METHODS FOR COMPUTATIONAL ENZYME DESIGN AND APPLICATION TO THE CHORISMATE-PREPHENATE REARRANGEMENT

Thesis by

Jonathan Kyle Lassila

In Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2006

Defended November 15, 2006

ACKNOWLEDGEMENTS

I thank my advisor, Stephen Mayo, for providing unwavering support and encouragement throughout my somewhat wavering research pathway.

I thank my thesis committee for also being uniformly supportive and encouraging. I looked forward to my TA assignment every year thanks to the enthusiasm of Pamela and Doug. Thank you to Frances for some very helpful and crucial advice.

I thank all Mayo lab members, past and present, for taking on difficult problems and making this thesis possible in many ways. In particular, I am grateful to (in order of appearance) Deepshikha, Pavel, Dan, Geoff, Jennifer, Peter, Heidi, Ben, Christina, and Corey for working on projects with me. For this I also thank Michelle from the Arnold lab. Thanks to Cynthia and Rhonda for all of their help along the way and to Shannon, Shira, Jessica, Possu, Oscar, Jessica, Eun Jung, and Eric for important contributions.

I am grateful for the friends and community that I gained in Los Angeles. In particular, for a variety of contributions to my time here (in order of appearance): Dave, Ren, Oriane, Max, Daniel, Masen, Ed, Christian, Tina, Zane, Jody B., Nina, Christine, and Jody M.

Thank you to my family, who instilled in me the idea that science is a good thing to do.

ABSTRACT

The Claisen rearrangement of chorismate to prephenate has become an important model system for developing understanding of enzymatic catalysis as well as for computational treatment of enzyme active sites. This thesis presents general methods for the computational design of enzyme active sites and applies these methods to the design of catalysts for the chorismate-prephenate rearrangement. The computational methods described allow the incorporation of transition-state structures and other small molecules into protein design calculations. These design procedures were tested through redesign of the active site of *Escherichia coli* chorismate mutase. The six predicted mutations were experimentally characterized and most maintained or increased the catalytic activity of the enzyme. To further investigate the context of the mutations predicted in the calculation and the tolerance of a natural enzyme to secondary active site mutations, extensive substitution experiments were performed. The effect of every amino acid in five active site hydrophobic positions and one N-capping position was evaluated. These experiments clarified some of the strengths and weaknesses of the computational modeling procedure. Finally, attempts to design a completely new enzyme for catalysis of the chorismate-prephenate rearrangement are discussed.

CONTENTS

FIGURES, TABLES, AND SCHEMES

Figure, Table, or Scheme *Page*

Chapter 1

Introduction

Enzymes are excellent catalysts, accelerating the rates of chemical reactions by as much as 10^{19} fold.[1] A great deal is now known about the structures of enzymes[2] and the chemical mechanisms of the reactions they catalyze.[3] It has been appreciated for more than 80 years that catalysts can accelerate the rates of chemical reactions by facilitating the formation of their transition states.[4-6] However, the questions that remain unanswered are likely to motivate enzymologists and protein design groups for some time:

How do enzymes facilitate formation of the transition state of a reaction?

How can these physical principles be employed to design new enzymes?

While these two questions form central themes in this thesis, the dominant focus of the work presented here may be the degree to which the questions remain largely unanswered.

The primary objective of the thesis research was to use computational protein design methods to develop a new catalyst for the Claisen rearrangement of chorismate to prephenate. The rearrangement offered a compelling test case for the development and testing of computational enzyme design approaches. Its status as a rare "pure" enzyme-catalyzed reaction, an intramolecular reaction that does not require general acids or bases, has no intermediate steps, and proceeds through the same mechanism in enzyme and in solution,[7-9] has made it a benchmark for development of hybrid quantum mechanical

methods for studying enzymes and the focus of renewed debate about the fundamental origins of catalysis.[10] Transition state structures for the rearrangement have been well characterized by computational and experimental methods,[7-11] and their compact configuration makes them especially suited for modeling in the protein design framework.

While the rational, computational design of completely new protein catalysts has been accomplished,[12,13] the application of protein design methods to enzyme catalysis is still in its earliest stages of development. Computational protein design methods optimize the identity of amino acid side chains with respect to an energy function typically containing terms for van der Waals, hydrogen bonding, electrostatics, and solvation energies.[14-17] Because these calculations represent very difficult optimization problems, discrete representations of side chain conformations must be used[18-20] and a number of specialized search algorithms have been developed.[21-23] The process has proven effective in creating thermostable proteins, ligand-binding proteins, and even new folding topologies.[24-26] In applying these methods to the design of enzymes, however, new approaches must be developed to reflect the special energetics of the catalytic process. This last element represents the most difficult challenge of computational enzyme design. Because we do not yet have a complete understanding—let alone a consistent quantitative description—of the molecular and atomic interactions that enable enzyme catalysis,[27,28] it might seem premature to attempt to rationally design new enzymes. However, I think that it is worthwhile to try and I also believe that the two questions highlighted above will be answered gradually and in concert as we refine and experimentally test theoretical models of enzymatic reactions.

While a computationally designed new enzyme was not realized in this thesis research, the following five chapters present progress toward the goal and lessons learned along the way.

Chapter 2 discusses methods for modeling small molecules during the process of an active site design calculation. The placement of small molecules such as enzymatic transition state structures represents an important obstacle in computational protein design methodology. If a small molecule placement within the binding site is not known, as in attempts to design new enzymes, the possible translational, rotational, and

conformational variations of the small molecule must be considered in combination with side chain conformations and identities. Because sequence design calculations already represent large and difficult optimization problems, the added combinatorial complexity introduced with the small molecule must be dealt with creatively to make the calculation tractable.

In chapter 3, an experiment is described in which the computational design methods described in chapter 2 were applied to the active site of *E. coli* chorismate mutase. The six mutations predicted in the calculation were tested experimentally. One mutation slightly increased the catalytic efficiency of the enzyme.

Chapters 4 and 5 investigate the context of the mutations described in chapter 3. A series of exhaustive substitutions were performed, where the effect of every amino acid in each of the six positions previously mutated was investigated. For each variant with chorismate mutase activity *in vivo*, thermostabilities and kinetic parameters were measured. These studies were motivated by two objectives. First, they provide information about the tolerance of a natural active site to the mutation of hydrophobic and polar residues. Second, they allow feedback with which to assess the energy functions used in protein design experiments. Chapter 4 investigates the five hydrophobic positions that were mutated in the design experiment of chapter 3. An active site N-capping residue is studied in chapter 5; this residue is particularly interesting because the mutation predicted by the computational procedure in this position lost measurable catalytic activity.

Chapter 6 presents an overview of lessons from this thesis research. The measured stabilities and activities of active site mutations from chapter 3 and 4 are compared to energies from computational design calculations. Some reflections on the goal of *de novo* design are offered.

The appendices describe work related to the design and energetics of protein structure in general. Appendix A addresses the question of how specific side chains on the surfaces of proteins contribute to protein stability. Specifically, the contribution of surface salt bridges to protein stability has been debated, in part because fixing a side chain in position for such an interaction could lead to a destabilizing effect when the loss of conformational entropy of that side chain is considered. Networks of ionic side chains

had been proposed to offer a greater stabilizing influence because each individual side chain might experience a greater electrostatic benefit for the same entropic cost. Appendix A describes how thermodynamic mutant cycles were used to look for this cooperative effect in one system. In Appendix B, the X-ray diffraction structure of a computationally designed protein is described. Surprisingly, despite suggestions that the protein adopted the correct fold in solution, the crystal structure showed a dramatically different fold. Although the origin of altered folding specificity is not known, evidence is presented that one of the crystallization reagents, dioxane, led to increased helicity and altered oligomerization state in solution.

References

1. Wolfenden, R., and Snider, M. J. (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res. 34,* 938-945.

2. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucl. Acids Res. 28,* 235-242.

3. Walsh, C. (1979) Enzymatic Reaction Mechanisms. Freeman, San Franscisco.

4. Polanyi, M. (1921) Uber adsorptionskatalyze. *Z. Elektrochem. 27,* 142-150.

5. Pauling, L. (1946) Molecular architecture and biological reactions. *Chem. Eng. News 24*, 1375-1377.

6. Wolfenden, R. (1976) Transition state analog inhibitors and enzyme catalysis. *Annu. Rev. Biophys. Bioeng. 5,* 271-306.

7. Addadi, L., Jaffe, E. K., and Knowles, J. R. (1983) Secondary tritium isotope effects as probes of the enzymic and nonenzymic conversion of chorismate to prephenate. *Biochemistry 22,* 4494-4501.

8. Gustin, D. J., Mattei, P., Kast, P., Wiest, O., Lee, L., Cleland, W. W., and Hilvert, D. (1999) Heavy atom isotope effects reveal a highly polarized transition state for chorismate mutase. *J. Am. Chem. Soc. 121*, 1756-1757.

9. Wright, S. K., DeClue, M. S., Mandal, A., Lee, L., Wiest, O., Cleland, W. W., and Hilvert, D. (2005) Isotope effects on the enzymatic and nonenzymatic reactions of chorismate. *J. Am. Chem. Soc. 127*, 12957-12964.

10. Gao, J., Ma, S., Major, D.T., Nam, K., Pu, J., and Truhlar, D.G. (2006) Mechanisms and free energies of enzymatic reactions. *Chem. Rev. 106*, 3188-3209.

11. Wiest, O., and Houk, K.N. (1994) On the transition state of the chorismate-prephenate rearrangement. *J. Org. Chem. 59*, 7582-7584.

12. Bolon, D. N. and Mayo, S. L. (2001) Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. USA 98*, 14274-14279.

13. Dwyer, M.A., Looger, L.L. and Hellinga, H.W. (2004) Computational design of a biologically active enzyme. *Science 304,* 1967-1971.

14. Dahiyat, B.I. and Mayo, S.L. (1997) *De novo* protein design: Fully automated sequence selection. *Science 278*, 82-87.

15. Dahiyat, B.I. and Mayo, S.L. (1997) Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. USA 94*, 10172-10177.

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

17. Street, A.G. and Mayo, S.L. (1998) Pairwise calculation of protein solvent-accessible surface areas. *Fold. Des. 3*, 253-258.

18. Ponder, J.W. and Richards, F.M. (1987) Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol. 193*, 775-791.

19. Dunbrack, R.L, Jr. and Cohen, F.E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Prot. Sci. 6*, 1661-1681.

20. Lovell, S.C., Word, J.M., Richardson, J.S. and Richardson, D.C. (2000) The penultimate rotamer library. *Proteins, 40*, 389-408.

21. Desmet, J., Demaeyer, M., Hazes, B., and Lasters, I. (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature 356*, 539-542.

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

23. Desmet, J., Spriet, J. and Lasters, I. (2002) Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Prot. Struct. Funct. Genet., 48*, 31-43.

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

25. Looger, L. L., Dwyer, M. A., Smith, J. J. and Hellinga, H. W. (2003) Computational design of receptor and sensor proteins with novel functions. *Nature 423*, 185-190.

26. Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L. and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science 302*, 1364-1368.

27. Kraut, D.A., Carroll, K.S., and Herchlag, D. (2003) Challenges in enzyme mechanism and energetics. *Annu. Rev. Biochem. 72*, 517-571.

28. Borman, S. (2004) Much ado about enzyme mechanisms. *Chem. Eng. News 82*, 35-39.

Chapter 2

Combinatorial Methods for Small Molecule Placement in Computational Enzyme Design

The text of this chapter was adapted from a manuscript coauthored with Heidi K. Privett, Benjamin D. Allen, and Stephen L. Mayo.

Lassila, J. K., Privett, H. K., Allen, B. D., and Mayo, S. L. (2006) Combinatorial methods for small molecule placement in computational enzyme design. *Proceedings of the National Academy of Sciences USA 103*, 16710-16715.

Abstract

The incorporation of small molecule transition state structures into protein design calculations poses special challenges because of the need to represent the added translational, rotational, and conformational freedoms within an already difficult optimization problem. Successful approaches to computational enzyme design have focused on catalytic side-chain contacts to guide placement of small molecules in active sites. We describe a process for modeling small molecules in enzyme design calculations that extends previously described methods, allowing favorable small molecule positions and conformations to be explored simultaneously with sequence optimization. Because all current computational enzyme design methods rely heavily on sampling of possible active site geometries from discrete conformational states, we tested the effects of discretization parameters on calculation results. Rotational and translational step sizes as well as side-chain library types were varied in a series of computational tests designed to identify nativelike binding contacts in three natural systems. We find that conformational parameters, especially the type of rotamer library used, significantly affect the ability of design calculations to recover native binding-site geometries. We describe the construction and use of a crystallographic conformer library, and find that it more reliably captures active-site geometries than traditional rotamer libraries in the systems tested.

Introduction

As catalysts, enzymes offer advantageous properties including dramatic rate enhancements, complete control over absolute stereochemistry, and nontoxic biodegradation. Yet a fundamental limiting factor in the use of enzymes for chemical synthesis, bioremediation, therapeutics, and other applications is the availability of enzymes with the required activities, specificities, and tolerances to reaction conditions. It is therefore a major goal of computational protein design to be able to reliably create completely new protein catalysts with specific properties on demand.

