

Rank	ASresidue	residueE	totalE	#mutations	%b-H	%b-P	%b-T
1	38	-22.41	-1371.34	6	67.5	34.6	65
2	61	-17.84	-1363.4	6	73.7	69.1	73.3
3	232	-11.1	-1358.49	8	83.9	100	84.8
4	178	-10.87	-1355.94	6	77.1	92.1	78.4
5	55	-0.25	-1348.79	5	57.4	85	59.2
6	31	-3.68	-1345.92	2	59.7	100	63.6
7	5	-5.16	-1344.64	3	68.7	33.3	65.2
8	250	-3.31	-1340.65	3	54.7	24	53.3
9	130	-12.08	-1337.31	6	67.8	99.6	71.1
10	104	-16.94	-1336.55	4	85.4	97.7	86.2

Table 5-4. Top 10 results from active site scan of the open conformation of TIM with hapten-like rotamers. Results are sorted by residue energy of the hapten-like rotamer at the active site, or total energy of the molecule. ASresidue: active site residue; %b-P: fraction polar burial of rotamer; %b-T: fraction total burial of rotamer; #mutations: number of mutations ORBIT predicts. The energies (kcal mol^{-1}) are calculated by ORBIT using the DREIDING force field. They are not absolute energies. Residues that are returned in both lists are highlighted in yellow.

Benzal Library (HESR)

Sorted by Residue Energy

Rank	ASresidue	residueE	totalE	#mutations	%b-H	%b-P	%b-T
1	242	-39.36	-1339.86	10	100	100	100
2	150	-35.09	-1322.73	8	100	100	100
3	154	-32.94	-1323.87	6	100	100	100
4	51	-24.05	-1333.91	9	100	100	100
5	162	-23.92	-1332.6	8	99.9	100	99.9
6	38	-23.04	-1342.78	4	84.1	58.5	78.3
7	10	-20.78	-1310.41	9	100	100	100
8	246	-20.69	-1299.04	10	100	100	100
9	52	-19.66	-1335.85	4	64.7	29.8	55.1
10	125	-19.58	-1307.44	7	93.1	100	94.3

Sorted by Total Energy

Rank	ASresidue	residueE	totalE	#mutations	%b-H	%b-P	%b-T
1	145	-7.04	-1372.96	5	61	13.2	50
2	179	-5.92	-1368.23	4	82	27.5	72.8
3	5	-17.58	-1365.37	5	64.1	8.5	52.2
4	106	-11.71	-1364.67	5	71.4	12.4	61.9
5	182	-17.52	-1363.92	4	81.2	17.3	70.7
6	185	-11	-1361.87	5	63.1	42.4	59
7	148	-5.78	-1357.62	4	50.7	0.8	40.8
8	55	-10.57	-1356.58	5	66.6	25.2	58.4
9	118	-8.77	-1352.98	3	68.5	7	55.9
10	122	-2.31	-1351.16	4	64.7	39.6	58.9

Table 5-5. Top 10 results from active site scan of the open conformation of TIM with HESR. Results are sorted by residue energy of the hapten-like rotamer at the active site, or total energy of the molecule. ASresidue: active site residue; %b-P: fraction polar burial of rotamer; %b-T: fraction total burial of rotamer; #mutations: number of mutations ORBIT predicts. The energies (kcal mol^{-1}) are calculated by ORBIT using the DREIDING force field. They are not absolute energies. Residues that are returned in both scans with HESR and scans with hapten-like romaters are highlighted in light yellow.

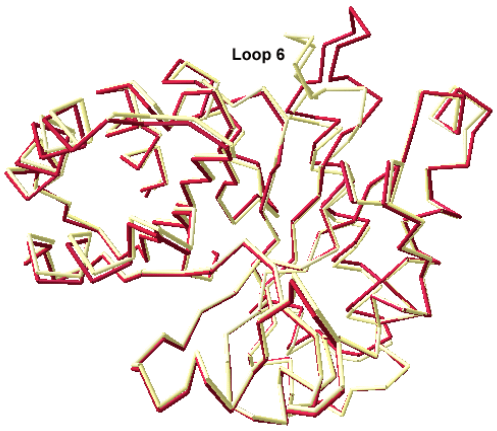


Figure 5-10. Superposition of backbone atoms of "open" and "almost closed" conformations of TIM. Cα trace is shown for each subunit. "Open" conformation (subunit A) is shown in red, and "almost closed" conformation (subunit B) is in yellow. Loop 6 on subunit B folds to trap a sulfate ion.

Benzal Library (HESR)
Sorting by Residue Energy

Rank	ASresidue	residueE	totalE	#mutations	%b-H	%b-P	%b-T
1	242	-36.91	-1346.72	10	100.0	99.8	99.9
2	21	-31.56	-1287.37	10	99.5	99.9	99.6
3	150	-31.11	-1354.54	7	100.0	100.0	100.0
4	154	-27.6	-1335.81	8	100.0	100.0	100.0
5	142	-23.7	-1391.89	4	82.5	54.0	75.3
6	246	-22.46	-1305.21	9	100.0	99.7	99.9
7	28	-22.41	-1344.82	10	99.1	100.0	99.2
8	194	-21.99	-1301.1	8	100.0	100.0	100.0
9	147	-21.51	-1334.22	10	100.0	100.0	100.0
10	164	-21.29	-1342.59	9	100.0	100.0	100.0

Sorting by Total Energy

Rank	ASresidue	residueE	totalE	#mutations	%b-H	%b-P	%b-T
1	146	-13.91	-1419.67	5	68.4	70.6	68.8
2	191	-13.88	-1414.36	2	67.0	38.8	61.2
3	148	-7.92	-1411.45	4	58.9	2.5	46.8
4	145	-9.22	-1405.24	4	63.6	11.4	53.8
5	111	-16.47	-1397.32	5	82.9	25.0	72.9
6	185	-8.55	-1397.06	3	80.3	34.8	71.0
7	55	-17.24	-1395.29	4	74.8	49.7	68.8
8	38	-14.03	-1394.82	5	76.4	15.1	63.8
9	115	-8.06	-1394.22	3	63.0	5.0	50.3
10	188	-2.87	-1393.53	3	59.2	10.0	50.5

Table 5-6. Top 10 results from active site scan of the almost-closed conformation of TIM with HESR. Results are sorted by residue energy of the hapten-like rotamer at the active site, or total energy of the molecule. ASresidue: active site residue; %b-P: fraction polar burial of rotamer; %b-T: fraction total burial of rotamer; #mutations: number of mutations ORBIT predicts. The energies (kcal mol^{-1}) are calculated by ORBIT using the DREIDING force field. They are not absolute energies. Residues that are highlighted have appeared in scans with HESR on the open conformation of TIM. Residues 55 and 38 have appeared in in both scans with HESR and hapten-like rotamers.

Protein	Titratable groups	pK_a^{exp}	pK_a^{calc}
Ribonuclease T_1 (9RNT)	His 40	7.9	8.5
	His 92	7.8	6.3
Phosphatidylinositol-specific phospholipase C (PI-PLC: 1GYM)	His 32	7.6	< 0.0
	His 82	6.9	7.8
	His 92	5.4	5.8
	His 227	6.9	7.3
Xylanase (1XNB)	Glu 78	4.6	7.9
	Glu 172	6.7	5.8
	His 149	< 2.3	< 0.0
	His 156	6.5	6.1
	Asp 4	3.0	3.9
	Asp 11	2.5	3.4
	Asp 83	< 2	6.1
	Asp 101	< 2	9.8
	Asp 119	3.2	1.8
	Asp 121	3.6	4.6
Cat. Ab. 33F12 (1AXT)	Lys H99	5.5	2.1

Table 5-7. Results of MCCE pK_a calculations on test proteins. Of the 17 titratable groups, 9 were within 1 pH unit of the experimentally determined pK_a (highlighted in red).

Table 5-8. **Results of modeling the HESR at Lys 13, the natural catalytic residue.** Definitions and format are same as table 5-6.

Catalytic residue	Residue energy	Total energy	#mutations	%b-H	%b-P	%b-T
13A (open)	6.5577	-240.824	19 (1)	84	73.4	82.3
13B (almost closed)	19.6671	-236.83	16 (0)	67.8	65.1	67.3

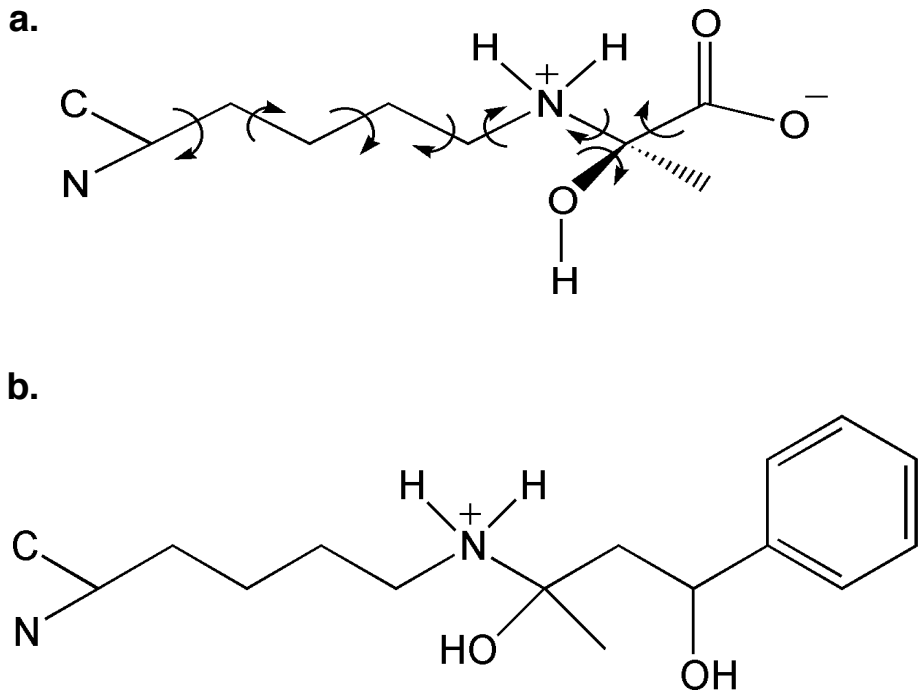


Figure 5-11. KPY rotamer and the HESR benzal rotamer. **a**. new rotamer library generated for the testing of GBIAS on KDPG aldolase. The intermediate is the carbinolamine intermediate resulting from lysine and pyruvate. The new rotamer is named KPY. Arrows indicate the dihedral angle is varied. KPY is similar to the HESR for the benzaldehyde-acetone aldol reaction (**b**).

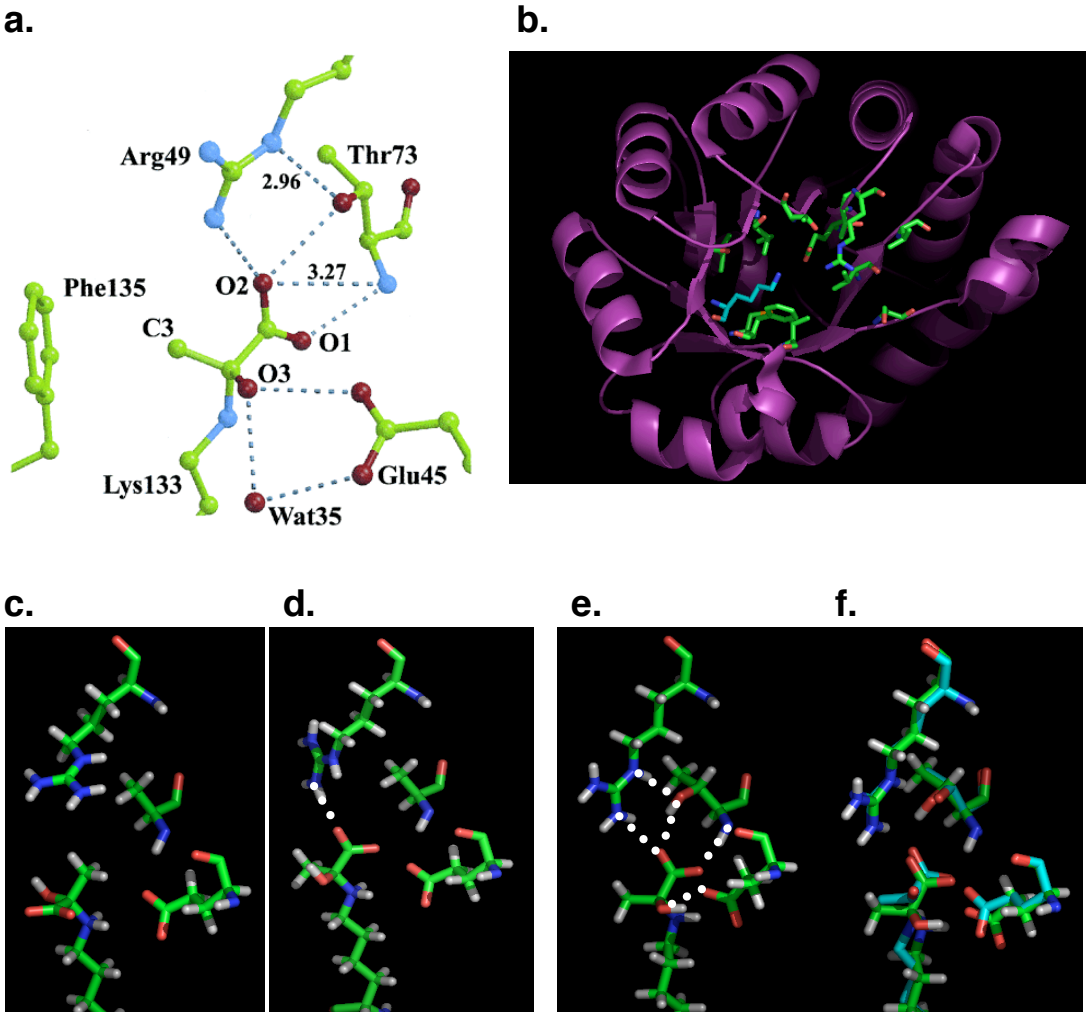
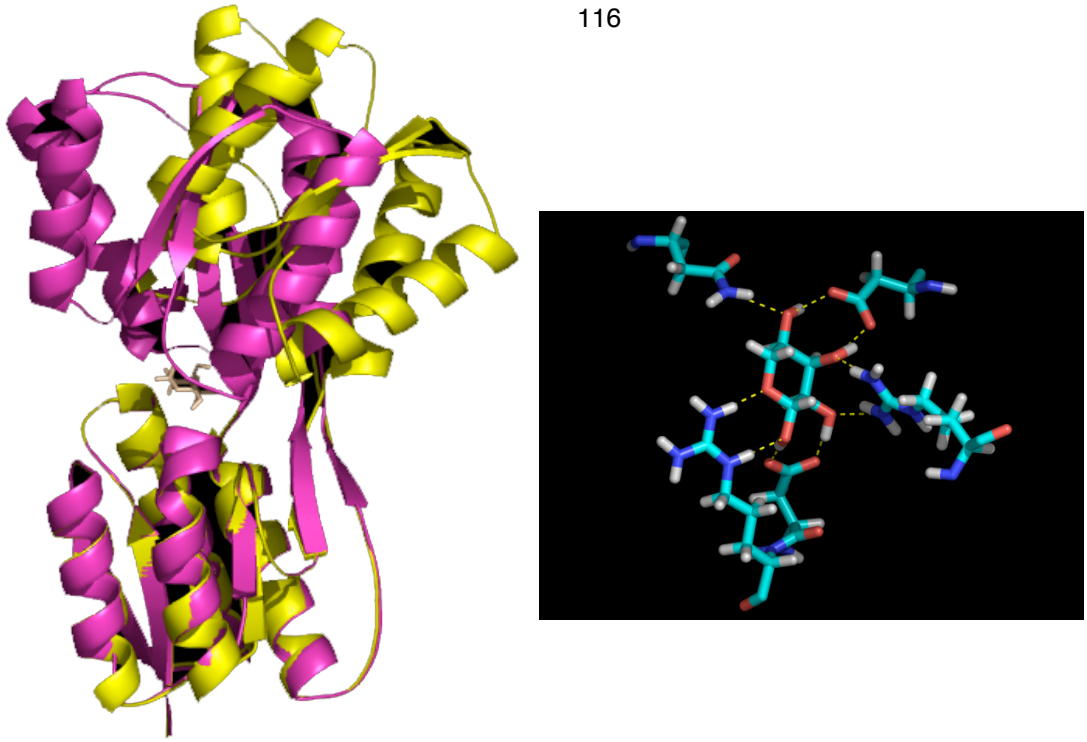


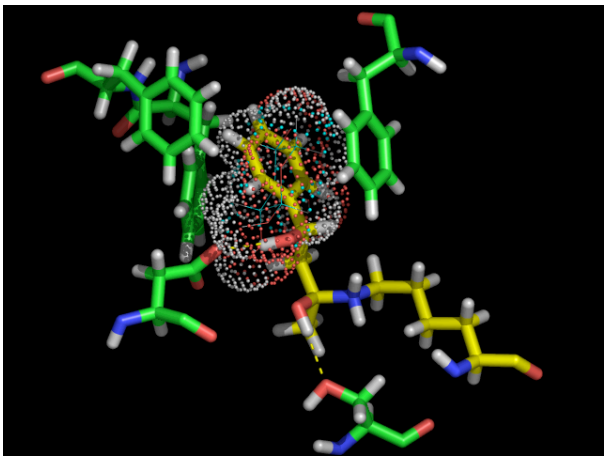
Figure 5-12. Using GBIAS to retain crystallographic hydrogen bonds in KDPG aldolase. **a.** Stick representation of the interactions of the trapped intermediate with surrounding residues. (Figure from Allard *et al.*, PNAS, 2002)[26] **b.** A subunit of KDPG aldolase used for design. Residues surrounding Lys133 were designed. **c.** Stick representation of the active site residues shown in the same orientation as in **a**. GBIAS energy=0, no hydrogen bonds retained. **d.** GBIAS energy=5, 1 hydrogen bond retained. **e.** GBIAS energy=10. Most hydrogen bonds from crystal structure are retained. **f.** Superimposition of the designed active site onto wild-type active site. KPY at 133 superimposes onto the trapped intermediate.



a. **b.**

Figure 5-13. Ribbon diagram of ribose binding protein in open and closed conformations. a. Open conformation is shown in yellow. Upon ligand binding, (ribose is shown in sticks) the two domains close in the closed conformation (magenta). The open conformation is 43º open compared to the closed form. **b.** The extensive hydrogen bond network employed to bind ribose in the RBP binding site.

a.



b.

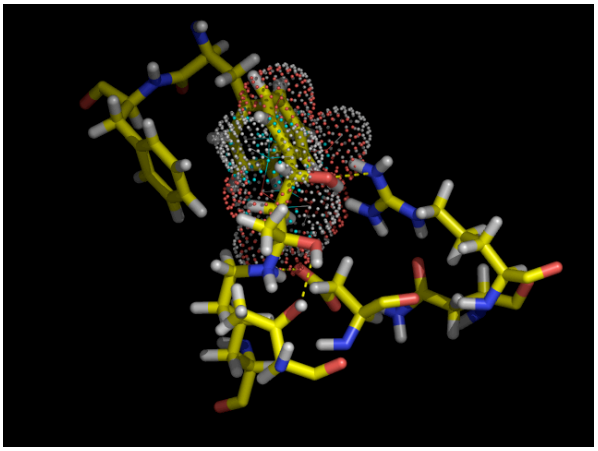


Figure 5-14. HESR in the binding pocket of RBP. a. HESR is placed in place of Arg141. **b.** HESR is placed in place of Arg90. Side chains are shown in sticks in CPK-inspired colors. The dot surface is where ribose binds in the crystal structure.

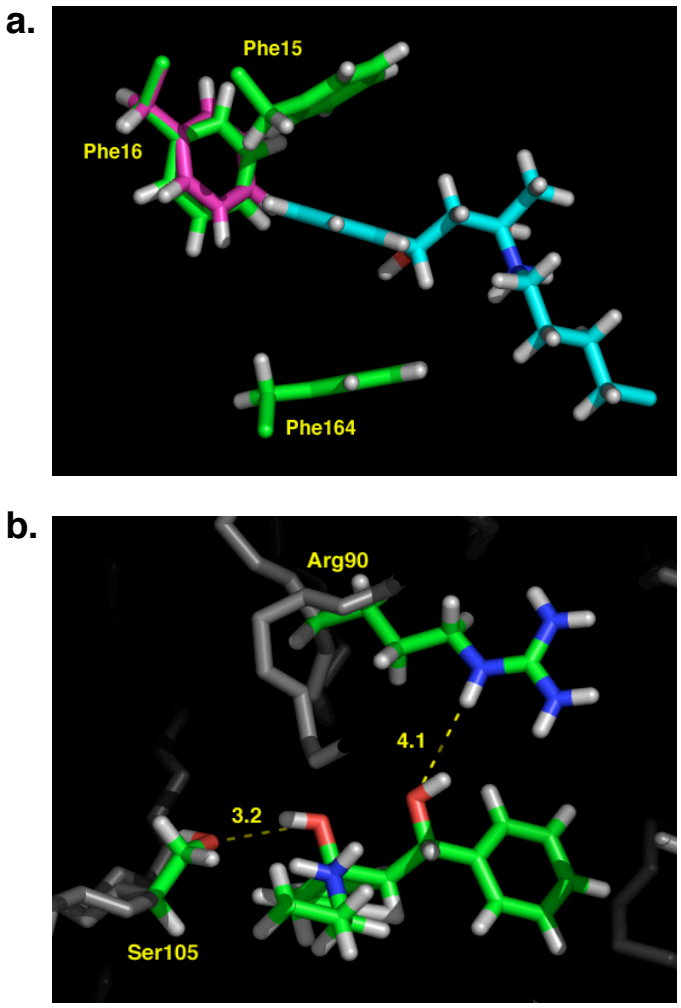


Figure 5-15. Modeled active site on RBP for aldol reaction. **a.** HESR is shown in cyan. The phenyl ring of HESR is "caged" in phenyl rings. It is stacked in between the phenyl rings of Phe15 and Phe164, and perpendicular to Phe16. **b.** The hydroxyl groups on HESR could form hydrogen bonds with Ser105 and possibly with Arg90.