A catalyst by definition must reduce the energy barrier for formation of the transition state. To design transition-state-stabilizing interactions, computational protein design groups have incorporated transition-state or high-energy intermediate state structures into design calculations. These efforts have yielded experimentally verified new catalytic proteins.[1,2] However, substantial challenges still prevent routine or reliable design of enzymes. One major challenge is in finding energy functions that are fast enough for large calculations but that still provide informative approximations of electrostatic and desolvation effects in the protein environment.[3,4] This paper focuses on another fundamental challenge, the need to represent the large translational, rotational, and conformational freedoms of a small molecule within already astronomically large sequence design calculations.

Here we define protein design as the selection of amino acid sequences such that the resulting protein occupies a given three-dimensional fold and has desired functional properties. Earlier experiments sought to redesign full protein sequences or confer increased thermostability,[5,6] but newer work has successfully introduced other properties including catalytic activity, conformational specificity, ligand affinity, and even novel protein folds.[1,2,7-9] In these examples, side-chain placement algorithms were used to select from a set of discrete, probable side-chain rotamers using energy functions tuned to produce thermostable proteins. These calculations represent difficult optimization problems[10] and they can also be large—a sample calculation performed on a typical enzyme active site yields more than 10^{65} possible sequence combinations, even when excluding movements of the small molecule.

The computational demands of sequence selection prevent ligand positioning using standard docking procedures, which often approximate or neglect side-chain flexibility.[11] Approaches developed specifically for the purpose of enzyme and binding site design have introduced other schemes to limit the calculation size. Looger *et al.* used stationary, inflexible ligand poses in a large number of individual protein design calculations and demonstrated experimentally that several of the resulting proteins had high ligand affinity.[8] Lilien *et al.* reported and experimentally validated an ensemble-based method that allows ligand translation and rotation simultaneously with side-chain optimization but only permits mutation of two or three amino acid positions at a time.[12] Chakrabarti *et al.* described a method for sequence design that neglects conformational and positional ligand flexibility and has not been experimentally tested.[13,14]

To design new enzyme active sites, a ligand placement method must be able to select side chains in many positions and must consider rotational, translational, and conformational freedom of the small molecule. The new catalytic proteins of Bolon and Mayo[1] and Dwyer *et al*[2] were designed by treating high-energy-state structures of the reacting molecules as extensions of contacting amino acid side-chain rotamers. In the latter case, a two-step procedure was utilized, where ligands, anchoring side chains, and other catalytic side chains were placed through a geometric screening procedure and surrounding side chains were designed in a second step.[2,15] We have developed a process for ligand placement in computational protein design calculations that expands upon previous work and that allows ligand rotation, translation, and conformational freedom to be explored combinatorially within the sequence design calculation itself. The implementation of ligand placement procedures within the context of the pairwise-decomposable protein design framework makes it possible to use a single energy function that can be parameterized as needed to reproduce experimental data.

We tested both a simple rotational and translational process for ligand placement as well as the previously used targeted ligand placement approach. A contact-based screening method is described that allows selection of ligand positions and confomations compatible with catalytic contacts. Test calculations in three systems, *E. coli* chorismate mutase, *S. cerevisiae* triosephosphate isomerase, and *S. avidinii* streptavidin, suggest that the success of ligand placement procedures can be quite sensitive to conformational

sampling parameters including rotational and translational step sizes and the types of rotamer libraries used. We evaluated the efficacy of two standard rotamer libraries and two crystallographic conformer libraries. Traditional rotamers are constructed from canonical χ angles determined by statistical analysis of the Protein Data Bank,[16-18] whereas conformers have Cartesian coordinates taken directly from high-resolution structures. Conformer libraries may allow more accurate modeling because they are not limited to ideal geometries and their sizes can be tuned more easily and naturally.[19,20] In our tests, a backbone-independent conformer library recovered wild-type-like active site geometries more successfully than the other libraries, despite smaller size.

Results and Discussion

We have implemented and tested a process for incorporation of small molecules into computational protein design calculations. The procedure is general and may be used to place ground-state ligands or transition-state structures. It is also amenable to multistate design methods that seek to explicitly reflect the energy difference between reactant and transition states or between alternative ligands.

General Calculation Procedure

Each ligand placement calculation comprised five steps. In the first step, a large number of discrete variations of ligand coordinates was created. Initial sets of orientations were created by one of two methods, either simple rotation and translation or a targeted placement approach, both of which are discussed in more detail in subsequent sections. In the tests described here, each set of ligand variations contained 10^6-10^9 members, reflecting rotational and translational movement as well as internal conformational flexibility.

Next, the large number of substrate orientations was reduced to a manageable number (< ~20,000) using both a simple hard-sphere steric potential to check for backbone clashes and a set of user-defined geometric criteria for side-chain/ligand contacts. In this work, geometric criteria were defined to reflect the distances, angles, and torsions characteristic of important catalytic contacts observed in the crystal structures (Figure 1). In designing an enzyme with no naturally existing precedent, ideal

contact geometries would be based on chemical intuition and/or quantum mechanical calculations. The geometric criteria were applied as follows. For every ligand variation, each of the geometric criteria was tested for satisfaction by contacts from any possible amino acid side-chain conformation in all designed protein positions. If a ligand variation was not able to make at least one of each type of user-specified contact, that ligand variation was discarded from the set. After geometric and steric pruning, the ligand variations remaining were only those theoretically capable of making each of the user-specified contacts.

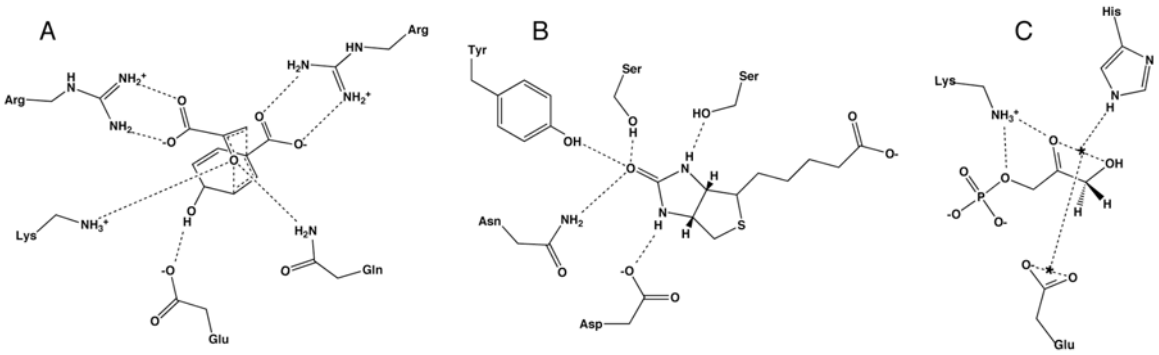


Figure 1. Contact geometries specified in small molecule pruning step. Ranges of distances, angles, and torsions were allowed that included the crystallographic geometries. (*A*) Chorismate mutase. (*B*) Biotin in streptavidin. (*C*) Triosephosphate isomerase Michaelis complex, modeled using an approach similar to that of reference 2. Asterisks indicate pseudoatoms used in geometry definitions.

In the third step, pairwise energies for all side-chain/side-chain, side-chain/backbone, backbone/ligand, and side-chain/ligand interactions were calculated using the full force field. In our work, this normally includes a scaled van der Waals term,[21] hydrogen-bonding and electrostatic terms,[22] and a solvation potential.[23,24]

The fourth step is an optional energy biasing that favors side-chain/ligand contacts deemed necessary for catalysis or binding. This energy biasing step helps to overcome the shortcomings of molecular mechanics energy functions as well as the inherent limitation of treating a multi-state design problem—*differential* stabilization of transition state relative to substrate in protein versus solution—using single-state design algorithms. As methods for modeling electrostatics and solvation and for designing over multiple states improve, the need for this biasing step should be reduced. Previous work

utilized selective application of solvation energy[1] or an additional search algorithm step[8] for the same purpose. We favor the use of adjustable bias energies that can be tailored for specific purposes and investigated as a design variable.

To implement the bias, user-specified energies were added or subtracted from pairwise side-chain/ligand interaction energies. We use the energy bias under two regimes, one for normal design calculations and another for rapid assessment of catalytic residue arrangements within a protein scaffold. In normal design calculations, a small energy benefit is simply applied to favor specified types of side-chain/ligand contacts. Alternatively, to quickly identify potential catalytic residues, exaggerated energetic benefits and penalties are applied together. A very large energy benefit is given for desired types of pairwise interactions (100 kcal/mol was used in the test cases reported here). An even larger energy penalty (10,000 kcal/mol here) is applied to all other pairwise side-chain/ligand interactions, except when the side chain is alanine or glycine. In other words, the energy penalty forces all designed side chains to alanine or glycine unless they participate in user-specified catalytic contacts with the ligand. Although this process clearly does not yield physically relevant energetics, it offers a useful tool to investigate the catalytic conformational space within a binding pocket. The tests performed here to study the effect of sampling parameters on calculation results took advantage of this second approach. Calculations performed to demonstrate sequence selection utilized the normal design approach of applying a simple energy benefit to catalytic contacts.

Finally, in the fifth step, optimal sequences were identified using the FASTER[25,26] or HERO[27] search methods. In the test cases described here, the result reported is the lowest-energy sequence with the maximal number of specified contacts.

Rotation-Translation Search

Simple rotation and translation can be used to fill the active site with an initial set of ligand variations in the first step of the process described. Because discrete steps must be used to rotate and translate the ligand, we evaluated the sensitivity of the calculation results to rotational and translational step sizes. A series of calculations was performed using an alanine-containing active-site background, as discussed in step 4 above. We

first tested different rotational step sizes using the crystallographic translational starting position with three initial random rotations. Backbone-dependent and backbone-independent rotamer and conformer libraries were tested. Each side-chain library was tested with and without inclusion of the specific crystallographic side-chain rotamers from the structure under examination.

As seen in Table 1, the results of these calculations (in terms of both RMSD relative to crystallographic position and number of wild-type contacts) were strongly dependent on the both the rotational step size and the rotamer library used. In the case of chorismate mutase, only the backbone-independent conformer library was able to find nativelike geometry and contacts. Figure 2 shows results from this library with the 5° step size. When the crystallographic rotamers were included in the calculation, however, all four libraries returned nativelike results. It should be noted that none of the three test case structures were included in the set of structures used to create the conformer libraries. The backbone-independent conformer library appeared the most consistently successful with the other two test cases as well, although it showed strong dependence on rotational step size in streptavidin.

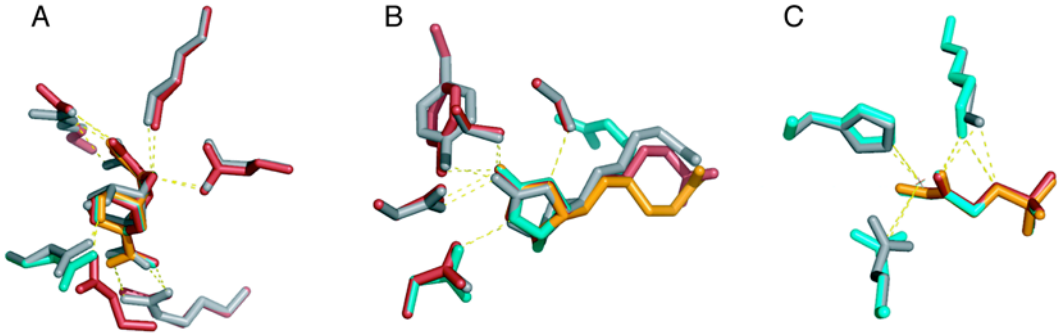


Figure 2. Sample results from test calculations presented in Table 1. Crystallographic side chains and ligands are shown in gray. Results from three trials using different initial random rotational positions are shown in red, teal, and orange. In cases where three colors are not visible, the selected rotamers from two or more calculations were identical. Results are shown from calculations with 5° rotation and the backbone-independent conformer library. (*A*) Chorismate mutase. An alternate backbone position was chosen for a glutamate-hydroxyl contact in one trial (red side chain, lower left). (*B*) Biotin in streptavidin. Note that the biotin pentanoic acid moiety samples different conformations in the calculation and the surrounding side chains were not designed. (*C*) Triosephosphate isomerase.