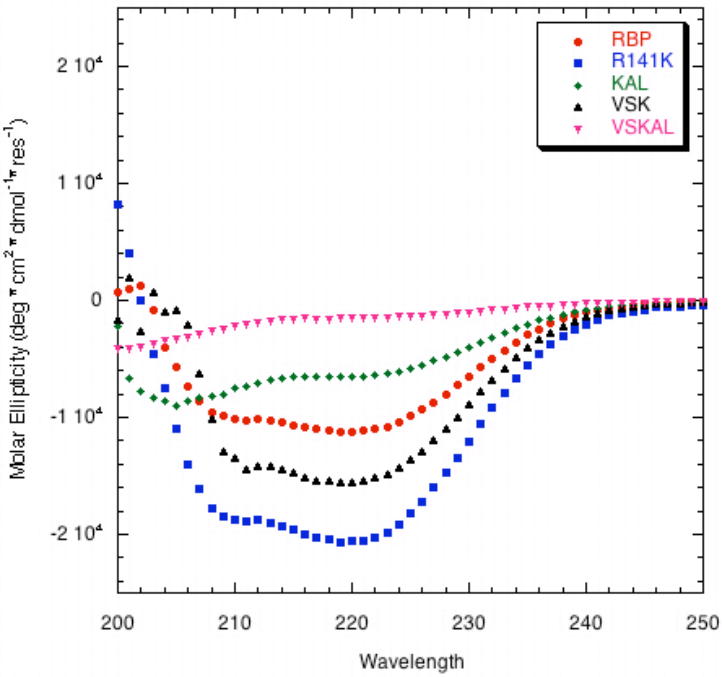


Figure 5-16. CD wavelength scan of RBP and mutants.
KAL: R141K/D215A/Q235L. VSK: D89V/N105S/R141K. VSKAL: D89V/N105S/R141K/D215A/Q235L. KAL and VSKAL do not appear to be folded correctly. R141K, VSK have more intense signal than wild-type RBP, with minimums at 208nm and 222nm as is characteristic of proteins with mostly helices.

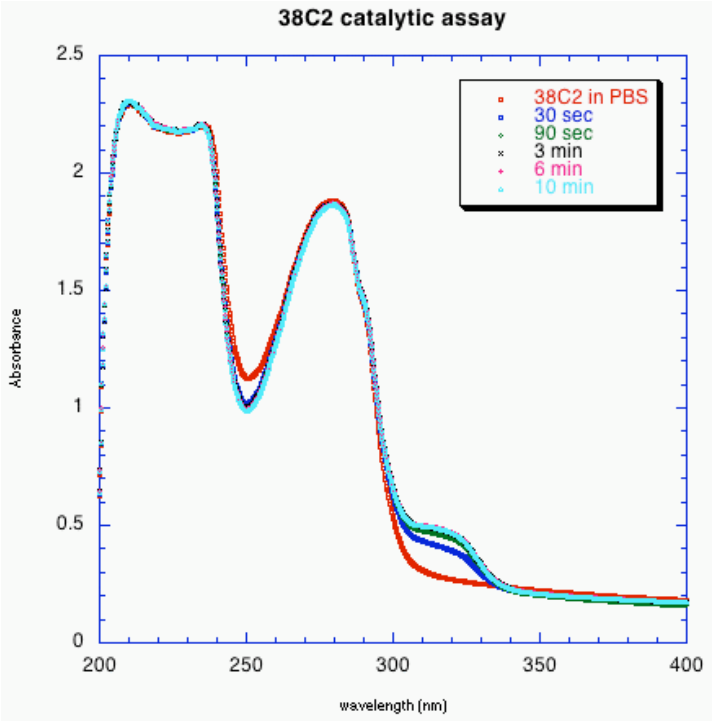


Figure 5-17. Catalytic assay of 38C2. Absorbance at 318nm increased upon addition of acetylacetone, in accordance with the formation of the vinylogous amide. Calculation of the actual binding site shows 38C2 to be 73% active.

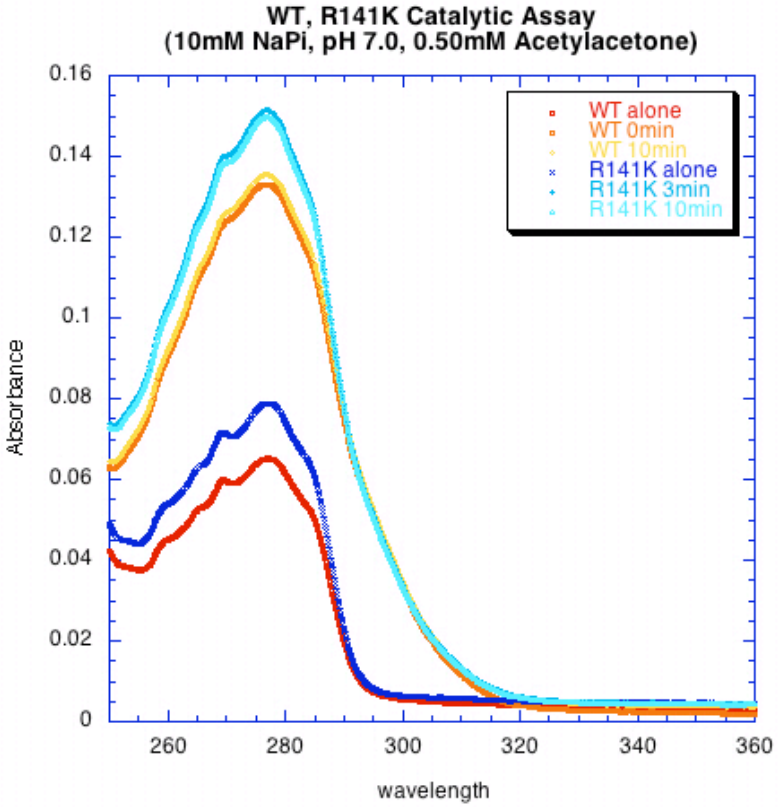


Figure 5-18. Catalytic assay of RBP and R141K. This is representative of the catalytic assays performed with the remaining mutants of RBP. No vinylogous amide formation is observed.

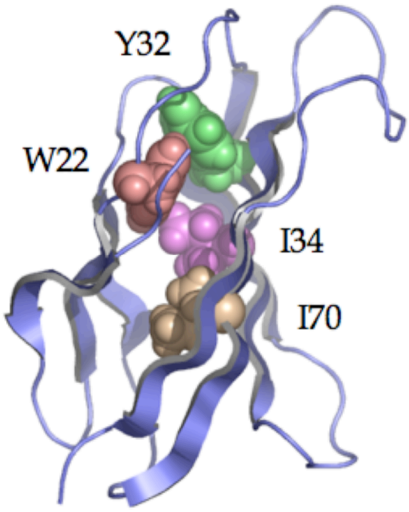


Figure 5-19. Ribbon diagram of tenth fibronectin type III domain. The four core residues, Y32, W22, I34 and I70 are shown in space filling model.

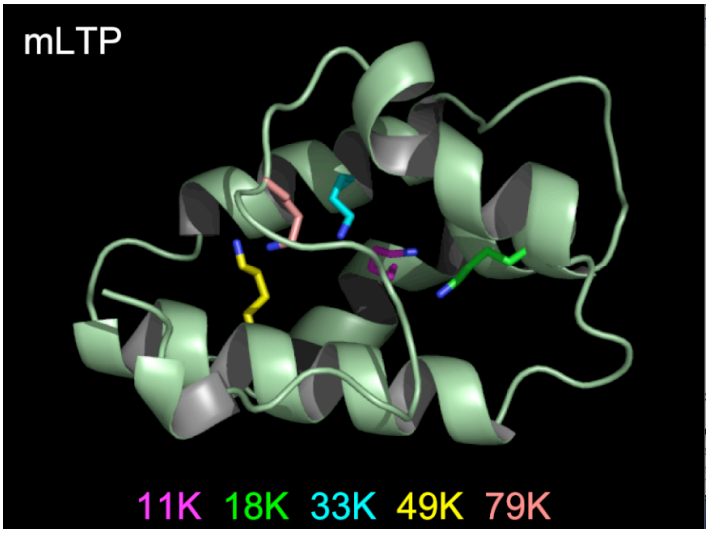


Figure 5-20. Ribbon diagram of mLTP. The five residue positions that are mutated to lysine are shown in sticks model. The Nε of the lysines are colored blue.

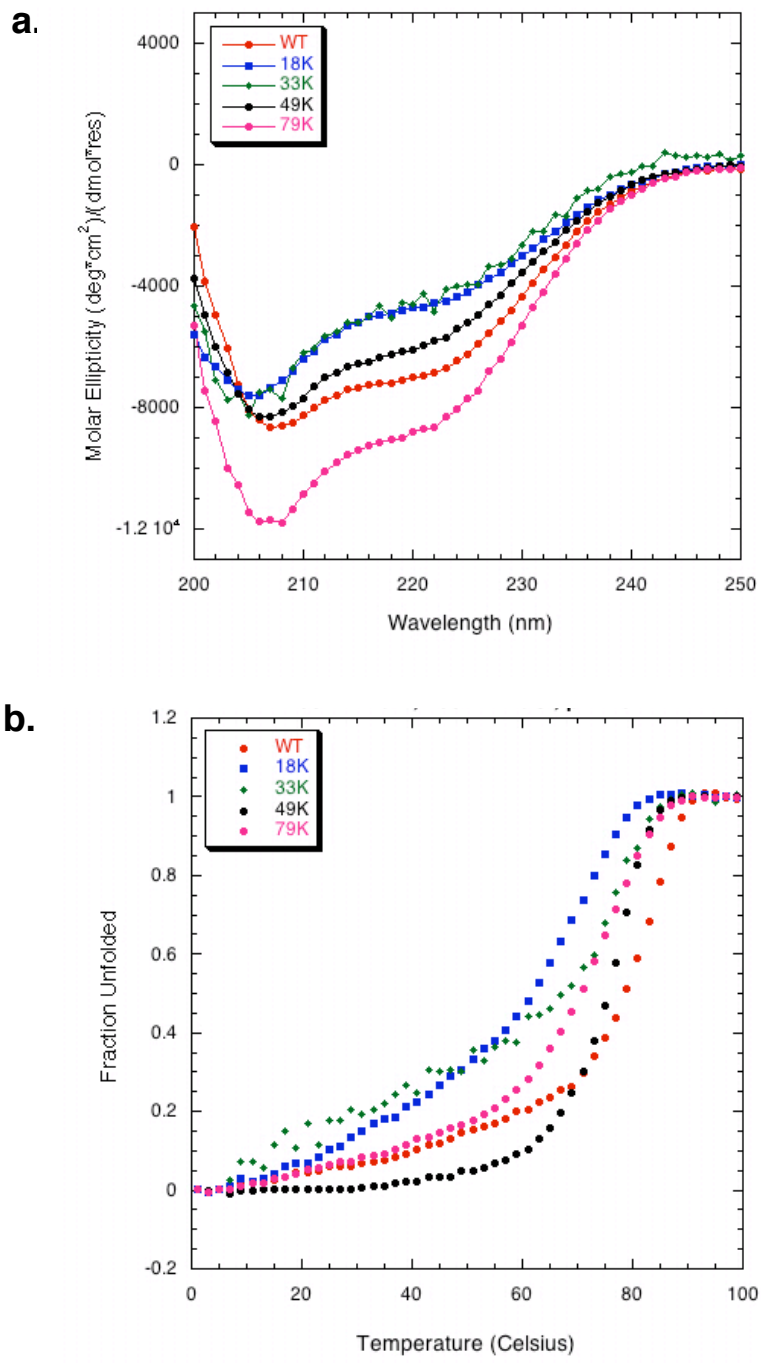


Figure 5-21. Circular dichroism spectroscopy of mLTP and mutants.
a. Wavelength scans of wild-type (WT) mLTP and the four folded mutants, 18K, 33K, 49K, and 79K. The scans show the characteristic minimus at 208nm and 222nm for helical proteins. **b.** Thermal denaturations of the five proteins. Of the mutants, 18K is most destabilized, with an apparent T_m of 74 °C. 33K: 78 °C. 49K: 78 °C. 79K: 76 °C.

Chapter 6

Double Mutant Cycle Study of Cation-π Interaction

This work was done in collaboration with Shannon Marshall.

Introduction

The marginal stability of a protein is not due to one dominant force, but to a balance of many non-covalent interactions between amino acids arising from hydrogen bonding, electrostatics, van der Waals interaction, and hydrophobic interactions.[1] These forces confer secondary and tertiary structure to proteins, allowing amino acid polymers to fold into their unique native structures. Even though hydrogen bonding is electrostatic by nature, most would think of electrostatics as the nonspecific repulsion between like charges, and the specific attraction between oppositely charged side chains, referred to as a salt bridge. The cation-π interaction is another type of specific attractive electrostatic interaction. It was experimentally validated to be a strong non-covalent interaction in the early 1980s using small molecules in the gas phase. Evidence of cation-π interactions in biological systems was provided by Burley and Petsko.[2,3] They discovered a prevalence of aromatic-aromatic and amino-aromatic interactions, and found them to be stabilizing forces.

Cation-π interactions are defined as the favorable electrostatic interactions between a positive charge and the partial negative charge of the quadrupole moment of an aromatic ring (Figure 6-1). In this view, the π system of the aromatic side chain contributes partial negative charges above and below the plane, forming a permanent quadrupole moment that interacts favorably with the positive charge. The aromatic side chains are viewed as polar, yet hydrophobic residues. Gas phase studies established the interaction energy between K^+ and

benzene to be 19 kcal mol^{-1}, even stronger than that of K^+ and water.[4] In aqueous media, the interaction is weaker.

Evidence strongly indicates this interaction is involved in many biological systems where proteins bind cationic ligands or substrates.[4] In unliganded proteins, the cation-π interaction is typically between a cationic side chain (Lys or Arg) and an aromatic side chain (Trp, Phe, or Tyr). Gallivan and Dougherty[5] used an algorithm based on distance and energy to search through a representative dataset of 593 protein crystal structures. They found that ~21% of all interacting pairs involving K, R, F, Y, and W are significant cation-π interactions. Using representative molecules, they also conducted a computational study of cation-π interactions vs. salt bridges in aqueous media. They found that the well depth of the cation-π interaction was 5.5 kcal mol^{-1} in water compared to 2.2 kcal mol^{-1} for salt bridges, even though salt bridges are much stronger in gas phase studies. The strength of the cation-π interaction in water led them to postulate that cation-π interactions would be found on protein surfaces, where they contribute to protein structure and stability. Indeed, cation-π pairs are rarely completely buried in proteins.[6]

There are six possible cation-π pairs resulting from two cationic side chains (K, R) and three aromatic side chains (W, F, Y). Of the six, the pair with the most occurrences is RW, accounting for 40% of the total cation-π interactions found in a search of the PDB database. In the same study, Gallivan and Dougherty also found that the most common interaction is between neighboring

residues, with *i* and (*i+4*) the second most common.[5] This suggests cation-π interactions can be found within α-helices. A geometry study of the interaction between R and aromatic side chains showed that the guanidinium group of the R side chain stacks directly over the plane of the aromatic ring in a parallel fashion more often than would be expected by chance.[7] In this configuration, the R side chain is anchored to the aromatic ring by the cation-π interaction, but the three nitrogen atoms of the guanidinium group are still free to form hydrogen bonds with any neighboring residues to further stabilize the protein.

In this study we seek to experimentally determine the interaction energy between a representative cation-π pair, R and W, in positions *i* and (*i+4*). This will be done using the double mutant cycle on a variant of the all α-helical protein engrailed homeodomain. The variant is a surface and core designed engrailed homeodomain (sc1) that has been extensively characterized by a former Mayo group member, Chantal Morgan.[8] It exhibits increased thermal stability over the wild type. Since cation-π pairs are rarely found in the core of the protein, we chose to place the pair on the surface of our model system.

Materials and Methods

Computational Modeling

In order to determine the optimal placement of the cation-π interacting pair, the ORBIT (Optimization of Rotamers by Iterative Techniques) suite of protein design software developed by the Mayo group was used. The

coordinates of the 56-residue engrailed homeodomain structure were obtained from PDB entry 1enh. Residues 1-5 are disordered in the absence of DNA, and thus were removed from the structure. The remaining 51 residues were renumbered, explicit hydrogens were added using the program BIOGRAF (Molecular Simulations, Inc., San Diego, California), and the resulting structure was minimized for 50 steps using the DREIDING forcefield.[9] The surface-accessible area was generated using the Connolly algorithm.[10] Residues were classified as surface, boundary, or core as described.[11]

Engrailed homeodomain is composed of three helices. We considered two sites for the cation-π interaction, residue pairs 9 and 13, and 42 and 46 (Figure 6-2). Both pairs are in the middle of their respective α-helix on the protein surface. Discrete rotamers from the Dunbrack and Karplus backbone-dependent rotamer library[12] were used to represent the side-chains. Rotamers at ±1 standard deviation about χ_1 and χ_2 were also included. Four calculations were performed at each site. For the 9 and 13 pair, R was placed at position 9, W at position 13, and the surrounding positions (*i-4, i-1, i+1, j-1, j+1, j+4* where *i*=9 and *j*=13) were mutated to A. The interaction energy was then calculated. This approach allowed the best conformations of R and W to be chosen for maximal cation-π interaction. Next, the conformations of R and W at positions 9 and 13 were held fixed, while the conformations of the surrounding residues, but not the identity, were allowed to change. This way, the interaction energy between the cation-π pair and the surrounding residues was calculated. The same

calculations were performed with W at position 9, and R at position 13, and likewise for both possibilities at sites 42 and 46.

The geometry of the cation-π pair was optimized using van der Waals interactions scaled by 0.9,[13] and electrostatic interactions were calculated using Coulomb's law with a distance-dependent dielectric of 2r. Partial atomic charges from the OPLS force field,[14] which reflect the quadropole moment of aromatic groups, were used. The interaction energies between the cation-π pair and the surrounding residues were calculated using the standard ORBIT parameters and charge set.[15] Pairwise energies were calculated using a force field containing van der Waals, Coulombic, hydrogen bond, and polar hydrogen burial penalty terms.[16] The optimal rotameric conformations were determined using the dead-end elimination (DEE) theorem with standard parameters.[17]

Of the four possible combinations at the two sites chosen, two pairs had good interaction energies between the cation-π pair, and with the surrounding residues: W42-R46 and R9-W13. A visual examination of the resulting models showed that R9-W13 exhibited optimal cation-π geometry (Figure 6-3); this pair was therefore investigated experimentally using the double-mutant cycle.

Protein Expression and Purification

For ease of expression and protein stability, sc1, the core- and surface-optimized variant of homeodomain was used instead of wild-type homeodomain. Four variants of sc1 were made for the double mutant cycle: 9A/13A, 9A/13W,

9R/13A, and 9R/13W. All variants were generated by site-directed mutagenesis using inverse PCR, and the resulting plasmids were transformed into XL1 Blue cells (Stratagene) by heat shock. The cells were grown for approximately 40 minutes at 37 ºC and plated on agarose containing ampicillin. The plasmids also contained a gene conferring ampicillin resistance, allowing only cells with successful transformations to survive. After overnight growth at 37 ºC, colonies were picked and grown in 10 ml LB with ampicillin. The plasmids were extracted from the cells, purified, and verified by DNA sequencing. Plasmids with correct sequences were then transformed into competent BL21 (DE3) cells (Stratagene) by heat shock for expression.

One liter LB with cells for each mutant was grown at 37º C to an OD of 0.6 at 600 nm. Cells were then induced with IPTG and grown for 4 hours. The recombinant proteins were isolated from cells using the freeze-thaw method[18] and purified by reverse-phase HPLC. HPLC was performed using a C8 prep column (Zorbax) and linear water-acetonitrile gradients with 0.1% trifluoroacetic acid. The identities of the proteins were checked by MALDI-TOF; all masses were within one unit of the expected weight.