Table 1. RMSD and number of wild-type contacts as a function of rotational step size and rotamer library[a,b]

Chorismate Mutase

	Rotational step size				
Rotamer Library[c]	30°	20°	15°	10°	5°
Conformer: bb-ind	-	-	0.61 ± 0.03 (4.0)	0.55 ± 0.05 (4.0)	0.47 ± 0.04 (4.7)
with xtal rotamers	-	-	0.61 ± 0.03 (4.0)	0.55 ± 0.05 (4.0)	0.47 ± 0.04 (4.7)
Rotamer: bb-ind	-	-	3.88 ± 0.37 (0.0)	2.88 ± 1.44 (0.0)	3.01 ± 1.61 (0.0)
with xtal rotamers	-	-	1.57 ± 1.70 (2.7)	0.51 ± 0.00 (4.0)	0.52 ± 0.01 (4.0)
Conformer: bb-dep	-	-	3.66 ± 0.11 (1.0)	3.59 ± 0.08 (1.0)	3.60 ± 0.09 (1.0)
with xtal rotamers	-	1.67 ± 1.78 (3.3)	1.57 ± 1.83 (3.7)	0.60 ± 0.08 (4.3)	0.54 ± 0.06 (5.0)
Rotamer: bb-dep	-	-	-	-	-
with xtal rotamers	-	-	-	0.49 ± 0.04 (4.3)	0.52 ± 0.01 (4.0)

Streptavidin-Biotin

	Rotational step size				
Rotamer Library[c]	30°	20°	15°	10°	5°
Conformer: bb-ind	-	-	-	-	0.27 ± 0.09 (5.0)
with xtal rotamers	-	0.24 ± 0.09 (5.0)	0.24 ± 0.07 (5.0)	0.26 ± 0.06 (5.0)	0.20 ± 0.13 (5.0)
Rotamer: bb-ind	-	-	0.77 ± 0.42 (2.3)	0.60 ± 0.14 (3.0)	0.60 ± 0.05 (2.7)
with xtal rotamers	0.37 ± 0.17 (5.0)	0.24 ± 0.09 (5.0)	0.24 ± 0.07 (5.0)	0.26 ± 0.06 (5.0)	0.30 ± 0.17 (5.0)
Conformer: bb-dep	-	-	-	0.25 ± 0.12 (5.0)	0.20 ± 0.07 (5.0)
with xtal rotamers	-	0.24 ± 0.09 (5.0)	0.24 ± 0.07 (5.0)	0.22 ± 0.03 (5.0)	0.29 ± 0.09 (4.0)
Rotamer: bb-dep	-	-	-	0.82 ± 0.28 (2.3)	0.66 ± 0.02 (3.0)
with xtal rotamers	-	0.24 ± 0.09 (5.0)	0.24 ± 0.07 (5.0)	0.26 ± 0.06 (5.0)	0.16 ± 0.06 (5.0)

Triosephosphate Isomerase

	Rotational step size				
Rotamer Library[c]	30°	20°	15°	10°	5°
Conformer: bb-ind	-	1.87 ± 1.07 (0.7)	3.59 ± 2.28 (1.0)	0.28 ± 0.07 (3.0)	0.24 ± 0.05 (3.0)
with xtal rotamers	-	1.31 ± 0.29 (1.0)	1.95 ± 2.28 (1.3)	0.27 ± 0.06 (3.0)	0.15 ± 0.02 (3.0)
Rotamer: bb-ind	5.09 ± 0.05 (0.3)	0.60 ± 0.12 (1.7)	0.55 ± 0.25 (2.3)	0.34 ± 0.04 (2.3)	0.25 ± 0.08 (3.0)
with xtal rotamers	5.06 ± 0.05 (0.3)	0.60 ± 0.12 (2.0)	0.37 ± 0.04 (3.0)	0.25 ± 0.04 (3.0)	0.15 ± 0.02 (3.0)
Conformer: bb-dep	-	-	-	-	-
with xtal rotamers	-	-	-	-	0.15 ± 0.02 (3.0)
Rotamer: bb-dep	3.28 ± 0.73 (1.7)	0.60 ± 0.12 (1.7)	0.37 ± 0.05 (2.3)	0.31 ± 0.04 (2.3)	0.25 ± 0.08 (3.0)
with xtal rotamers	3.28 ± 0.73 (2.3)	0.60 ± 0.12 (2.3)	0.37 ± 0.05 (3.0)	0.29 ± 0.03 (3.0)	0.15 ± 0.02 (3.0)

[a] Dashes indicate that required contacts were not satisfied in at least one of three trials.
[b] Values are non-hydrogen-atom RMSD in Ångstroms relative to crystallographic ligands or bicyclic ring atom RMSD relative to crystallographic ligand for biotin (i.e., the pentanoic acid moiety was not considered in biotin RMSDs). Averages and standard deviations from three random initial positions are reported. Numbers in parentheses are the number of contacts where the amino acid position was the same as in the wild-type structure, averaged over the three trials. Maximum possible number of wild-type contacts: chorismate mutase, 5; streptavidin, 5; triosephosphate isomerase, 3.
[c] bb-ind: backbone-independent, bb-dep: backbone-dependent.

Next, we tested various combinations of rotational and translational step sizes starting from random initial ligand positions and using only the backbone-independent conformer library (Figure 3; Table 2). The crystallographic rotamers from the structures under investigation were not included in these calculations. The results show that, subject to the constraints imposed by the geometries defined in the pruning step and the biasing step, more than one combination of rotational and translational step size is viable for each test case and the sensitivity of the result to step size varies among the test cases.

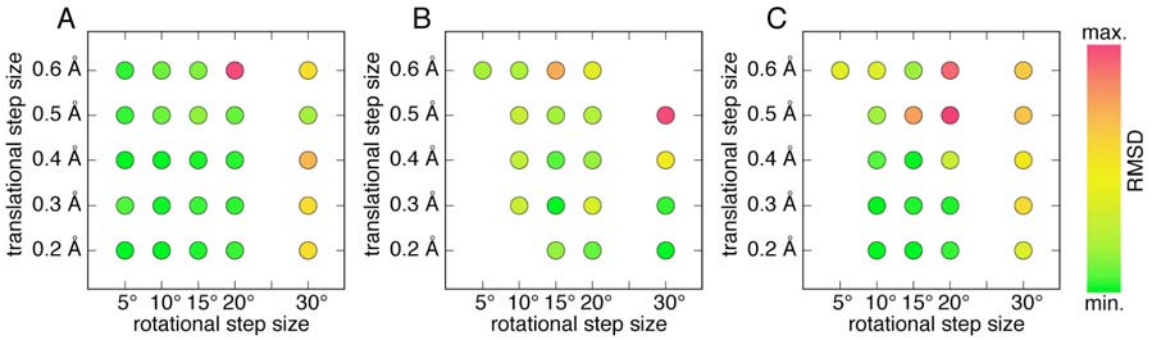


Figure 3. Effect of rotational and translational step sizes. Each spot represents the average of three trials with initial random starting positions. Missing points indicate that one or more trials could not identify wild-type-like contacts or else that the calculation was prohibitively large; no calculations were performed using a 25° rotational step size. Colors indicate non-hydrogen atom RMSD as described in the tables. (*A*) Chorismate mutase (min., 0.53 Å; max., 2.61 Å) (*B*) Streptavidin-biotin (min., 0.57 Å; max., 2.05 Å) (*C*) triosephosphate isomerase (min., 0.44 Å; max., 5.64 Å).

The rotation/translation tests were performed using three initial random starting positions for each system. The starting positions were created by randomly rotating and translating the ligand within a 1 $Å^3$ box around the ligand centroid (or the centroid of the bicyclic ring system in biotin). Using the same atom comparisons as described in the tables, the nine initial positions had RMSDs relative to crystallographic positions of between 2.1 Å and 4.5 Å, with an average of 3.2 Å. These tests do not provide full, unbiased searches of the active sites. Full active site searches could be conducted using this method by performing separate calculations for grid points distributed evenly through the active site. Given the time required to perform these smaller calculations (Table 2), searching an entire active site using rotational and translational perturbations would be computationally expensive. For example, examining a 3.6 x 3.6 x 3.6 Å grid

using the 10° and 0.3 Å step sizes would require an estimated 324 hours on a 16-processor cluster for placement of ligands and catalytic side chains in the chorismate mutase active site. Thus, for initial positioning of a ligand within an active site, rotational and translational placement is inefficient. However, the ability to adjust small molecule position and conformation simultaneously with side-chain optimization should be extremely valuable for refining an initial position identified from a coarser search method.

Table 2. RMSD and number of wild-type contacts as a function of rotational and translational step sizes[a,b]

Chorismate mutase

Translational step size (Å)	Rotational step size 30°	20°	15°	10°	5°	Time (10°, hours)[c]
0.6	1.69 ± 1.54 (2.3)	2.61 ± 1.67 (1.3)	0.77 ± 0.10 (4.3)	0.73 ± 0.02 (4.0)	0.61 ± 0.06 (4.7)	3
0.5	0.91 ± 0.20 (3.7)	0.72 ± 0.07 (4.0)	0.83 ± 0.06 (3.3)	0.74 ± 0.05 (4.0)	0.60 ± 0.13 (4.3)	10
0.4	2.02 ± 1.99 (2.3)	0.60 ± 0.04 (4.7)	0.59 ± 0.13 (4.0)	0.57 ± 0.12 (4.3)	0.53 ± 0.13 (4.3)	11
0.3	1.73 ± 1.51 (2.3)	0.61 ± 0.07 (4.3)	0.62 ± 0.15 (4.3)	0.58 ± 0.07 (4.0)	0.65 ± 0.04 (4.0)	12
0.2	1.71 ± 1.53 (2.3)	0.62 ± 0.10 (4.0)	0.60 ± 0.09 (4.0)	0.54 ± 0.07 (4.0)	0.56 ± 0.05 (4.0)	33

Streptavidin-biotin

Translational step size (Å)	Rotational step size 30°	20°	15°	10°	5°	Time (10°, hours)[c]
0.6	-	1.16 ± 0.60 (3.7)	1.67 ± 1.02 (3.7)	0.88 ± 0.44 (4.3)	0.84 ± 0.48 (4.3)	5
0.5	2.05 ± 0.59 (1.7)	0.91 ± 0.44 (5.0)	0.84 ± 0.61 (5.0)	0.99 ± 0.91 (3.7)	-	18
0.4	1.32 ± 1.39 (3.7)	0.80 ± 0.09 (5.0)	0.67 ± 0.28 (5.0)	0.96 ± 0.72 (3.7)	-	19
0.3	0.63 ± 0.16 (5.0)	1.08 ± 0.49 (5.0)	0.57 ± 0.21 (5.0)	1.03 ± 0.48 (4.3)	-	18
0.2	0.60 ± 0.32 (5.0)	0.70 ± 0.34 (5.0)	0.80 ± 0.24 (5.0)	-	-	-

Triosephosphate isomerase

Translational step size (Å)	Rotational step size 30°	20°	15°	10°	5°	Time (10°, hours)[c]
0.6	3.80 ± 2.14 (0.3)	5.22 ± 0.32 (0.0)	1.29 ± 0.91 (1.3)	2.39 ± 2.54 (1.7)	2.40 ± 2.58 (2.0)	0.4
0.5	3.92 ± 1.94 (0.0)	5.64 ± 0.45 (0.3)	4.47 ± 1.45 (0.0)	1.33 ± 1.01 (1.7)	-	2
0.4	3.13 ± 1.77 (0.3)	1.96 ± 1.05 (2.0)	0.47 ± 0.24 (1.7)	0.78 ± 0.66 (3.0)	-	2
0.3	3.44 ± 1.96 (0.3)	0.59 ± 0.18 (2.0)	0.60 ± 0.29 (2.3)	0.46 ± 0.11 (3.0)	-	2
0.2	2.33 ± 1.80 (0.7)	0.68 ± 0.10 (2.3)	0.49 ± 0.12 (3.0)	0.44 ± 0.11 (3.0)	-	5

[a] Dashes indicate that required contacts were not satisfied in at least one of three trials or that the calculation was too large to complete.

[b] Values are non-hydrogen atom RMSD in Ångstroms relative to crystallographic ligands or bicyclic atom RMSD relative to crystallographic ligand for biotin (i.e. the pentanoic acid moiety was not considered in biotin RMSDs). Averages and standard deviations from three random initial positions are reported. Numbers in parentheses are the number of contacts where the amino acid position was the same as in the wild-type structure, averaged over the three trials. Maximum possible number of wild-type contacts: chorismate mutase, 5; streptavidin, 5; triosephosphate isomerase, 3.

[c] Wall clock time; calculations performed on a 16-processor cluster.

Targeted Ligand Placement

A second approach places the small molecule with reference to a contacting side chain (Figure 4). In this approach, one or more small molecule variations are placed for every rotamer of the selected contacting side chain in every putative active-site position. This process has the advantage that ligand poses are targeted more efficiently to orientations that are able to make productive side-chain contacts. Previous computational enzyme design work utilized similar approaches.[1,2] In contrast to previous methods, however, our procedure does not maintain any association between the targeting rotamer and the small molecule—once the set of ligand conformations and orientations is constructed in step 1, the ligand variations are all subjected to pruning, pairwise energy calculations, and optimization as independent entities in the calculation. An implication of this procedure is that a ligand may engage in a catalytic contact with a rotamer, amino acid, or protein position that differs from those of the side-chain rotamer that was originally used to place that ligand.

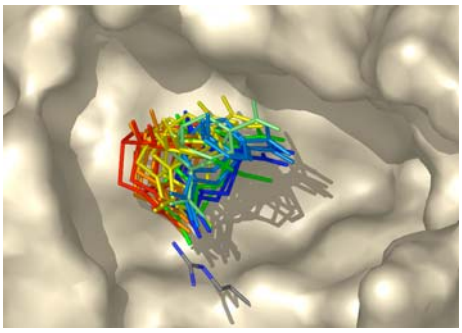


Figure 4. Targeted placement procedure. For a given side chain rotamer, small molecule ligands are placed such that they are able to meet specified geometric criteria. This is repeated for every possible conformation of the amino acid at every designed position. Shown is a subset of orientations of a chorismate mutase transition-state structure in contact with one conformation of arginine.

We tested the effect of four types of side-chain libraries on the ability of a targeted placement process to find wild-type-like ligand positions and contacts. For the three test cases, the following side-chain contacts were used to anchor the ligand: chorismate mutase, C11 carboxylate to arginine; streptavidin, N1 to aspartate; triosephosphate isomerase, O2 and O3 to histidine. For each contact type, variations were allowed in the geometry of the contact, including the contacting atoms (NH1-NH2

vs. NE-NH1 for arginine) and variations in defined distances, angles, and dihedrals of the contact.

As with the rotational and translational search, success in achieving native active-site conformations was highly dependent on the side-chain library used (Table 3). Only the backbone-independent conformer library yielded results for all three test cases that were comparable to those with crystallographic rotamers included. Using that library, all three systems returned all wild-type contacts with low ligand RMSD relative to the crystallographic position. As with the rotation/translation search, the chorismate mutase case showed the strongest sensitivity to rotamer library. Inspection of the structures revealed that an arginine side chain (Arg 28) occupies a conformation in the inhibitor-bound, active enzyme structure that was not well approximated in the other rotamer libraries.

The targeted placement approach allowed a thorough and directed search of active-site conformational space, including between 10^6 and 10^9 small molecule orientations and conformations spread throughout the active site. In contrast to the rotation/translation method, a full active-site search took between one and eighteen hours to complete using the backbone-independent conformer library and no initial starting position was required. This method offers an efficient first step for defining active-site geometry in a new protein scaffold. One shortcoming is that it may be difficult to sample the many geometrical variations of a flexible hydrogen-bonding interaction. For example, the 972 variations in guanidino-carboxylate contact geometry sampled in the chorismate mutase case are probably adequate to reflect flexibility in this relatively rigid dual hydrogen-bonding interaction. A less restrained interaction, however, such as a serine hydrogen bonding with a sterically unrestricted ligand carbonyl oxygen, results in a compromise between maintaining a manageable calculation size and modeling contact flexibility. One solution is to use a targeted method to find an initial ligand position within the binding site and then, in a second calculation, optimize both active-site packing and fine rotational and translational placement of the ligand.

Table 3. Results from targeted placement procedure as a function of rotamer library.

Chorismate mutase

Rotamer library[a]	log(initial ligand variations)	RMSD (Å)[b] (WT contacts)	Time (hours)[c]
Conformer: bb-ind	7.88	0.60 (5)	16
with xtal rotamers	7.88	0.68 (3)	18
Rotamer: bb-ind	8.18	3.61 (0)	51
with xtal rotamers	8.18	0.66 (4)	62
Conformer: bb-dep	7.64	3.62 (1)	8
with xtal rotamers	7.64	0.68 (4)	9
Rotamer: bb-dep	7.76	2.31 (1)	14
with xtal rotamers	7.76	0.66 (4)	16

Streptavidin-biotin

Rotamer library[a]	log(initial ligand variations)	RMSD (Å)[b] (WT contacts)	Time (hours)[c]
Conformer: bb-ind	7.07	0.64 (5)	1.4
with xtal rotamers	7.07	0.64 (5)	1.4
Rotamer: bb-ind	7.20	0.54 (4)	3.5
with xtal rotamers	7.20	0.34 (4)	3.4
Conformer: bb-dep	6.35	0.37 (5)	0.2
with xtal rotamers	6.35	0.54 (4)	0.2
Rotamer: bb-dep	7.17	3.50 (0)	2.6
with xtal rotamers	7.17	0.19 (5)	2.8

Triosephosphate isomerase

Rotamer library[a]	log(initial ligand variations)	RMSD (Å)[b] (WT contacts)	Time (hours)[c]
Conformer: bb-ind	7.31	0.49 (3)	1.3
with xtal rotamers	7.31	0.49 (3)	1.3
Rotamer: bb-ind	7.78	0.46 (3)	8.7[d]
with xtal rotamers	7.78	0.46 (3)	8.7[d]
Conformer: bb-dep	6.82	7.51 (0)	0.3
with xtal rotamers	6.82	0.78 (3)	0.3
Rotamer: bb-dep	7.58	0.51 (3)	4.3[d]
with xtal rotamers	7.58	0.51 (3)	4.9[d]

[a] bb-ind, backbone-independent; bb-dep, backbone-dependent.
[b] RMSDs calculated as described in Table 1. Maximum possible number of wild-type contacts: chorismate mutase, 5; streptavidin, 5; triosephosphate isomerase, 3.
[c] Wall clock time; calculations performed on a 16-processor cluster.
[d] Calculation was performed as a series of smaller calculations.

Sequence Design

The computational tests described in the previous sections were designed to evaluate the effects of calculation parameters on recovery of native enzyme geometries, and the design of active-site residues was limited to catalytic side chains. However, the general procedure described here is equally amenable to full active-site design calculations.

In previously published work, 18 active site residues of *E. coli* chorismate mutase were redesigned simultaneously with rotational and translational relaxation of the transition-state structure from the starting crystallographic position.[28] The six predicted mutations were experimentally investigated and some were found to confer increased catalytic efficiency[28] or thermostability (see chapter 4). A detrimental mutation predicted in the study underscored the importance of continued work on energy functions. In the calculation that motivated this experimental work, the initial starting position of the small molecule was taken from the crystal structure and a limited degree of rotational and translational optimization was employed.

We performed a test calculation to demonstrate that small molecules can be placed simultaneously with full active-site side-chain optimization, without reference to any known starting position. In a sample calculation using *E. coli* chorismate mutase, the targeted placement method was used to identify 10^7 small molecule variations. In this example, after the geometric pruning step and elimination of variants with backbone steric clashes, 155 small molecule variations remained. These variants were evaluated combinatorially with ten different side chain identities in twelve active-site positions. Using FASTER for optimization, the calculation took approximately 9 hours to complete on a 16-processor cluster with about 70% of the total calculation time consumed in calculating a surface-area-based solvation term.

Conclusions

The described procedures allow the incorporation of small molecule placement directly into sequence design calculations. The test calculations performed suggest that the results of computational enzyme design processes can be quite sensitive to calculation parameters including the rotamer library used and the coarseness of ligand positioning. These results emphasize that the conformational space of a calculation must be explored before meaningful conclusions can be reached about energy functions.

Given that we still have much to learn about the complex relationship between protein structure and catalytic activity,[29,30] luck and choice of system may continue to play a role in the success of *de novo* computational enzyme design efforts for some time. However, the power of computational enzyme design to stringently evaluate our

understanding of the energetics of catalysis should not be overlooked. Experimental feedback gained from both successful and unsuccessful designs will make it possible to critically examine energy functions for modeling active sites. Employing quality transition-state structures derived from *ab initio* calculations and experimental evidence will help computational design experiments to provide more meaningful information about the effectiveness of energy functions. The use of large side-chain structural libraries and fine movements of transition-state structures will help to reduce errors from conformational sampling. Backbone relaxation and multi-state design will offer other important tools to improve the value of design calculations. Finally, the construction of gene libraries or large numbers of computationally designed variants has great potential for overcoming the shortcomings of enzyme design models,[31] but results from these experiments will be most useful for furthering our understanding of catalysis and design if both active and inactive variants are reported. By critically evaluating current methods for computational enzyme design, we will move closer to a deeper and more practically useful understanding of the sequence determinants of enzyme activity in the future.

Methods

Structures and charges

PDB files were used without minimization (*E. coli* chorismate mutase,[32] 1ecm; *S. avidinii* streptavidin,[33] 1mk5; *S. cerevisiae* triosephosphate isomerase,[34] 1ney. Hydrogens were added with REDUCE.[35]

A library of ligand internal conformations was created for each system as follows. Chorismate mutase: An HF/6-31G* *ab initio* transition-state structure[36] was used with only one variation—the O4 hydroxyl proton was allowed to occupy three positions, 60°, 180°, and -35°, defined by the H-C-O-H dihedral angle. The minima in a torsional profile at the HF/6-31G* level were at approximately 180° and -35°, and 60° was included as an option because hydrogen-bonding patterns in chorismate mutases from other species suggested population of that region of torsional space. Streptavidin: Four rotatable bonds in biotin were allowed to occupy three positions each (60°, -60°, 180° for sp3-sp3 bonds and 30°, 90°, 150° for the symmetric carboxylate group). Thirty-four

conformations were excluded because of high internal energy calculated using the van der Waals component of the DREIDING force field.[37] Triosephosphate isomerase: The pdb structure used was the Michaelis complex with the substrate dihydroxyacetone phosphate. In ground-state dihydroxyacetone phosphate, two rotatable bonds (defined by the P-O-C-C and C-C-O-H dihedral angle) were allowed to occupy three positions each (60°, -60°, 180°). Three conformations were excluded because of high internal DREIDING van der Waals energy.

Ligand atomic charges were obtained by fitting charges to electrostatic potential from HF/6-31G* single-point energy calculations using the transition-state structure (chorismate mutase) or crystallographic ground-state structure (biotin, dihydroxyacetone phosphate). *Ab initio* calculations and charge determinations were performed using Spartan (Wavefunction, Inc.) or Jaguar (Schrödinger, Inc.).

Side-chain rotamer libraries

Standard backbone-dependent and backbone-independent rotamer libraries were used with expansion by one standard deviation about $\chi 1$ and $\chi 2$.[17]

Crystallographic conformer libraries were prepared using coordinates from 149,813 side chains selected from 1,011 unique structures. A clustering algorithm was developed based on ideas described by Shetty *et al*[20] and is described briefly here. Every side-chain conformation from the raw data set is assigned to exactly one cluster. Each cluster is represented by the centroid, which is the member with coordinates closest to the average coordinates of all cluster members. A conformer library consists of a list of all of the cluster representatives and their coordinates. In our clustering algorithm, clusters are assigned through discrete clustering moves: *Switch* allows a single raw conformer to leave one cluster and join another; *Merge* combines two clusters into one; *Split* allows a raw conformer to start a new cluster on its own. These moves are depicted in Figure 5.

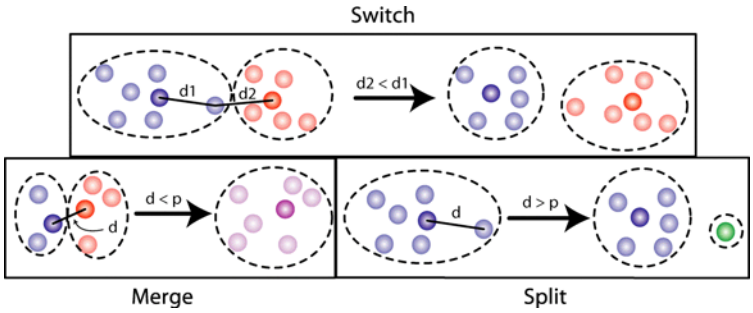


Figure 5. The three clustering moves are illustrated by showing the state of a sample system before and after the move is performed. Each dot represents a single side-chain conformation taken from the PDB. Distances represent side-chain RMSDs between pairs of conformers. Dots sequestered together by a dashed line and colored the same are members of the same cluster. Darker-colored dots denote cluster representatives.

RMSDs between pairs of conformers are compared to determine whether or not to apply a particular move. *Switch* is applied so that each raw conformer is a member of the cluster whose centroid is closest to it. *Merge* and *Split* are applied based on the value of the clustering parameter p: two clusters are merged if their centroids are within p of each other, whereas a conformer splits off and starts a new cluster if the closest centroid of any existing cluster is farther than p from it. The clustering moves are applied as follows until the number of clusters converges:

1. Start with a small number of clusters (1 was used in this work), and randomly assign a single raw conformer to each as the sole member and cluster representative.
2. Assign each raw conformer in the data set to the cluster whose centroid is closest.
3. While the number of clusters is not converged:
    a. Iteratively attempt to *Merge* pairs of clusters until no cluster can be further merged.
    b. For each conformer C:
        i. Measure the distance d between C and the centroid of every existing cluster.
        ii. If the distance d to the closest cluster centroid is greater than p, *Split* C off as its own cluster.
        iii. Else, *Switch* C to the closest cluster.
        iv. Recompute the centroid for every cluster that has changed membership.

The algorithm allows the construction of both backbone-dependent and backbone-independent libraries to custom sizes by using clustering factor *p* to define the desired degree of similarity between independent conformers. In this work, clustering factors of 0.3 Å and 1.0 Å were used for backbone-dependent and backbone-independent rotamer libraries, respectively.

For all calculation types, conformer libraries were smaller than the standard rotamer libraries. As an example, the number of side-chain conformations for the chorismate mutase calculations described in Table 3 were as follows: backbone-independent rotamer, 14229; backbone-independent conformer, 5955; backbone-dependent rotamer, 7945; and backbone-dependent conformer, 5539.

Calculation Parameters

All non-Gly, non-Pro residues reasonably within the natural active sites were included in calculations. Residues with any atom within a 5 Å radius from any atom in the crystallographic ligands were included, less those residues separated from the natural ligand by backbone elements and plus a few adjacent residues not within the 5 Å cutoff. The positions designed were (all in chain A unless otherwise designated): chorismate mutase, 28, 32, 35, 39, 46, 47, 48, 51, 52, 55, 81, 84, 85, 88, 7B, 11B, 14B, 18B; streptavidin, 23, 24, 25, 27, 43, 45, 46, 47, 49, 50, 79, 86, 88, 90, 92, 108, 110, 112, 128, 130; and triosephosphate isomerase, 10, 12, 95, 97, 165, 170, 211, 230.