Circular Dichroism (CD)

CD data were collected using an Aviv 62A DS spectropolarimeter equipped with a thermoelectric cell holder and an autotitrator. Urea denaturation data was acquired every 0.2 M from 0.0 M to 9.0 M with a 9 minute mixing time

and 100 second averaging time at 25° C. Samples contained 5 μM protein and 50 mM sodium phosphate adjusted to pH 4.5. Protein concentration was determined by UV spectrophotometry. To maintain constant pH, the urea stock solution also was adjusted to pH 4.5. Protein unfolding was monitored at 222 nm. Urea concentration was measured by refractometry. ΔG_u was calculated assuming a two-state transition and using the linear extrapolation model.[19]

Double Mutant Cycle Analysis

The strength of the cation-π interaction was calculated using the following equation:

$$\Delta G^{\text{cation-}\pi} = (\Delta G^{RW} - \Delta G^{AA}) - [(\Delta G^{RA} - \Delta G^{AA}) + (\Delta G^{AW} - \Delta G^{AA})] \quad (6\text{-}1)$$

ΔG^{RW} = free energy of unfolding of the R9/W13 mutant
ΔG^{AA} = free energy of unfolding of the A9/A13 mutant
ΔG^{RA} = free energy of unfolding of the R9/A13 mutant
ΔG^{AW} = free energy of unfolding of the A9/W13 mutant

Results and Discussion

The urea denaturation transitions of all four homeodomain variants were similar, as shown in Figure 6-4 and Table 6-1. The cation-π interaction energy, determined using the double mutant cycle, indicates that it is unfavorable on the order of 1.4 kcal mol^{-1}. However, additional factors must be considered. First, the cooperativity of the transitions, given by the *m*-value, ranges from 0.73 to 0.91 kcal mol^{-1} M^{-1}. The low *m*-values suggest that the transitions may not be two state. Therefore, free energies calculated assuming a two-state transition may

not be accurate, affecting the interaction energy calculated from the double mutant cycle.[20] Second, the urea denaturation curves for all four variants lack a well-defined post-transition, which makes fitting of the experimental data to a two-state model difficult.

In addition to low cooperativity, analysis of the surrounding residues of Arg and Trp provided further insight. In the sc1 variant, the (*i-4, i-1, i+1, j-1, j+1,* and *j+4*) residues are E, K, R, E, E, and R, respectively. R9 and W13 are in a very charged environment. In the R9/W13 variant, the cation-π interaction is in conflict with the local interactions that R9 and W13 can form with E5 and R17. The double mutant cycle is not appropriate for determining an isolated interaction in a charged environment. The charged residues surrounding R9 and W13 need to be mutated to provide a neutral environment.

The cation-π interaction introduced to homeodomain mutant sc1 does not contribute to protein stability. Several improvements can be made for future studies. First, since sc1 is the experimental system, the sc1 sequence should be used in the modeling studies. Second, to achieve a well-defined post-transition, urea denaturations could be performed at a higher temperature. pH of protein could be adjusted to 7.0 instead of 4.5. Because sc1 is a stable protein, perhaps the 9 minute mixing time with denaturant is not long enough to reach equilibrium. Longer mixing times could be tried. Third, the immediate surrounding residues of the cation-π pair can be mutated to Ala to provide a neutral environment to

isolate the interaction. This way, the interaction energy of a cation-π pair can be accurately determined.

References

1. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133-55 (1990).
2. Burley, S. K. & Petsko, G. A. Amino-Aromatic Interactions in Proteins. *Febs Letters* **203**, 139-143 (1986).
3. Burley, S. K. & Petsko, G. A. Aromatic-Aromatic Interaction - a Mechanism of Protein- Structure Stabilization. *Science* **229**, 23-28 (1985).
4. Ma, J. C. & Dougherty, D. A. The Cation-π Interaction. *Chem Rev* **97**, 1303-1324 (1997).
5. Gallivan, J. P. & Dougherty, D. A. Cation- π interactions in structural biology. *PNAS* **96**, 9459-9464 (1999).
6. Gallivan, J. P. & Dougherty, D. A. A computation study of Cation-π interations vs salt bridges in aqueous media: Implications for protein engineering. *JACS* **122**, 870-874 (2000).
7. Flocco, M. M. & Mowbray, S. L. Planar stacking interactions of arginine and aromatic side-chains in proteins. *J Mol Biol* **235**, 709-17. (1994).
8. Morgan, C. Ph.D. Thesis, California Institute of Technology (2000).
9. Mayo, S. L., Olafson, B. D. & Goddard III, W. A. DREIDING: A generic force field for molecular simulations. *J Phys Chem* **94**, 8897-8909 (1990).
10. Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709-713 (1983).

11. Marshall, S. A. & Mayo, S. L. Achieving stability and conformational specificity in designed proteins via binary patterning. *J Mol Biol* **305**, 619-31. (2001).

12. Dunbrack, R. L., Jr. & Karplus, M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J Mol Biol* **230**, 543-74. (1993).

13. Dahiyat, B. I. & Mayo, S. L. Probing the role of packing specificity in protein design. *PNAS* **94**, 10172-7. (1997).

14. Jorgensen, W. L. & Tirado-Rives, J. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *JACS* **110**, 1657-1666 (1988).

15. Dahiyat, B. I., Gordon, D. B. & Mayo, S. L. Automated design of the surface positions of protein helices. *Protein Science* **6**, 1333-7. (1997).

16. Gordon, D. B., Marshall, S. A. & Mayo, S. L. Energy functions for protein design. *Curr Opin Struct Biol* **9**, 509-13 (1999).

17. Pierce, N. A., Spriet, J. A., Desmet, J. & Mayo, S. L. Conformational splitting: A more powerful criterion for dead-end elimination. *J Comp Chem* **21**, 999-1009 (2000).

18. Johnson, B. H. & Hecht, M. H. Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. *Biotechnology* **12**, 1357-1360 (1994).

19. Santoro, M. M. & Bolen, D. W. Unfolding free-energy changes determined by the linear extrapolation method. 1.unfolding of phenylmethanesulfonyl a-chymotrpsin using different denaturants. *Biochemistry* **27** (1988).

20. Marshall, S. A. Ph.D Thesis, California Institute of Technology (2001).

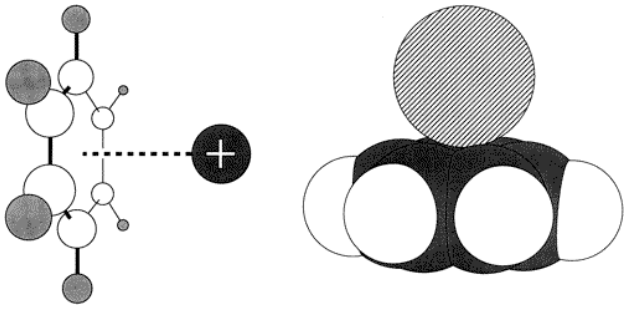


Figure 6-1 Schematic of the cation-π interaction.

Left: a generic cation is shown positioned along a benzene ring. Right: space-filling model of the $K^+\cdots$benzene complex; the optimal geometry has the cation interacting with the face of the aromatic ring, not the edge. Adapted from Ma & Dougherty, 1997.[4]

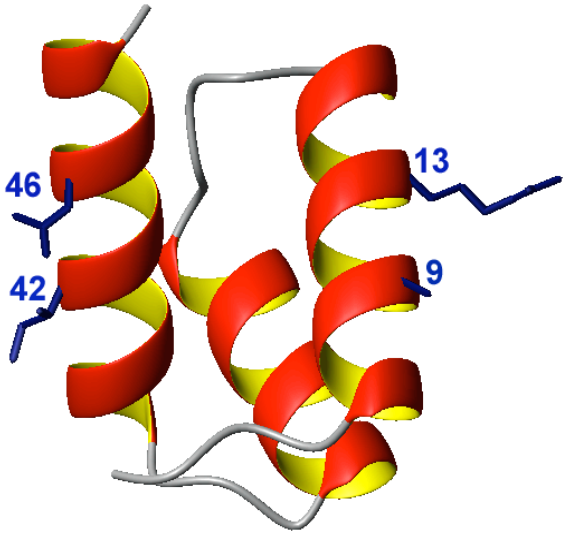


Figure 6-2 Ribbon diagram of engrailed homeodomain.

The tertiary structure of engrailed homeodomain, with positions 9, 13, 42, and 46 labeled. Side-chains shown are wild type.

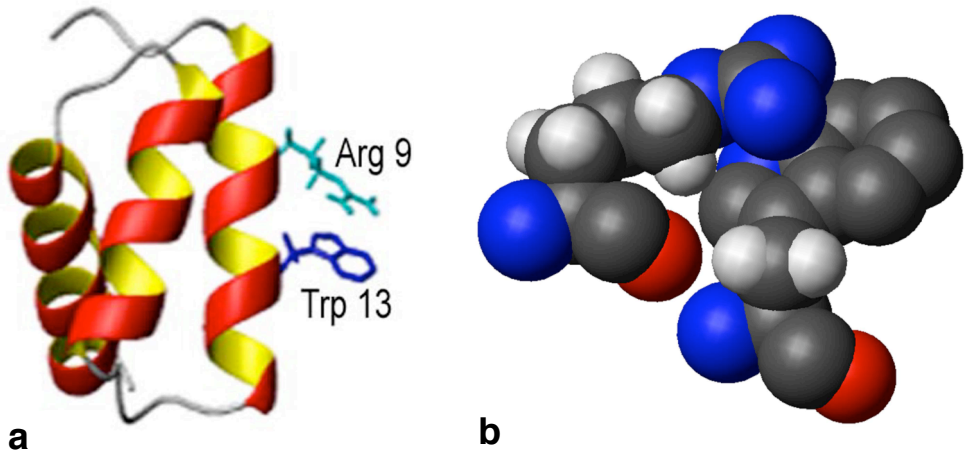


Figure 6-3 Modelled Arg9-Trp13 in engrailed homeodomain.

a. Modelled Arg9-Trp13 pair with planar stacking of the guanidinium group of Arg with the aromatic ring of Trp. **b**. The two groups are in close van der Waals contact, which should allow optimal cation-π contact.

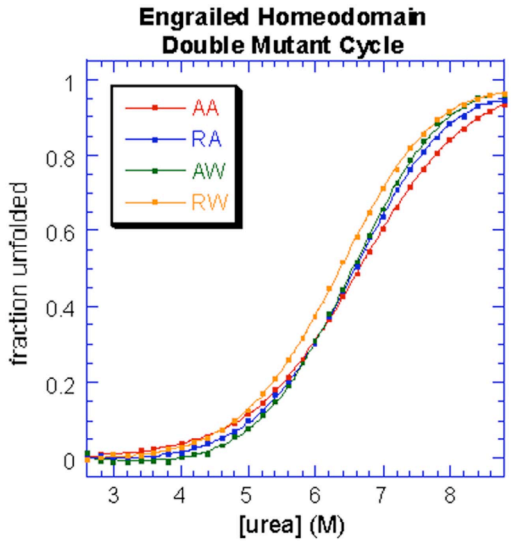


Figure 6-4 Urea denaturation of homeodomain variants.