In ligand placement test cases, designed residues were restricted to ligand-contacting residues or alanine as follows: Arg, Lys, Gln, Glu, or Ala in chorismate mutase; Ser, Asn, Tyr, Asp, or Ala in streptavidin, and Glu, His, Lys, or Ala in triosephosphate isomerase. Four calculations on triosephosphate isomerase were run as smaller component calculations, as indicated in Table 2, because of prohibitive size as a single calculation.

Energy Functions and optimization

Energy functions included scaled van der Waals,[21] hydrogen-bonding, and electrostatic terms.[22] A surface-area based solvation potential[23] was used in sequence design calculations but not for ligand placement, where solvation energy would have

been heavily outweighed by geometric considerations. Sequences were optimized with respect to the energy function using FASTER[25,26] or HERO.[27] On occasion, a top-ranked sequence contained more than one instance of a given specified geometric contact, owing to the energy benefit applied for these contacts. In these cases, Monte Carlo[38,39] was used to sample around the global minimum energy sequence and the top-ranked sequence with a single instance of each geometric contact was reported.

References:

1. Bolon, D. N. and Mayo, S. L. (2001) Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. USA 98*, 14274-14279.

2. Dwyer, M.A., Looger, L.L. and Hellinga, H.W. (2004) Computational design of a biologically active enzyme. *Science* 304, 1967-1971.

3. Mendes, J., Guerois, R. and Serrano, L. (2002) Energy estimation in protein design. *Curr. Opin. Struct. Biol. 12*, 441-446.

4. Vizcarra, C.L. and Mayo, S.L. (2005) Electrostatics in computational protein design. *Curr. Opin. Chem. Biol. 9*, 622-626.

5. Dahiyat, B.I. and Mayo, S.L. (1997) De novo protein design: Fully automated sequence selection. *Science 278*, 82-87.

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

7. Shimaoka, M., Shifman, J.M., Jing, H., Takagi, J., Mayo, S.L. and Springer, T.A. (2000) Computational design of an integrin I domain stabilized in the open high affinity conformation. *Nat. Struct. Biol.*, *7*, 674-678.

8. Looger, L. L., Dwyer, M. A., Smith, J. J. and Hellinga, H. W. (2003) Computational design of receptor and sensor proteins with novel functions. *Nature 423*, 185-190.

9. Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L. and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science 302*, 1364-1368.

10. Pierce, N.A. and Winfree, E. (2002) Protein Design is NP-hard. *Prot. Eng. 15*, 779-782.

11. Taylor, R.D., Jewsbury, P.J. and Essex, J.W. (2002) A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des. 16*, 151-166.

12. Lilien, R.H., Stevens, B.W., Anderson, A.C. and Donald, B.R. (2005) A novel ensemble-based scoring and search algorithm for protein redesign and its application to modify the substrate specificity of the gramicidin synthetase A phenylalanine adenylation enzyme. *J. Comput. Biol. 12*, 740-761.

13. Chakrabarti, R., Klibanov, A.M. and Friesner, R.A. (2005) Computational prediction of native protein ligand-binding and enzyme active site sequences. *Proc. Natl. Acad. Sci. USA 102*, 10153-10158.

14. Chakrabarti, R., Klibanov, A.M. and Friesner, R.A. (2005) Sequence optimization and designability of enzyme active sites. *Proc. Natl. Acad. Sci. USA 102*, 12035-12040.

15. Hellinga, H.W. and Richards, F.M. (1991) Construction of new ligand binding sites in proteins of known structure I. Computer-aided modeling of sites with pre-defined geometry. *J. Mol. Biol. 222*, 763-785.

16. Ponder, J.W. and Richards, F.M. (1987) Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol. 193*, 775-791.

17. Dunbrack, R.L, Jr. and Cohen, F.E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Prot. Sci. 6*, 1661-1681.

18. Lovell, S.C., Word, J.M., Richardson, J.S. and Richardson, D.C. (2000) The penultimate rotamer library. *Proteins 40*, 389-408.

19. Xiang, Z. and Honig, B. (2001) Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol. 311*, 421-430.

20. Shetty, R.P., de Bakker, P.I.W., DePristo, M.A. and Blundell, T.L. (2003) Advantages of fine-grained side chain conformer libraries. *Prot. Eng. 16*, 963-969.

21. Dahiyat, B.I. and Mayo, S.L. (1997) Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. USA 94*, 10172-10177.

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

23. Street, A.G. and Mayo, S.L. (1998) Pairwise calculation of protein solvent-accessible surface areas. *Fold. Des. 3*, 253-258.

24. Lazaridis, T. and Karplus, M. (1999) Effective energy functions for proteins in solution. *Prot. Struct. Funct. Genet. 35*, 133-152.

25. Desmet, J., Spriet, J. and Lasters, I. (2002) Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Prot. Struct. Funct. Genet. 48*, 31-43.

26. Allen, B.D. and Mayo, S.L. (2006) Dramatic performance enhancements for the FASTER optimization algorithm. *J. Comput. Chem. 27*, 1071-1075.

27. Gordon, D.B., Hom, G.K., Mayo, S.L. and Pierce, N.A. (2003) Exact rotamer optimization for protein design. *J. Comput. Chem. 24*, 232-243.

28. Lassila, J.K., Keeffe, J. R., Oeschlaeger, P. and Mayo, S.L., (2005) Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity. *Protein Eng. Des. Sel. 18*, 161-163.

29. Kraut, D.A., Carroll, K.S. and Herschlag, D. (2003) Challenges in enzyme mechanism and energetics. *Annu. Rev. Biochem. 72*, 517-571.

30. Benkovic, S.J. and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis. *Science 301*, 1196-1202.

31. Bolon, D.N., Voigt, C.A. and Mayo, S.L. (2002) *De novo* design of biocatalysts. *Curr. Opin. Chem. Biol. 6*, 125-129.

32. Lee, A.Y., Karplus, P.A., Ganem, B. and Clardy, J. (1995) Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase. *J. Am. Chem. Soc. 117*, 3627-3628.

33. Hyre, D.E., Le Trong, I., Merritt, E.A., Eccleston, J.F., Green, N.M., Stenkamp, R.E. and Stayton, P.S. (2006) Cooperative hydrogen bond interactions in the streptavidin-biotin system. *Protein Sci. 15*, 459-467.

34. Jogl, G., Rozovsky, S., McDermott, A.E. and Tong, L. (2003) Optimal alignment for enzymatic proton transfer: Structure of the Michaelis complex of triosephosphate isomerase at 1.2-Ångstrom resolution. *Proc. Natl. Acad. Sci. USA 100*, 50-55.

35. Word, J.M., Lovell, S.C., Richardson, J.S. and Richardson, D.C. (1999) Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol. 285*, 1735-1747.

36. Wiest, O. and Houk, K. N. (1994) On the transition state of the chorismate-prephenate rearrangement. *J. Org. Chem. 59*, 7582-7584.

37. Mayo, S. L., Olafson, B. D. and Goddard, W. A. (1990) DREIDING: A generic force field for molecular simulations. *J. Phys. Chem. 94*, 8897-8909.

38. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys. 21*, 1087-1092.

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

Chapter 3

Computationally Designed Variants of *Escherichia Coli* Chorismate Mutase Show Altered Catalytic Activity

The text of this chapter has been adapted from a manuscript coauthored with Jennifer R. Keeffe, Peter Oelschlaeger, and Stephen L. Mayo.

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

Abstract

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

Introduction

The Claisen rearrangement of chorismate to prephenate (Figure 1) is a rare enzyme-catalyzed pericyclic reaction that proceeds through the same mechanism uncatalyzed in solution. Chorismate mutases from various organisms provide rate enhancements of around 10^6 despite strong dissimilarities in three-dimensional fold.[1-3] The metabolic importance of chorismate as the key branch point in the shikimate pathway has prompted extensive experimental investigation of the chorismate-prephenate rearrangement since the 1960s[4] and has driven complementation experiments to probe the structural determinants of enzyme catalysis.[5] The concerted, unimolecular nature of the rearrangement and the lack of covalent protein contacts have encouraged numerous theoretical studies of the catalyzed and uncatalyzed reactions. The question of how chorismate mutases achieve rate enhancement has been actively discussed in recent years.[6-10]

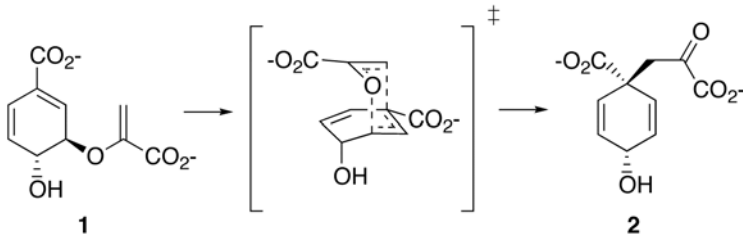


Figure 1. Claisen rearrangement of chorismate (1) to prephenate (2).

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

Materials and Methods

Computational design

An *ab initio* calculated transition state structure[12] was modeled at the position of a transition state analog in the EcCM crystal structure, pdb code 1ecm.[3] Translation (± 0.2 Å each axis) and rotation (± 5° each axis) of the transition state structure was allowed and all amino acids except glycine, proline, cysteine, and methionine were permitted in the following positions: 28, 32, 35, 39, 46, 47, 48, 51, 52, 55, 81, 84, 85, 88 (chain A); 7, 11, 14, 18 (chain B). Residues from both chains compose each of the two active sites in this symmetric homodimer. A backbone-dependent side chain rotamer library[13] was used with expansion by one standard deviation about $\chi 1$ and $\chi 2$ for all amino acids except arginine and lysine. The energy function used in the ORBIT protein design software[14] was based on the DREIDING force field[15] and includes a scaled van der Waals term,[16] hydrogen bonding and electrostatic terms,[17] and a surface-area based solvation potential.[18] Transition state partial atomic charges were obtained as previously reported.[12] An additional energy penalty was applied to effectively eliminate from consideration all sequences that could not maintain key contacts between the transition state structure and Arg 11 and Arg 28. These contacts were previously demonstrated to be necessary for catalysis.[19] The HERO rotamer optimization method was used to obtain the minimum energy amino acid sequence and conformations.[20] Six mutations were predicted in the optimized structure: Leu7Ile, Ala32Ser, Val35Ile, Asp48Ile, Ile81Leu, and Val85Ile. Visual inspection and subsequent calculations indicated that four of the mutations were independent, so they were treated separately.

Protein expression and kinetic characterization

The gene encoding EcCM residues 1-109 of the bifunctional chorismate mutase-prephenate dehydratase (P protein) was amplified from genomic DNA (ATCC 700926D) and inserted into the pTYB11 vector from the IMPACT-CN intein fusion system (New England Biolabs) between the *Sap*I and *Xho*I sites. Inverse PCR mutagenesis[21] was used to construct five variants: Leu7Ile, Ala32Ser, Val35Ile, Asp48Ile, and Ile81Leu/Val85Ile. Constructs were verified by DNA sequencing. Variant and wild-type proteins were expressed in *E. coli* BL21 (DE3). Expression was induced with 1 mM IPTG at OD_{600} =

0.6 and cells were grown for 18 hours at 22 °C. Cells were harvested and lysed by French press in 20 mM Tris-Cl pH 8.0, 500 mM NaCl, 1 mM EDTA, 10 mM $MgCl_2$, 1 mM PMSF, with DNase and RNase. Chitin affinity purification was conducted with a column buffer containing 20 mM Tris-Cl, pH 8.0, 500 mM NaCl, and 1 mM EDTA. On-column cleavage proceeded in 50 mM DTT at 25 °C for 18 hours. Mass spectrometry and SDS-PAGE verified that this protocol results in successful cleavage and yields the expected product. Wild-type and variant proteins were further purified by gel filtration on a HiPrep Sephacryl S-100 high resolution column (Amersham Biosciences) with 20 mM Tris-Cl pH 7.8, 100 mM NaCl as the running buffer and final storage buffer. Protein characterization followed procedures recently reported for the same construct.[22] Protein concentration was determined by Bradford assay using BSA as a standard. Chorismate mutase activity was determined by following the disappearance of chorismate with UV absorbance at 275 nm. Activity assays were conducted at 37 °C and contained 20 nM protein in 50 mM Tris pH 7.8, 2.5 mM EDTA, 20 mM β-mercaptoethanol, and 0.01 % BSA. Initial velocities were buffer-corrected and were determined with less than 6% depletion of initial substrate concentration. All proteins were initially tested using a substrate concentration range of about 50-500 μM. The wild-type and the Ala32Ser mutant were further assayed with substrate ranges of approximately 20-2000 μM and a minimum of five trials including two separate protein preparations each. Kinetic parameters were determined by nonlinear fitting to the Michaelis-Menten equation.