Urea denaturation of homeodomain variants for double mutant cycle analysis. A9/A13 is shown in red, R9/A13 in blue, A9/W13 in green, and R9/W13 in orange.

Table 6-1: **Thermodynamic parameters of homeodomain variants from urea denaturation**.[20]

	ΔG_u[a] (kcal mol^{-1})	C_m[b] (M)	*M*[c] (kcal mol^{-1} M^{-1})
AA	4.82	6.6	0.73
AW	5.99	6.6	0.91
RA	5.58	6.6	0.85
RW	5.36	6.4	0.84

[a]Free energy of unfolding at 25 ºC

[b]Midpoint of the unfolding transition

[c]Slope of ΔG_u versus denaturant concentration

Chapter 7

Modulating nAChR Agonist Specificity by Computational Protein Design

The text of this chapter and work described were done in collaboration with Amanda L. Cashin.

Introduction

Ligand gated ion channels (LGIC) are transmembrane proteins involved in biological signaling pathways. These receptors are important in Alzheimer's, Schizophrenia, drug addiction and learning and memory.[1] Small molecule neurotransmitters bind to these transmembrane proteins, induce a conformational change in the receptor, and allow the protein to pass ions across the impermeable cell membrane. A number of studies have identified key interactions that lead to binding of small molecules at the agonist binding site of LGICs. High-resolution structural data on neuroreceptors are only just becoming available,[2-4] and functional data are still needed to further understand the binding and subsequent conformational changes that occur during channel gating.

Nicotinic acetylcholine receptors (nAChR) are one of the most extensively studied members of the Cys-loop family of LGICs, which include γ-aminobutyric, glycine and serotonin receptors. The embryonic mouse muscle nAChR is a transmembrane protein composed of five subunits, $(\alpha_1)_2\beta\gamma\delta$.[5] Biochemical studies [6,7] and the crystal structure of the acetylcholine binding protein (AChBP),[2] a soluble protein highly homologous to the ligand binding domain of the nAChR (Figure 7-1), identified two agonist binding sites at the α/γ and α/δ interfaces on the muscle type nAChR that are defined by an aromatic box of conserved amino acid residues. The principal face of the agonist binding site contains four of the five conserved aromatic box residues, while the complementary face contains the remaining aromatic residue.

Structurally similar nAChR agonists acetylcholine (ACh), nicotine (Nic) and epibatidine (Figure 7-2) bind to the same aromatic binding site with differing activity. Recently, Sixma and co-workers published a nicotine bound crystal structure of AChBP[3] which reveals additional agonist binding determinants. To verify the functional importance of potential agonist-receptor interactions revealed by the AChBP structures, chemical scale investigations were performed to identify mechanistically significant drug-receptor interactions at the muscle-type nAChR.[8,9] These studies identified subtle differences in the binding determinants that differentiate ACh, Nic and epibatidine activity.

Interestingly, these three agonists also display different relative activity among different nAChR subtypes. For example, the neuronal α7 nAChR subtype displays the following order of agonist potency: epibatidine > nicotine >ACh.[10] For the mouse muscle subtype the following order of agonist potency is observed: epibatidine > ACh >> nicotine.[8,11] A better understanding of residue positions that play a role in agonist specificity would provide insight into the conformational changes that are induced upon agonist binding. This information could also aid in designing nAChR subtype specific drugs.

The present study probes the residue positions that affect nAChR agonist specificity for acetylcholine, nicotine, and epibatidine. To accomplish this goal, we utilized AChBP as a model system for computational protein design studies to improve the poor specificity of nicotine at the muscle type nAChR.

Computational protein design is a powerful tool for the modification of protein-protein,[12] protein-peptide,[13] protein-ligand[14] interactions. For example, a designed calmodulin with 13 mutations from the wild-type protein showed a 155-fold increase in binding specificity for a peptide.[13] In addition, Looger et al. engineered proteins from the periplasmic binding protein superfamily to bind trinitrotoluene at nanomolar affinity, and lactate and serotonin at micromolar affinity.[14] These studies demonstrate the ability of computational protein design to successfully predict mutations that dramatically affect binding specificity of proteins.

With the availability of the 2.2 Å crystal structure of AChBP-nicotine complex,[3] the present study predicted mutations in efforts to stabilize AChBP in the nicotine preferred conformation by computational protein design. AChBP, although not a functional full-length ion-channel, provides a highly homologous model system to the extracellular ligand binding domain of nAChRs. The present study utilizes mouse muscle nAChR as the functional receptor to experimentally test the computational predictions. By stabilizing AChBP in the nicotine-bound conformation, we aim to modulate the binding specificity of the highly homologous muscle type nAChR for three agonists nicotine, acetylcholine and epibatidine.

Materials and Methods

Computational Protein Design with ORBIT

The AChBP-nicotine structure (PDB ID 1UWA) was obtained from the Protein Data Bank.[3] The subunits forming the binding site at the interface of B and C were selected for our design while the remaining three subunits (A, D, E) and the water molecules were deleted. Hydrogens were added with the Reduce program of MolProbity (http://kinemage.biochem.duke.edu/molprobity) and minimized briefly with ORBIT. The ORBIT protein design suite uses a physically based force-field and combinatorial optimization algorithms to determine the optimal amino acid sequence for a protein structure.[15,16] A backbone dependent rotamer library with χ1 and χ2 angles expanded by ±15° around all residues except Arg and Lys was used.[17] Charges for nicotine were calculated *ab initio* with Jaguar (Shrodinger) using density field theory with the exchange-correlation hybrid B3LYP and 6-31G** basis set. Nine residues (chain B: 89, 143, 144, 185, 192. chain C: 104, 112, 114, 53) interacting directly with nicotine are considered the primary shell, and were allowed to be all amino acids except Gly. Residues contacting the primary shell residues are considered the secondary shell (chain B: 87, 139, 141, 142, 146, 149, 182, 183, 184. chain C: 33, 34, 36, 51, 55, 57, 75, 98, 99, 102, 106, 110, 113, 116). Wild-type prolines and glycines were not designed. 87B, 33C and 113C were allowd to be all nonpolar amino acids except methionine, and 144B, 146B, 182B, 34C, 57C, 75C and 116C were allowed to be all polar residues. A tertiary shell includes residues within 4 Å of primary and secondary shell residues, and they were allowed to change in amino acid conformation but not identity. A bias towards the wild-type sequence using the

SBIAS module was applied at 1, 2, and 4 kcal*mol^{-1}. An algorithm based on the dead end elimination theorem (DEE) was used to obtain the global minimum energy amino acid sequence and conformation (GMEC).[18]

Mutagenesis and Channel Expression

In vitro runoff transcription using the AMbion mMagic mMessage kit was used to prepare mRNA. Site-directed mutagenesis was performed using Quick-Change mutagenesis and was verified by sequencing. For nAChR expression, a total of 4.0 ng of mRNA was injected in the subunit ration of 2:1:1:1 α:β:γ:δ. The β subunit contained a L9'S mutation, as discussed below. Mouse muscle embryonic nAChR in the pAMV vector was used, as reported previously.

Electrophysiology

Stage VI oocytes of *Xenopus laevis* were harvested according to approved procedures. Oocyte recordings were made 24 to 48 h post-injection in two-electrode voltage clamp mode using the OpusXpress^{TM} 600A (Molecular Devices Corporation, Union City, California).[8,19] Oocytes were superfused with calcium-free ND96 solution at flow rates of 1ml/min, 4 ml/min during drug application and 3 ml/min wash. Cells were voltage clamped at –60 mV. Data were sampled at 125 Hz and filtered at 50 Hz. Drug applications were 15 s in duration. Agonists were purchased from Sigma/Aldrich/RBI: 9([-]-nicotine tartrate), (acetylcholine chloride) and ([±] epibatidine). Epibatidine was also purchased from Tocris ([±]

epibatidine). All drugs were prepared in calcium-free ND96. Dose-response data were obtained for a minimum of 10 concentrations of agonists and for a minimum of 4 different cells. Curves were fitted to the Hill equation to determine EC_{50} and Hill coefficient.

Results and Discussion

Computational Design

The design of AChBP in the nicotine bound state predicted 10 mutations. To identify those predicted mutations that contribute the most to the stabilization of the structure, we used the SBIAS module of ORBIT, which applies a bias energy toward wild-type residues. We identified two predicted mutations, T57R and S116Q (AChBP numbering will be used unless otherwise stated) in the secondary shell of residues, with strong interaction energies. They are on the complementary subunit of the binding pocket (chain C) and formed inter-subunit side chain to backbone hydrogen bonds to the primary shell residues (Figure 7-3). S116Q reaches across the interface to form a hydrogen bond with a donor to acceptor distance of 3.0 Å with the backbone oxygen of Y89, one of the aromatic box residues important in forming the binding pocket. T57R makes a network of hydrogen bonds. E110 flips from the crystallographic conformation to form a hydrogen bond with a donor to acceptor distance of 3.0 Å with T57R, which also hydrogen bonds with E157 in its crystallographic conformation. T57R could also form a potential hydrogen bond, with a donor to acceptor distance of 3.6 Å, to the

backbone oxygen of C187, part of a disulfide cysteine bond on a principal loop in the binding domain. Most of the nine primary shell residues kept the crystallographic conformations, a testament to the high affinity of AChBP for nicotine (K_d=45nM).[3]

Interestingly, T57 is naturally R in AChBP from *Aplysia californica*, a different species of snail. It is not a conserved residue. From the sequence alignment (Figure 7-1) residue 57 is Q, E, Q, A in the alpha, beta, gamma, and delta subunits respectively. In addition, the S116Q mutation is at a highly conserved position in nAChRs. In all four mouse muscle nAChR subunits, residue 116 is a proline, part of a PP sequence. The mutation study will give us important insight into the necessity of the PP sequence for the function of nAChRs.

Mutagenesis

Conventional mutagenesis for T57R was performed at the equivalent position of AChBP's complementary face on the mouse muscle nAChR at γQ59R and δA61R subunits. The mutant receptor was evaluated using electrophysiology. When studying weak agonists and/or receptors with diminished binding capability, it is necessary to introduce a Leu-to-Ser mutation at a site known as 9' in the second transmembrane region of the β subunit.[8,9] This 9' site in the β subunit is almost 50 Å from the binding site, and previous work has shown that a L9'S mutation lowers the effective concentration at half

maximal response (EC_{50}) by a factor of roughly 10.[9,20] Results from earlier studies[9,20] and data reported below demonstrate that trends in EC_{50} values are not perturbed by L9'S mutations. In addition, the alpha subunits contain an HA epitope between M3 and M4. Control experiments show a negligible effect of this epitope on EC_{50}. Measurements of EC_{50} represent a functional assay; all mutant receptors reported here are fully functioning ligand-gated ion channels. It should be noted that the EC_{50} value is not a binding constant, but a composite of equilibria for both binding and gating.

Nicotine Specificity Enhanced by 59R Mutation

The ability of the γ59Rδ61R mutant to impact nicotine specificity at the muscle type nAChR was tested by determining the EC_{50} in the presence of acetylcholine, nicotine and epibatidine (Figure 7-4). The EC_{50} values for the wild-type and mutant receptors are show in Table 7-1. The computational design studies predict this mutation will help stabilize the nicotine bound conformation by enabling a network of hydrogen bonds with side chains of E110 and E157 as well as the backbone carbonyl oxygen of C187.

Upon mutation, the EC_{50} of nicotine *decreases* 1.8-fold compared to the wild-type value, thus improving the potency of nicotine for the muscle-type nAChR. Conversely, ACh shows 3.9-fold *increase* in EC_{50} compared to the wild-type value, thus decreasing the potency of ACh for the nAChR. The values for epibatidine are relatively unchanged in the presence of the mutation in

comparison to wild-type. Interestingly, these data show a change in agonist specificity of ACh and epibatidine in comparison to nicotine for the nAChR. The wild-type receptor prefers ACh 69-fold more than nicotine and epibatidine 95-fold more than nicotine. The agonist specificity is significantly changed with the γ59Rδ61R mutant where the receptor's preference for ACh decreases to 10-fold over nicotine and epibatidine decreases to 44-fold over nicotine. The specificity change can be quantified in the ΔΔG values from Table 7-1. These values indicate a more favorable interaction for nicotine (-0.3 kcal/mol) than for ACh (0.8 kcal/mol) and epibatidine (0.1 kcal/mol) in the presence of the γ59Rδ61R mutant compared to wild-type receptors.

The ability of this single mutation to enhance nicotine specificity of the mouse nAChR demonstrates the importance of the secondary shell residues surrounding the agonist binding site in determining agonist specificity. Because the aromatic box is nearly 100% conserved among nAChRs, we hypothesize the agonist specificity does not depend on the amino acid composition of the binding site itself, but on specific conformations of the aromatic residues. It is possible that the secondary shell residues, significantly less conserved among nAChR sub-types, play a role in stabilizing unique agonist preferred conformations of the binding site. The T57R mutation, a secondary shell residue on the complementary face of the binding domain, was designed to interact with the primary face shell residue C187 across the subunit interface to stabilize the

nicotine preferred conformation. These data demonstrate the importance of this secondary shell residue in determining agonist activity and selectivity.

Because the nicotine bound conformation was used as the basis for the computational design calculations, the design generated mutations that would further stabilize the nicotine bound state. The 57R mutation electrophysiology data demonstrate an increase in preference in nicotine for the receptor compared to wild-type receptors. The activity of ACh, structurally different from nicotine, decreases possibly because it undergoes an energetic penalty to reorganize the binding site into an ACh preferred conformation or to bind to a nicotine preferred confirmation. The changes in ACh and nicotine preference for the designed binding pocket conformation leads to a 6.9-fold *increase* in specificity for nicotine in the presence of 57R. The activity of epibatidine, structurally similar to nicotine, remains relatively unchanged in the presence of the 57R mutation. Perhaps the binding site conformation of epibatidine more closely resembles that of nicotine and therefore does not undergo a significant change in activity in the presence of the mutation. Therefore only a 2.2-fold increase in agonist specificity is observed for nicotine over epibatidine.

Conclusions and Future Directions

The present study aimed to utilize computational protein design to modulate the agonist specificity of nAChR for nicotine, acetylcholine and epibatidine. By stabilizing nAChR in the nicotine-bound conformation, we

predicted two mutations to stabilize the nAChR in the nicotine preferred conformation. The initial data has corroborated our design. The T57R mutation is responsible for a 6.9-fold *increase* in specificity of nicotine over acetylcholine, and 2.2-fold *increase* for nicotine over epibatidine. The S116Q mutations experiments are currently underway. Future directions could include probing agonist specificity of these mutations at different nAChR subtypes and other Cys-loop family members. As future crystallographic data become available this method could be extended to investigate other ligand-bound LGIC binding sites.

References

1. Paterson, D. & Nordberg, A. Neuronal nicotinic receptors in the human brain. *Prog Neurobiol* **61**, 75-111 (2000).
2. Brejc, K. et al. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269-76 (2001).
3. Celie, P. H. N. et al. Nicotine and Carbamylcholine Binding to Nicotinic Acetylcholine Receptors as Studied in AChBP Crystal Structures. *Neuron* **41**, 907-914 (2004).
4. Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J Mol Biol* **346**, 967-89 (2005).
5. Miyazawa, A., Fujiyoshi, Y., Stowell, M. & Unwin, N. Nicotinic acetylcholine receptor at 4.6 Å resolution: transverse tunnels in the channel wall. *J Mol Biol* **288**, 765-86 (1999).
6. Grutter, T. & Changeux, J. P. Nicotinic receptors in wonderland. *Trends in Biochemical Sciences* **26**, 459-463 (2001).
7. Karlin, A. Emerging structure of the nicotinic acetylcholine receptors. *Nat Rev Neurosci* **3**, 102-14 (2002).
8. Cashin, A. L., Petersson, E. J., Lester, H. A. & Dougherty, D. A. Using physical chemistry to differentiate nicotinic from cholinergic agonists at the nicotinic acetylcholine receptor. *Journal of the American Chemical Society* **127**, 350-356 (2005).

9. Beene, D. L. et al. Cation-pi interactions in ligand recognition by serotonergic (5-HT3A) and nicotinic acetylcholine receptors: the anomalous binding properties of nicotine. *Biochemistry* **41**, 10262-9 (2002).

10. Gerzanich, V. et al. Comparative pharmacology of epibatidine: a potent agonist for neuronal nicotinic acetylcholine receptors. *Mol Pharmacol* **48**, 774-82 (1995).

11. Rush, R., Kuryatov, A., Nelson, M. E. & Lindstrom, J. First and second transmembrane segments of alpha3, alpha4, beta2, and beta4 nicotinic acetylcholine receptor subunits influence the efficacy and potency of nicotine. *Mol Pharmacol* **61**, 1416-22 (2002).

12. Kortemme, T. et al. Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* **11**, 371-9 (2004).

13. Shifman, J. M. & Mayo, S. L. Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc Natl Acad Sci U S A* **100**, 13274-9 (2003).

14. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185-90 (2003).

15. Dahiyat, B. I. & Mayo, S. L. *De novo* protein design: fully automated sequence selection. *Science* **278**, 82-7 (1997).

16. Mayo, S. L., Olafson, B. D. & Goddard, W. A. Dreiding: a Generic Force-Field for Molecular Simulations. *Journal of Physical Chemistry* **94**, 8897-8909 (1990).

17. Dunbrack, R. L., Jr. & Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* **6**, 1661-81 (1997).

18. Pierce, N. A., Spriet, J. A., Desmet, J. & Mayo, S. L. Conformational splitting: A more powerful criterion for dead-end elimination. *Journal of Computational Chemistry* **21**, 999-1009 (2000).

19. Lummis, S. C., D, L. B., Harrison, N. J., Lester, H. A. & Dougherty, D. A. A cation-pi binding interaction with a tyrosine in the binding site of the GABAC receptor. *Chem Biol* **12**, 993-7 (2005).

20. Kearney, P. C. et al. Agonist binding site of the nicotinic acetylcholine receptor: Tests with novel side chains and with several agonists. *Molecular Pharmacology* **50**, 1401-1412 (1996).

```
AChBP-L         LDRADILYN-IRQTSR----PDVIPTQRDR-PVAVSVSLKFINILEVNEITNEVDVVFWQ
AChBP-A         --QANLMRLKSDLFNR----SPMYPGPTKDDPLTVTLGFTLQDIVKVDSSTNEVDLVYYE
alpha-m         LGSEHETRLVAKLFED--YSSVVRPVEDHREIVQVTVGLQLIQLINVDEVNQIVTTNVRL
beta-m          RGSEAEGQLIKKLFSN--YDSSVRPAREVGDRVGVSIGLTLAQLISLNEKDEEMSTKVYL
gamma-m         QSRNQEERLLADLMRN--YDPHLRPAERDSDVVNVSLKLTLTNLISLNEREEALTTNVWI
delta-m         WGLNEEQRLIQHLFNEKGYDKDLRPVARKEDKVDVALSLTLSNLISLKEVEETLTTNVWI

AChBP-L         QTTWSDRTLAWNSSHSP--DQVSVPISSLWVPDLAAYNAISKPEVLTPQLARVVS-DGEV
AChBP-A         QQRWKLNSLMWDPNEYGNITDFRTSAADIWTPDITAYSSTRPVQVLSPQIAVVTH-DGSV
alpha-m         KQQWVDYNLKWNPDDYGGVKKIHIPSEKIWRPDVVLYNNADGDFAIVKFTKVLLDYTGHI
beta-m          DLEWTDYRLSWDPAEHDGIDSLRITAESVWLPDVVLLNNNDGNFDVALDINVVVSFEGSV
gamma-m         EMQWCDYRLRWDPKDYEGLWILRVPSTMVWRPDIVLENNVDGVFEVALYCNVLVSPDGCI
delta-m         DHAWVDSRLQWDANDFGNITVLRLPPDMVWLPEIVLENNNDGSFQISYACNVLVYDSGYV

AChBP-L         LYMPSIRQRFSCDVSGVDTESG-ATCRIKIGSWTHHSREISVDPTTEN-----------S
AChBP-A         MFIPAQRLSFMCDPTGVDSEEG-VTCAVKFGSWVYSGFEIDLKTDTDQ-----------V
alpha-m         TWTPPAIFKSYCEIIVTHFPFDEQNCSMKLGTWTYDGSVVAINPESDQ---------P--D
beta-m          RWQPPGLYRSSCSIQVTYFPFDWQNCTMVFSSYSYDSSEVSLKTGLDPE---GEERQEVY
gamma-m         YWLPPAIFRSSCSISVTYFPFDWQNCSLIFQSQTYSTSEINLQLSQED----GQAIEWIF
delta-m         TWLPPAIFRSSCPISVTYFPFDWQNCSLKFSSLKYTAKEITLSLKQEEENNRSYPIEWII

AChBP-L         DDSEYFSQYSRFEILDVTQKKNSVTYSC--C-PEAYEDVEVSLNFRKKGRSEIL------
AChBP-A         DLSSYYAS-SKYEILSATQTRQVQHYSC--C-PEPYIDVNLVVKFRERRAGNGFFRNLFD
alpha-m         LSN--FMESGEWVIKEARGWKHWVFYSC--CPTTPYLDITYHFVMQRLPLYFIVNVIIPC
beta-m          IHEGTFIENGQWEIIHKPSRLIQLPGDQRGGKEGHHEEVIFYLIIRRKPLFYLVNVIAPC
gamma-m         IDPEAFTENGEWAIRHRPAKMLLDSVAP--AEEAGHQKVVFYLLIQRKPLFYVINIIAPC
delta-m         IDPEGFTENGEWEIVHRAAKLNVDPSVP--MDSTNHQDVTFYLIIRRKPLFYIINILVPC
```

Figure 7-1. Sequence alignment of AChBP with nAChR subunits from mouse muscle. AChBP-L (AChBP *Lymnaea*) and AChBP-A (AChBP *Aplysia*) are soluble proteins that bind acetylcholine. The predicted mutations are from design calculations on AChBP-L and nicotine complex. The binding pockets on nAChR on mouse muscle are formed between the principle subunit, alpha, and complementary subunits, beta, gamma and delta. The highly conserved aromatic box residues are highlighted in magenta, and the residue positions of the predicted mutations are in cyan.