Molecular dynamics simulations

Molecular dynamics simulations of EcCM and mutant Asp48Ile as free enzymes and in complex with the oxabicyclic transition state analog[23] were carried out using the *Sander* Program of the *AMBER 7.0* software package.[24] For symmetry, both chains were truncated to residues 6 through 95. The transition state analog was minimized and the electrostatic potential surface was calculated using DFT with B3LYP and the lacvp** basis set in the Jaguar 5.5 package (Schrödinger LLC). Atomic charges were obtained with *RESP*. Parameters for the transition state analog were defined according to similar structures in the *AMBER* libraries. The binding sites of the free enzymes were filled with water by solvating the molecules with a TIP3P[25] water shell using a closeness parameter

of 0.4 in *xLEaP*. All systems were then solvated by a truncated octahedron of TIP3P water 8 Å around the protein and neutralized with counter ions. The ff02EP version of the ff99 force field[26,27] was used to represent the protein and the ligand. After minimization, all systems were heated to 300 K with constant volume, starting at 10 K with three different initial velocities. Molecular dynamics simulations were prolonged at 300 K and constant pressure until the backbone RMSD reached a constant value of about 1.5 Å compared to the first frame; this was observed after 500 to 800 ps. Average structures over 10 ps of the equilibrated systems were generated and analyzed visually.

Results and Discussion

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

Table 1. Kinetic parameters of wild-type and mutant EcCM[a]

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

[a] Reported errors are standard deviations from at least three trials.
[b] Three mutants were assayed with a limited substrate concentration range, see text.
[c] Reaction rates with the Asp48Ile mutant were within error of the uncatalyzed solution reaction.

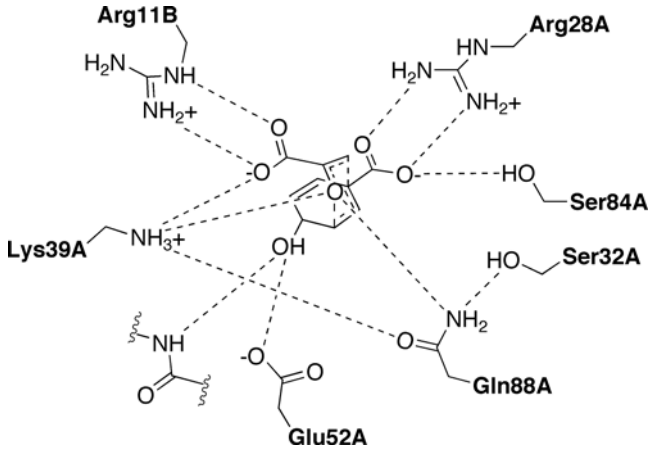


Figure 2. Predicted hydrogen bonding in the Ala32Ser chorismate mutase variant. Interactions in the wild type crystal structure are the same except for position 32A.

Val35Ile shows increased k_{cat} but also increased K_M, resulting in k_{cat}/K_M similar to wild type. The Leu7Ile mutation also does not have a significant effect on catalytic efficiency. Both mutations create slightly different hydrophobic packing environments near the essential residue Arg 11 in the predicted structures. The double mutant, Ile81Leu/Val85Ile, was predicted to alter packing against the hydrophobic ring face of the reacting molecule. This rearrangement of hydrophobic amino acids does not result in substantial change in k_{cat}. However, the mutations result in increased K_M and reduced catalytic efficiency. The relative insensitivity of the enzyme to changes in some amino acids is not surprising given the recent finding that the reaction is efficiently catalyzed by a protein exhibiting all the characteristics of a molten globule.[29]

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

that one role of Asp 48 may be to stabilize this key active site side chain in a conformation compatible with substrate binding and catalysis.

While the choice of mutations in this experiment was based solely on computational modeling and no sequence alignment information was used in the process, a BLAST search[31] using the EcCM sequence as the query showed that sequence variations corresponding to the Ala32Ser, Val35Ile, Val85Ile, and Ile81Leu mutations were observed in chorismate mutases from related organisms.

Our design procedure stabilizes a static active site configuration with a bound transition state structure. Although the substrate of the reaction is not explicitly considered, we expect that modeling interactions using the structure and charges of the transition state should promote some degree of differential stabilization of the transition state relative to substrate. The present study demonstrates that this approach can be used effectively to represent the active site of a natural enzyme. In this case, the predicted mutations were in residues not directly contacting the reacting molecule. The favorable result from the Ala32Ser mutation suggests that such secondary contacts are important in the enzyme design process. The complete loss of catalytic activity from the Asp48Ile mutation implies that improved treatment of electrostatics and consideration of the unbound enzyme could offer some benefit in future design efforts.

References

1. Chook, Y. M., Hengming, K. and Lipscomb, W. N. (1993) Crystal-structures of the monofunctional chorismate mutase from *Bacillus-subtilis* and its complex with a transition-state analog. *Proc. Natl Acad. Sci. USA 90*, 8600-8603.

2. Xue, Y., Lipscomb, W. N., Graf, R., Schnappauf, G. and Braus, G. (1994) The crystal-structure of allosteric chorismate mutase at 2.2-Ångstrom resolution. *Proc. Natl Acad. Sci. USA 91*, 10814-10818.

3. Lee, A. Y., Karplus, P. A., Ganem, B. and Clardy, J. (1995) Atomic-structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase. *J. Am. Chem. Soc. 117*, 3627-3628.

4. Gibson, M. I. and Gibson, F. (1964) Preliminary studies on isolation + metabolism of intermediate in aromatic biosynthesis - chorismic acid. *Biochem. J. 90*, 248-256.

5. Woycechowsky, K. J. and Hilvert, D. (2004) Deciphering enzymes – Genetic selection as a probe of structure and mechanism. *Eur. J. Biochem. 271*, 1630-1637.

6. Guimarães, C. R. W., Repasky, M. P., Chandrasekhar, J., Tirado-Rives, J. and Jorgensen, W. L. (2003) Contributions of conformational compression and preferential transition state stabilization to the rate enhancement by chorismate mutase. *J. Am. Chem. Soc. 125*, 6892-6899.

7. Strajbl, M., Shurki, A., Kato, M. and Warshel, A. (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J. Am. Chem. Soc. 125*, 10228-10237.

8. Hur, S. and Bruice, T. C. (2003) The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proc. Natl Acad. Sci. USA 100*, 12015-12020.

9. Crespo, A., Scherlis, D. A., Martí, M. A., Ordejón, P., Roitberg, A. E. and Estrin, D. A. (2003) A DFT-based QM-MM approach designed for the treatment of large molecular systems: Application to chorismate mutase. *J. Phys. Chem. B 107*, 13728-13736.

10. Ranaghan, K. E., Ridder, L., Szefczyk, B., Sokalski, W. A., Hermann, J. C. and Mulholland, A. J. (2004) Transition state stabilization and substrate strain in enzyme catalysis: *Ab initio* QM/MM modeling of the chorismate mutase reaction. *Org. Biomol. Chem. 2*, 968-980.

11. Stewart, J., Wilson, D. B. and Ganem, B. (1990) A genetically engineered monofunctional chorismate mutase. *J. Am. Chem. Soc. 112*, 4582-4584.

12. Wiest, O. and Houk, K. N. (1994) On the transition-state of the chorismate-prephenate rearrangement. *J. Org. Chem. 59*, 7582-7584.

13. Dunbrack, R. L. and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci. 6*, 1661-1681.

14. Dahiyat, B. I. and Mayo, S. L. (1997) De novo protein design: Fully automated sequence selection. *Science 278*, 82-87.

15. Mayo, S. L., Olafson, B. D. and Goddard, W. A. (1990) DREIDING – A generic force-field for molecular simulations. *J. Phys. Chem. 94*, 8897-8909.

16. Dahiyat, B. I. and Mayo, S. L. (1997) Probing the role of packing specificity in protein design. *Proc. Natl Acad. Sci. USA 94*, 10172-10177.

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

18. Street, A. G. and Mayo, S. L. (1998) Pairwise calculation of protein solvent-accessible surface areas. *Fold. Des. 3*, 253-258.

19. Liu, D. R., Cload, S. T., Pastor, R. M. and Schultz, P. G. (1996) Analysis of active site residues in *Escherichia coli* chorismate mutase by site-directed mutagenesis. *J. Am. Chem. Soc. 118*, 1789-1790.

20. Gordon, D. B., Hom, G. K., Mayo, S. L. and Pierce, N. A. (2002) Exact rotamer optimization for protein design. *J. Comput. Chem. 24*, 232-243.

21. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. and Galas, D. J. (1989) A simple method for site-directed mutagenesis using the polymerase chain-reaction. *Nucleic Acids Res. 17*, 6545-6551.

22. Zhang, S., Wilson, D. B. and Ganem, B. (2003) An engineered chorismate mutase with allosteric regulation. *Bioorg. Med. Chem. 11*, 3109-3114.

23. Bartlett, P. A., Johnson, C. R. (1985) An inhibitor of chorismate mutase resembling the transition-state conformation. *J. Am. Chem. Soc. 107,* 7792-7793.

24. Case, C. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M. Jr., Stanton, R. V., Cheng, A. L., Vincent, J. J.; Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, J., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., Kollman, P. A. AMBER 7, University of California, San Francisco, 2002.

25. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926-935.

26. Wang, J., Cieplak, P., Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem. 21,* 1049-1074.

27. Cornell, W. D., Cieplak, J. W., Bayly, C. I., Gould, I. R., Merz, K. M. Jr., Ferguson, D. M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A. (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc. 117,* 5179-5197.

28. Strajbl, M., Shurki, A., Kato, M. and Warshel, A. (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J. Am. Chem. Soc. 125*, 10228-10237.

29. Vamvaca, K., Vögeli, B., Kast, P., Pervushin, K. and Hilvert, D. (2004) An enzymatic molten globule: Efficient coupling of folding and catalysis. *Proc. Natl Acad. Sci. USA 101*, 12860-12864.

30. Kast, P., Hartgerink, J. D., Asif-Ullah, M. and Hilvert, D. (1996) Electrostatic catalysis of the Claisen rearrangement: probing the role of Glu78 in *Bacillus subtilis* chorismate mutase by genetic selection. *J. Am. Chem. Soc. 118*, 3069-3070.

31. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389-3402.

Chapter 4

Importance to Stability and Catalysis of Hydrophobic Active Site Residues in *Escherichia Coli* Chorismate Mutase

This work was performed in the Mayo laboratory. Jennifer Keeffe contributed to the purification and analysis of proteins.

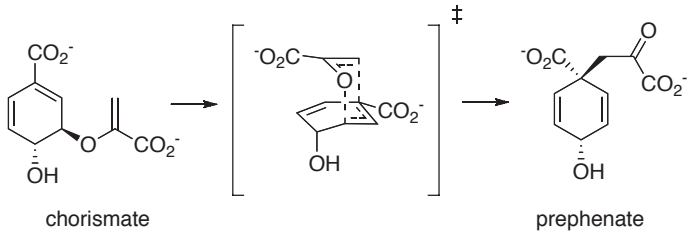
Abstract

Hydrophobic side chains in enzyme active sites may contribute to catalysis through critical interactions that organize the active site and position the reacting molecule. To evaluate the sequence-space landscape for stability and catalysis in hydrophobic active site positions of a natural enzyme, we used site-saturation mutagenesis to place all 20 amino acids in five residues of the *E. coli* monofunctional P-protein chorismate mutase. The ability of each mutant to promote viability in a chorismate mutase deletion strain was assessed, and active variants were purified and evaluated with respect to *in vitro* enzyme activity, far-UV circular dichroism, and thermal denaturation profiles. We found that substitutions in many cases led to decreased activity and increased stability, in accord with previous descriptions of stability-activity trade-offs in active site residues. A few mutations led to increased catalytic efficiency relative to wild type. Altogether, the enzyme tolerated about 34% of the possible single mutations, demonstrating that the selection of hydrophobic residues is important in the design of functional catalysts, yet there can be considerable tolerance to substitution in some active site positions.

Introduction

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

SCHEME 1.



As a first-order intramolecular rearrangement, catalyzed without intermediate steps, covalent catalysis, or modification of the reaction pathway, the reaction has become an important model system for computational studies of enzymatic catalysis (recently reviewed in ref. 6). These same features make the reaction a compelling target for testing computational enzyme design procedures. Because the active site environments of enzymes are likely to present substantial challenges to protein design energy functions that are normally parameterized for global stability,[7,8] we were interested in defining the effects on stability and catalysis of a large number of mutations in secondary active site positions. Active site residues that do not directly hydrogen bond with the reacting molecule can provide information relevant to energy functions because they are subject to the electrostatic and desolvation factors of the active site, yet their mutation does not always lead to immediate loss of activity.