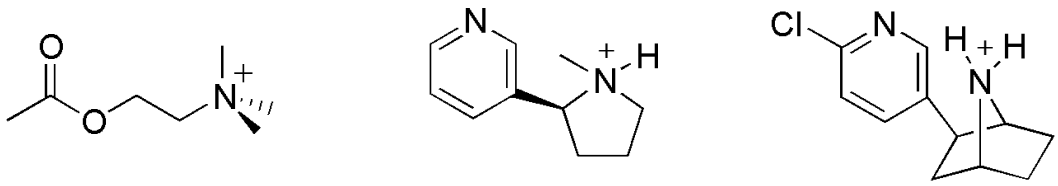


Figure 7-2. Structures of nAChR agonists, acetylcholine, nicotine, and epibatidine. Epibatidine is a nicotine-like agonist.

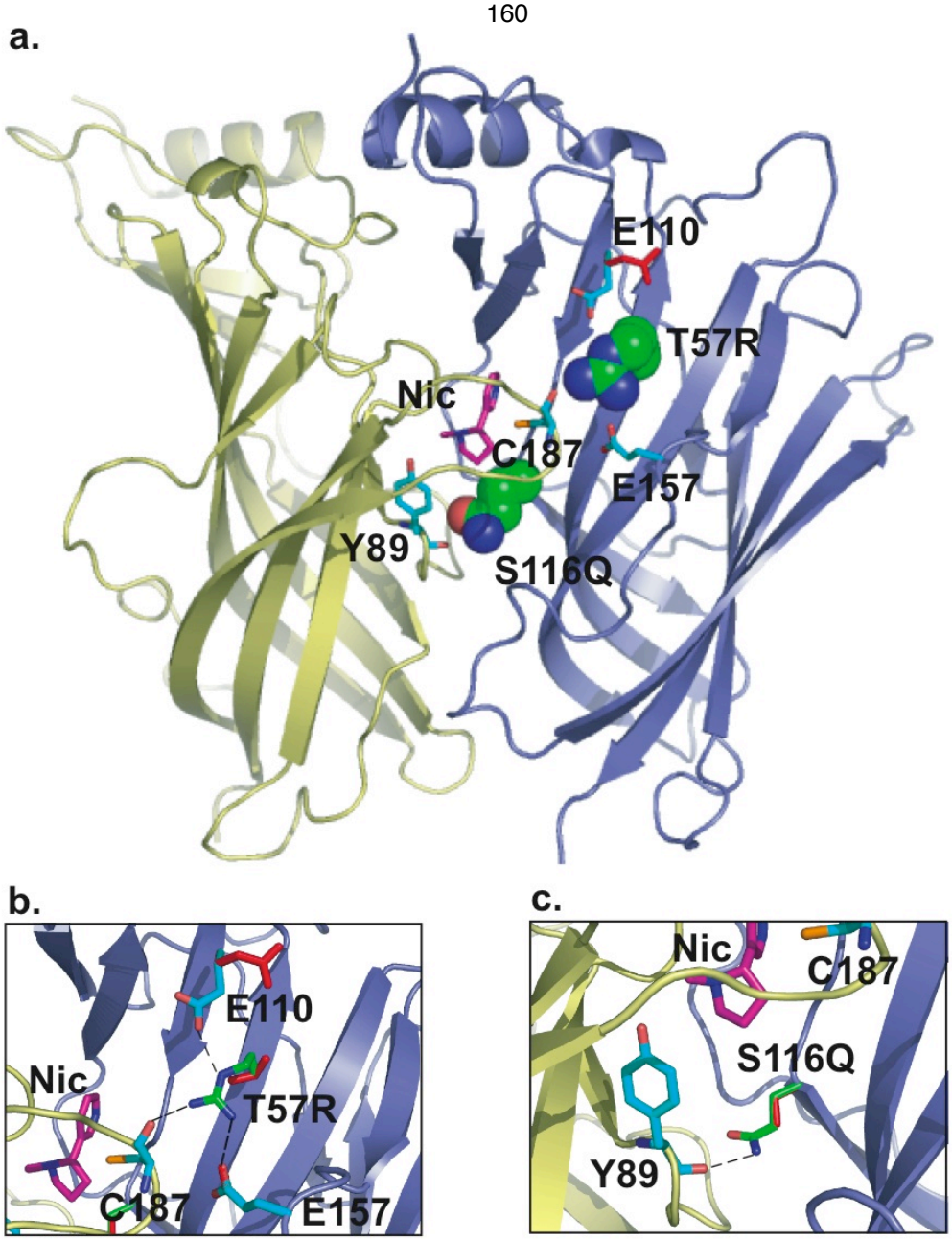


Figure 7-3. Predicted mutations from computational design of AChBP. **a.** Ribbon diagram of two AChBP subunits. Yellow: principle subunit. Blue: complementary subunit. Nicotine, the predicted mutations, and interacting sidechains are shown in CPK-inspired colors. Nicotine: magenta. Predicted mutations: green in space-filling model. Interacting residues: cyan. Crystallographic conformations are shown in red. **b**. Close-up view of T57R interactions. **c**. Close-up view of S116Q . Hydrogen bonds are shown as black dashed lines.

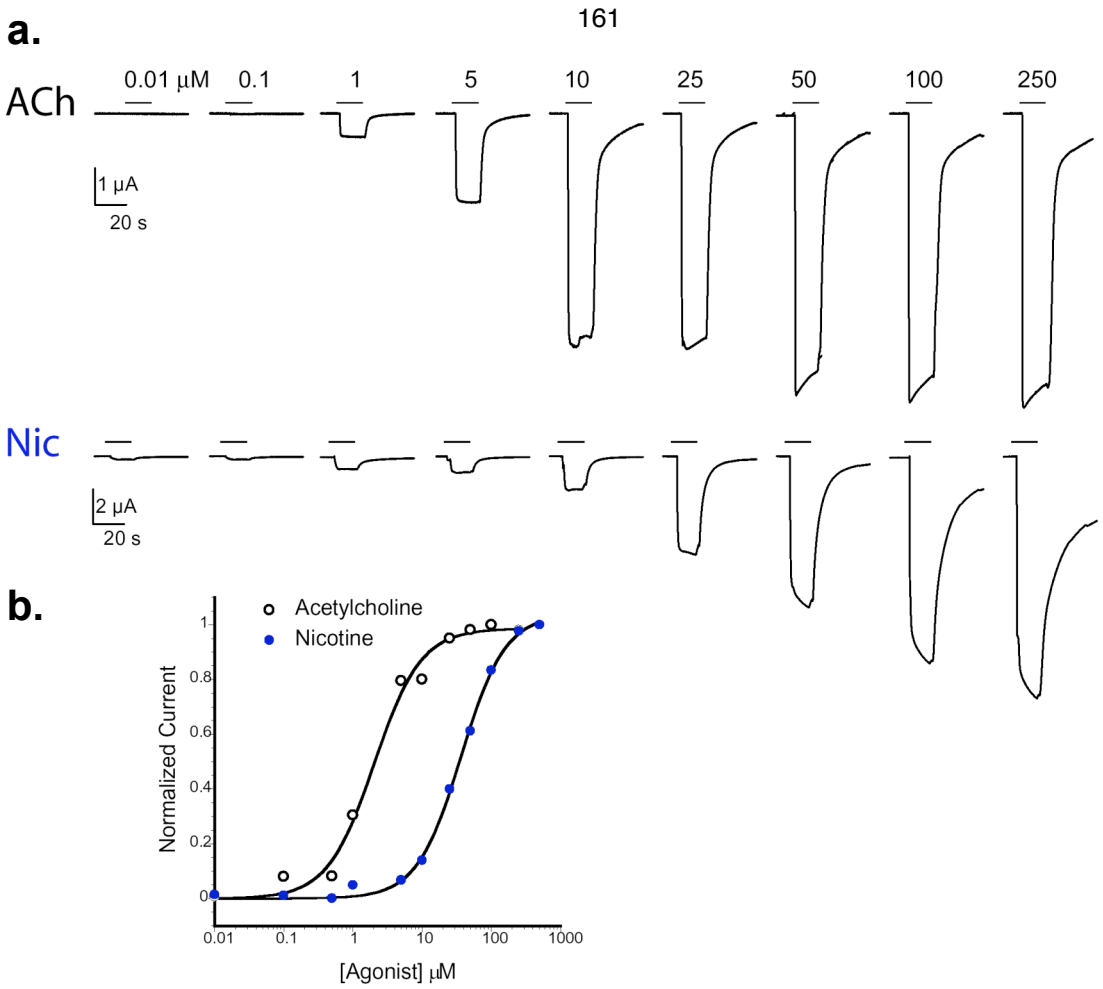


Figure 7-4. Electrophysiology data. Electropysiological analysis of ACh and nicotine. **a**. Representative voltage clamp current traces for oocytes expressing mutant muscle nAChRs $(\alpha 1)\beta_9\gamma 59R\delta 61R$. Bars represent application of ACh and nicotine at the concentrations noted. **b**. Representative ACh (○)and nicotine (●) dose-response relations and fits to the Hill equation for oocytes expressing $(\alpha 1)\beta_9\gamma 59R\delta 61R$ nAChRs.

Table 7-1. Mutation enhancing nicotine specificity

Agonist	Wild-type EC_{50}[a]	γ59Rδ61R EC_{50}[a]	Wild-type Nic/Agonist	γ59Rδ61R Nic/Agonist	γ59Rδ61R ΔΔG[b]
ACh	0.83 ± 0.04	3.2 ± 0.4	69	10	0.8
Nicotine	57 ± 2	32 ± 3	1	1	-0.3
Epibatidine	0.60 ± 0.04	0.72 ± 0.05	95	44	0.1

[a]EC_{50} (μM) ± standard error of the mean. (-) Nicotine nicotine and racemic epibatidine were used in these experiments. The receptor has a Leu9'Ser mutation in M2 of the β subunit. [b]ΔΔG (kcal/mol).