We constructed a series of mutations in the active site of a monofunctional *Escherichia coli* chorismate mutase. The *E. coli* enzymes fall within the AroQ class of

chorismate mutases,[9-12] which are all-helical proteins. The *Bacillus subtilis* enzyme (BsCM) and other AroH mutases, however, adopt a homotrimeric pseudo-α/β-barrel fold.[13] In *E. coli*, chorismate mutases are fused to their downstream biosynthetic partners, forming bifunctional chorismate mutase-prephenate dehydratase (P-protein) and chorismate mutase-prephenate dehydrogenase (T-protein) enzymes.[14] Mechanistic investigations in the bifunctional T-protein enzyme were complicated by a rate-limiting step other than the Claisen rearrangement.[3] However, truncated, monofunctional constructs of the P-protein (Figure 1A, referred to as EcCM herein) have made structural and mechanistic studies possible.[15] The rearrangement catalyzed by EcCM shows expected patterns of kinetic isotope effects and no sensitivity to viscogens, strongly suggesting that the Claisen rearrangment is the rate-determining step.[5,16] In contrast, BsCM appears to be limited partially by a diffusive process and partially by the chemical step.[4,16]

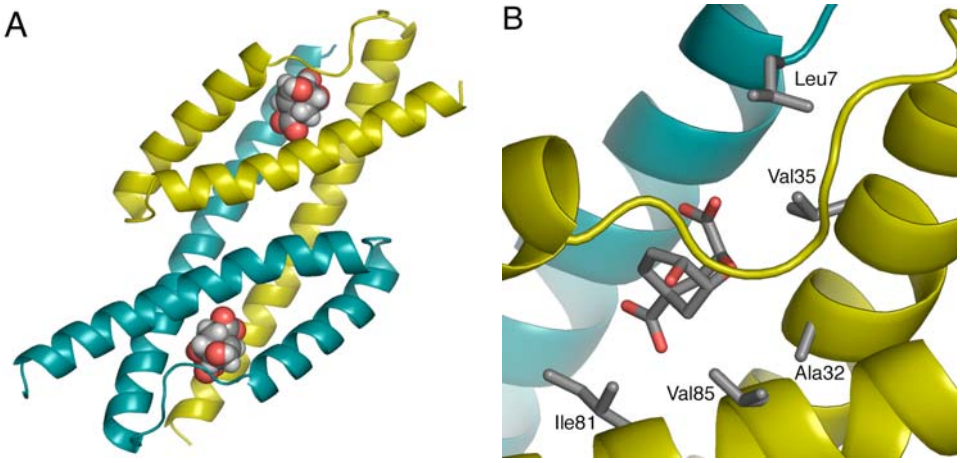


Figure 1. Truncated, monofunctional chorismate mutase construct from *Escherichia coli* chorismate mutase-prephenate dehydratase, from pdb code 1ecm. (A) Overall fold, showing the two chains of the symmetric homodimer in teal and gold. The crystallographic transition state analog is shown in spheres. (B) Active site residues investigated in this study. The crystallographic transition state analog is shown as well.

Previous mutagenesis and selection studies of both EcCM and BsCM enzymes have demonstrated the importance of positively charged active site residues poised to interact with the carboxylates and ether oxygen as well as negatively charged side chains

able to contact the hydroxyl group.[17-22] The present work focuses on five hydrophobic active site positions in EcCM that were previously selected for mutation in a computational design experiment (see chapter 3): Leu7, Ala32, Val35, Ile81, and Val85 (Figure 1B). These side chains do not hydrogen bond with the reacting molecule, but nonetheless may participate in catalysis by positioning the substrate, organizing charged and polar active site residues,[23] or defining the electrostatic environment of the active site[24]. The secondary active site residues investigated in this work may also have special consequences for enzyme stability. Prior mutagenesis studies have found inverse relationships between changes in protein stability and changes in activity following the mutation of active site residues, consistent with a notion that creation of a cavity with specialized electrostatic properties opposes the normal driving forces of protein stability.[25-29]

Here, we have evaluated the effects of all 19 possible single mutations in the five hydrophobic active site positions. Each of the 95 variants was tested for complementation of the auxotrophic *E. coli* KA12 pKIMP-UAUC system of Kast, Hilvert, and co-workers.[20,30] Biologically active variants were purified and their *in vitro* thermal stabilities and steady-state kinetic parameters were assessed.

Materials and Methods

Materials and Strains. *E. coli* KA12[30] and plasmid pKIMP-UAUC[20] were a gift from Donald Hilvert and Peter Kast. Chorismic acid (97.5%) was purchased from Sigma-Aldrich.

Construction and in vivo evaluation of mutants. The truncated, monofunctional N-terminal region of P-protein chrorismate mutase described by Ganem and co-workers[15] was previously isolated and fused to a self-cleavable chitin binding domain-intein fusion construct (IMPACT-CN, New England Biolabs).[31] Site-specific variants were produced in this construct by using single oligonucleotides with degenerate codons in combination with the Quikchange multi-site mutagenesis kit (Stratagene). DNA was prepared directly from plated colonies after transformation of XL10-gold cells (Stratagene) and subsequently used to transform KA12 pKIMP-UAUC. Cells were plated on both selective and non-selective M9c minimal agar plates, prepared as described[32] except that

20 μg/mL L-Phe was included in the selective plates. Non-selective plates contained both L-Phe and L-Tyr. To allow identification of active variants, mutagenesis reactions made use of inactive templates (i.e., stop codons in the position of interest) and colonies that were viable on selective media were selected for sequencing. For identification of inactive variants, reactions were performed using a wild-type template; colonies that were nonviable when picked onto selective media plates were sequenced. To speed identification of the remaining variants, non-degenerate and less-degenerate mutagenesis reactions were performed. All reported constructs were verified by DNA sequencing. Biological activity or inactivity was confirmed by retransformation and plating.

The possibility was investigated that disruption of activity might be caused by deleterious mutations to the promoter region or the N-terminal gene fusion rather than by mutation of the enzyme itself. For some inactive variants, evidence against spurious mutation was provided by multiple independent occurrences of the same mutation with identical behavior. For remaining inactive variants, expression of appropriately sized constructs was verified with small-scale cultures followed SDS-PAGE analysis. Frameshifts or deletions within the N-terminal gene fusion would have been readily apparent.

Protein expression and purification. Proteins were expressed in *E. coli* BL21 DE3, with induction at room temperature using 1 mM IPTG. Cell lysis utilized an Emulsiflex-C5 (Avestin) and was conducted in a buffer containing 20 mM Tris, pH 8, 500 mM NaCl, 1mM EDTA, 10 mM $MgCl_2$, and DNase, RNase, and PMSF. Affinity purification using chitin agarose (New England Biolabs) was performed according to the IMPACT-CN kit protocols. Affinity tag self-cleavage was induced with 50 mM DTT at 25° C. Mass spectrometry and SDS-PAGE previously indicated that this procedure results in the desired EcCM construct.[31] Proteins were further purified by gel filtration using 50 mM Tris pH 7.8, 100 mM NaCl. A well-resolved dimer peak was collected, and variants with significant higher-order aggregation were not further considered. Because the natural CM enzymes are considerably larger in size and do not have the affinity tag, affinity purification followed by gel filtration should have eliminated any significant catalytic contribution of natural enzymes. Protein concentration was determined by

Bradford assay using BSA as a standard for consistency with previously published kinetic studies of the same construct.[21,31,33]

CD spectrometry and thermal denaturation. All variants showed nearly superimposable far-UV circular dichroism wavelength scans at the activity assay temperature of 37° C. Thermostabilities were assessed by thermal denaturation monitored monitored with an Aviv 62DS circular dichroism spectrometer equipped with thermoelectric temperature control. CD was followed at 222 nm over a range of 15-99°C. Temperature steps of 1°C were used with averaging times of 30 sec. and equilibration times of 90 sec. Buffer containing 50 mM Tris pH 7.8, 100 mM NaCl was used. Midpoints of thermal denaturation (T_m) were obtained by fitting data to a two state model[34]. It should be noted that unfolding was not reversible and some samples showed clear evidence of non-two-state behavior, so T_m values should be taken as estimates of relative stabilities rather than thermodynamic parameters. CD data were collected at 22 ± 3 μM protein. Variations in T_m from different concentrations within the range were similar to variations across multiple trials.

Kinetic assays. The absorbance of chorismate was followed at 275 nm (ε = 2630 M^{-1} cm^{-1} (3)) to obtain intital reaction rates. Activity assays were performed at 37 °C in 50 mM Tris pH 7.8, 2.5 mM EDTA, 20 mM β-mercaptoethanol, and 0.01% BSA, conditions used in prior studies of the same construct.[21,31,33] Given pK_as for chorismate and prephenate of < 4.5,[35] these conditions ensure that the reaction of chorismate^{2-} to prephenate^{2-} was monitored. A protein concentration of 20 nM was generally used, however, less-active variants were assayed with 40 nM or 100 nM protein. For cases where k_{cat} and K_M are reported, at least three trials with 9 or more data points each were performed, and substrate concentration ranges of at least 0.3-3 times K_M were used. Kinetic parameters were determined by non-linear fitting of the buffer-corrected initial rates to the Michaelis-Menten equation.

Results and Discussion

Tolerance to mutation. We tested the effect of all possible single mutations to five hydrophobic amino acid residues of the monofunctional *E. coli* P-protein chroismate mutase (Table 1). Each variant was tested for its ability to promote growth in the KA12

Table 1. Results for complementation tests, kinetic parameters, and thermal stabilities[a]

pos 7		k_{cat}	K_M	k_{cat}/K_M	T_m
ALA	-				
ARG	-				
ASN	-				
ASP	-				
CYS	+		> 750	0.6	52.0
GLN	-				
GLU	-				
GLY	-				
HIS	-				
ILE	+	2188 ± 95	298 ± 82	7.3	65.8
LEU	**+**	**2330 ± 310**	**304 ± 52**	**7.7**	**61.4 ± 0.6**
LYS	-				
MET	+		> 750	2.6	60.3
PHE	+	2900 ± 140	350 ± 28	8.3	62.3
PRO	-				
SER	-				
THR	+		> 750	0.8	59.2
TRP	-				
TYR	-				
VAL	+	2140 ± 190	720 ± 110	3.0	61.9

pos 81		k_{cat}	K_M	k_{cat}/K_M	T_m
ALA	-				
ARG	-				
ASN	-				
ASP	-				
CYS	+				*aggr*
GLN	-				
GLU	-				
GLY	-				
HIS	-				
ILE	**+**	**2330 ± 310**	**304 ± 52**	**7.7**	**61.4 ± 0.6**
LEU	+		> 750	3.3	57.7
LYS	-				
MET	+	1610 ± 120	271 ± 23	5.9	64.6
PHE	+		> 750	0.5	59.0
PRO	-				
SER	-				
THR	-				
TRP	+				*aggr*
TYR	-				
VAL	+				*aggr*

pos 32		k_{cat}	K_M	k_{cat}/K_M	T_m
ALA	**+**	**2330 ± 310**	**304 ± 52**	**7.7**	**61.4 ± 0.6**
ARG	-				
ASN	-				
ASP	-				
CYS	+				*ppt*
GLN	-				
GLU	-				
GLY	+				*aggr*
HIS	-				
ILE	+				*ppt*
LEU	-				
LYS	-				
MET	-				
PHE	-				
PRO	-				
SER	+	2710 ± 360	220 ± 29	12.4	61.7
THR	+	3630 ± 240	384 ± 69	9.4	62.8
TRP	-				
TYR	-				
VAL	+				*ppt*

pos 85		k_{cat}	K_M	k_{cat}/K_M	T_m
ALA	+		> 750	0.7	62.7
ARG	+		> 750	2.5	66.9
ASN	+		> 750	0.7	66.8
ASP	-				
CYS	+	1890 ± 240	568 ± 87	3.3	54.7
GLN	-				
GLU	-				
GLY	-				
HIS	-				
ILE	+	3313 ± 64	550 ± 34	6.0	62.9
LEU	+		> 750	2.6	62.2
LYS	+		> 750	0.8	67.1
MET	+		> 750	0.9	65.6
PHE	+		> 750	1.2	61.7
PRO	-				
SER	-				
THR	+		> 750	1.3	62.1
TRP	+				*aggr*
TYR	+		> 750	1.6	65.7
VAL	**+**	**2330 ± 310**	**304 ± 52**	**7.7**	**61.4 ± 0.6**

pos 35		k_{cat}	K_M	k_{cat}/K_M	T_m
ALA	+	1220 ± 120	515 ± 86	2.4	61.2
ARG	-				
ASN	-				
ASP	-				
CYS	+	1115 ± 36	358 ± 37	3.1	61.8
GLN	-				
GLU	-				
GLY	-				
HIS	-				
ILE	+	3900 ± 510	441 ± 82	8.8	66.6
LEU	+		> 750	0.8	66.7
LYS	-				
MET	+		> 750	3.1	68.9
PHE	-				
PRO	-				
SER	-				
THR	+		> 750	1.2	56.5
TRP	-				
TYR	-				
VAL	**+**	**2330 ± 310**	**304 ± 52**	**7.7**	**61.4 ± 0.6**

[a] k_{cat} in μM/min; K_M in μM; k_{cat}/K_M in min^{-1} μM^{-1}. T_m in °C. Notes: *aggr/ppt,* protein aggregated or precipitated during purification. Where uncertainties are indicated, averages and standard deviations are reported from at least three trials. Where K_M was larger than 750 $min^{-1}\mu M^{-1}$, k_{cat}/K_M represents an estimate from a single trial.

pKIMP-UAUC chorismate mutase deletion system,[20,30] which requires chorismate mutase activity for viability in media lacking tyrosine and phenylalanine. Of the 95 possible single mutants, 35 (or 37%) were scored as active. Although it is difficult to compare results across systems given likely variations in cellular requirements, protein expression levels, and various other factors, this value is consistent with other extensive substitution experiments. A sampling from among the many exhaustive and random mutagenesis studies suggests that between 50% and 85% of single mutations typically maintain some level of function.[36-41] In catalytic sites, an average of 20%-60% of single mutations preserve some function.[41-44] Thus, the hydrophobic residues studied here have an average tolerance to substitution typical of active site residues in other proteins.

EcCM variants that allowed growth on selective media were purified and their *in vitro* stabilities and kinetic parameters were evaluated. Measured T_m values for biologically active variants ranged approximately ±10 °C relative to wild type, yet the average T_m (61.7° C) was similar to that of the wild type (61.4° C). *In vitro* activities were reduced relative to wild type, except in a few cases. The greatest increase in activity relative to wild type was found with the Ala32Ser mutation, previously reported as the result of a computational design approach.[31] This mutation results in an approximately 60% increase in catalytic efficiency with minimal change in stability.

Position 7. Leu7 lies near the enolpyruvyl moiety carboxylate and also the catalytically essential residue Arg11. Position 7 tolerated six relatively conservative side chain substitutions, with mutations to Ile and Phe offering slight increases in stability and Phe leading to slightly higher activity.

Position 32. Ala32 lies closest to the equivalent of the forming C-C bond and the cyclohexadienyl face carboxylate in the transition state analog of the 1ecm structure. Although six alternate side chains were tolerated in position 32, only the Ala32Ser and Ala32Thr variants could be readily purified without precipitation or aggregation. These substitutions increased catalytic efficiency slightly.

Position 35. In the 1ecm structure, Val35 is within van der Waals contact distance of the catalytically important residues Lys39 and Gln88, both of which contact the ether oxygen of the breaking bond. Val 35 also is expected to pack against the face of the enolpyruvyl moiety in the reacting molecule. Six mutations were tolerated in this

position. The Val35Met, Val35Leu, and Val35Ile mutations increased T_m by 7.5° C, 5.3° C, and 5.2° C, respectively. All mutations decreased catalytic efficiency except for Val35Ile, and all mutations increased K_M.

Position 81. In the 1ecm crystal structure, Ile81 packs closely against the cyclohexadienyl face carboxylate of the transition state analog. Six substitutions of this residue were tolerated, although the cysteine, tryptophan, and valine mutants were prone to aggregation *in vitro*. Aside from a slight increase in stability with the Ile81Met mutation, the substitutions for which data could be obtained led to reduced stability and catalytic efficiency.

Position 85. Based on the 1ecm structure, Val 85 lies adjacent to Ile81 and close to the cyclohexadienyl face of the reacting. The Cα-Cβ vector of Val85 is directed more towards solvent than Ile81; consistent with this orientation, the residue was broadly tolerant of substitution, allowing 12 of 19 possible mutations. Interestingly, 9 of the 11 mutants for which data were recorded showed increased T_m relative to wild type, while all of the mutants had increased K_M.

Comparison to AroQ family members. Table 2 shows structural equivalents from other solved AroQ mutase structures of each of the five residues investigated. A structural alignment was performed for each structure and sequence variations are reported from a BLAST[45] search of the RefSeq database[46] using the sequence from that structure as a query. A number of the observed sequence changes from other organisms were found to retain wild-type-like or have increased activity in the EcCM scaffold: Leu7Ile, Ala32Ser, Ala32Thr, and Val35Ile. Four mutations were observed in sequence alignments but led to precipitated or aggregated protein *in vitro*: Ala32Val, Ala32Gly, Ala32Cys, and Ile81Val. The majority of variations from sequence alignments led to reduction in activity or loss of biologically sufficient activity, however.

Table 2. Active site alignment of solved AroQ mutase structures and BLAST results[a]

***E. coli;* 1ecm**	**Leu 7**	**Ala 32** (STV)	**Val 35** (I)	**Ile 81** (L)	**Val 85**
***S. cerevisiae;* 4csm**	Leu 12	Gly 161	Val 164	Ile 239 (M)	Lys 243
***M. tuberculosis;* 2fp2**	Leu 130	Ala 53	Val 56 (IM)	Ile 102	Glu 106 (KV)
***P. furiosus;* 1ybz**	Leu 4	Ala 29 (CS)	Ile 32 (M)	Leu 67 (V)	Lys 71 (R)
***T. thermophilus;* 2d8d**	Ile 5 (LM)	Val 30 (A)	Ile 33 (V)	Phe 78	Leu 82 (T)

[a]Proteins were aligned over the transition state analog or over all residues. For each pdb protein sequence, protein-protein BLAST[44] was performed using the RefSeq database[45]. Variations of the position are reported in parentheses for matches with bitscores of 85 or greater (60 for 1ybz); these scores corresponded to sequences with ~40% sequence identity or greater.

The sequence alignments reflected the tolerance to mutation seen in position 85. Related organisms showed identities including Glu, Lys, and Arg, in accord with the more solvent-directed orientation of this position. A previous selection study reported that the equivalent of position 85 in *Methanococcus jannaschii* CM was highly conserved.[47] However, that observation may reflect interactions not present in the *E. coli* structure. The equivalent position in the *M. jannaschii* enzyme is expected to be occupied by a lysine that forms a salt bridge with a residue from a neighboring helix in crystal structures.[11] Curiously, the equivalent of position 85 in the *Mycobacterium tuberculosis* enzyme makes a catalytically important contact to the hydroxyl group of the transition state analog that EcCM makes with Glu52.[11]

Tradeoff between stability and activity. It has been proposed that evolutionary optimization of active site residues for catalysis would result in the selection of residues that are not optimal for protein stability.[25-29] Enzymes must devote stabilizing interactions to the reacting molecule that are left both unsatisfied and likely destabilizing in the unbound form of the protein. A laboratory-evolved chorismate mutase of Kast, Hilvert, and co-workers dramatically illustrated this principle by transforming from an apparent molten globule into a more ordered protein in the presence of a transition state analog.[48,49]

The results presented here are consistent with the idea of a stability-activity tradeoff in enzyme active sites. Of the biologically active variants, sixteen showed increased stability, while eleven showed decreased stability. In position 85 of EcCM, all of eleven biologically active mutants showed reduced catalytic efficiency and increased K_M and nearly all mutants showed increased thermostability. Val85 participates in van der Waals interactions with both the transition state analog and residues from neighboring helices in the 1ecm structure. Thus, stability could potentially be increased and chorismate affinity could be reduced through substitutions that partially fill the binding pocket, through helix-helix interactions that affect the geometry of the active site, or through a combination of the factors. Movement of the neighboring helix has been propsed to facilitate substrate entry[11] and helix-helix interaction could affect substrate affinity or cause substrate entry to impact the rate-limiting step. In position 35 as well, all biologically active mutations had increased K_M and a few showed large increases in

thermostability. In the 1ecm structure, Val35 appears to participate in van der Waals interactions with the transition state analog, catalytically important residues Lys39 and Gln88, and neighboring helix residues. While the present study cannot be expected to isolate the relative contributions of these different interactions, it clearly indicates that this residue contributes to the catalytic process and the global stability of the protein.

Relationship to models of EcCM catalytic mechanism. Among theorists, differing computational treatments of chorismate mutase systems have led to varied descriptions of the origin of enzymatic catalysis of the rearrangement. Despite an observation that the chemical step is not fully rate limiting in the enzyme,[16] the majority of work has modeled the BsCM active site.[50-68] The few computational investigations in the monofunctional EcCM construct have proposed, in opposition to most of the other studies, that both enzymes accelerate the rearrangement exclusively by restricting the conformational space of the substrate, not by stabilizing the transition state.[59,69-71] While these studies, and an earlier one,[72] explicitly identified Val35 and Ile81 as mediators of the substrate restriction, we do not expect the mutations identified here to differentiate between possible modes of EcCM catalysis.

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

Implications for enzyme design. The present mutational study was conducted to clarify the results of a previous computational design experiment.[31] Computational design procedures[73] were used to model an *ab initio* calculated transition state structure for the

rearrangement within the EcCM active site. Eighteen residues were designed, and the energy function selected for mutation only the five hydrophobic positions of the present study and Asp48. The replacement of Asp48 led to a loss of catalytic activity; this position is being further investigated with the aim of improving design models.

The results presented here suggest that current computational design methods are able to reasonably model hydrophobic side chains in enzyme active sites. The mutations selected in the previous study, Leu7Ile, Ala32Ser, Val35Ile, Ile81Leu, and Val85Ile, were all scored as biologically active and most performed well in terms of *in vitro* activity and stability. The availability of a large dataset for single mutations in these positions will enable evaluation and tuning of enzyme design energy functions in more detail.

Finally, the ability of EcCM to tolerate a number of mutations without reduction in stability or activity is in accord with previous descriptions of active site plasticity. Secondary active site residues that accept side chain substitutions can allow evolution or design of enzymes with new substrate specificities and chemical mechanisms, sometimes without disruption of the original function.[74-77] These secondary positions offer rich opportunities for testing and refining enzyme design methodologies through active site redesign experiments.

Acknowledgements

We thank Dr. Donald Hilvert and Dr. Peter Kast for generously sharing *E. coli* KA12 and plasmid pKIMP-UAUC. This work was supported by the Howard Hughes Medical Institute, the Defense Advanced Research Projects Agency, and the Institute for Collaborative Biotechnologies (ARO).

References

1. Sogo, S. G., Widlanski, T. S., Hoare, J. H., Grimshaw, C. E., Berchtold, G. A., and Knowles, J. R. (1984) Stereochemistry of the rearrangement of chorismate to prephenate: Chorismate mutase involves a chair transition state, *J. Am. Chem. Soc. 106*, 2701-2703.

2. Copley, S. D., and Knowles, J. R. (1985) The uncatalyzed Claisen rearrangement of chorismate to prephenate prefers a transition state of chairlike geometry, *J. Am. Chem. Soc. 107*, 5306-5308.

3. Addadi, L., Jaffe, E. K., and Knowles, J. R. (1983) Secondary tritium isotope effects as probes of the enzymic and nonenzymic conversion of chorismate to prephenate, *Biochemistry 22*, 4494-4501.

4. Gustin, D. J., Mattei, P., Kast, P., Wiest, O., Lee, L., Cleland, W. W., and Hilvert, D. (1999) Heavy atom isotope effects reveal a highly polarized transition state for chorismate mutase, *J. Am. Chem. Soc. 121*, 1756-1757.

5. Wright, S. K., DeClue, M. S., Mandal, A., Lee, L., Wiest, O., Cleland, W. W., and Hilvert, D. (2005) Isotope effects on the enzymatic and nonenzymatic reactions of chorismate, *J. Am. Chem. Soc. 127*, 12957-12964.

6. Gao, J., Ma, S., Major, D. T., Nam, K., Pu, J., and Truhlar, D. G. (2006) Mechanisms and free energies of enzymatic reactions, *Chem. Rev. 106*, 3188-3209.

7. Mendes, J., Guerois, R., and Serrano, L. (2002) Energy estimation in protein design, *Curr. Opin. Struct. Biol. 12*, 441-446.

8. Vizcarra, C. L., and Mayo, S. L. (2005) Electrostatics in computational protein design, *Curr. Opin. Chem. Biol. 9*, 622-626.

9. Lee, A. Y., Karplus, P. A., Ganem, B., and Clardy, J. (1995) Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase, *J. Am. Chem. Soc. 117*, 3627-3628.

10. Xue, Y., Lipscomb, W. N., Graf, R., Schnappauf, G., and Braus, G. (1994) The crystal structure of allosteric chorismate mutase at 2.2-Å resolution, *Proc. Natl. Acad. Sci USA 91*, 10814-10818.

11. Övkist, M., Dey, R., Sasso, S., Grahn, E., Kast, P, and Krengel, U. (2006) 1.6 Å crystal structure of the secreted chorismate mutase from *Mycobacterium tuberculosis*: Novel fold topology revealed, *J. Mol. Biol. 357*, 1483-1499.

12. Qamra, R., Prakash, P., Aruna, B., Hasnain, S. E., and Mande, S. C. (2006) The 2.15 Å crystal structure of *Mycobacterium tuberculosis* chorismate mutase reveals an

unexpected gene duplication and suggests a role in host-pathogen interactions, *Biochemistry 45*, 6997-7005.

13. Chook, Y. M., Ke, H., and Lipscomb, W. N. (1993) Crystal structures of the monofunctional chorismate mutase from *Bacillus subtilis* and its complex with a transition state analog, *Proc. Natl. Acad. Sci. USA 90*, 8600-8603.

14. Cotton, G. H., and Gibson, F. (1965) The biosynthesis of phenylalanine and tyrosine: Enzymes converting chorismic acid into prephenic acid and their relationships to prephenate dehydratase and prephenate dehydrogenase, *Biochim. Biophys. Acta 100*, 76-88.

15. Stewart, J., Wilson, D. B., and Ganem, B. (1990) A genetically engineered monofunctional chorismate mutase, *J. Am. Chem. Soc. 112*, 4582-4584.

16. Mattei, P., Kast, P., and Hilvert, D. (1999) *Bacillus subtilis* chorismate mutase is partially diffusion-controlled, *Eur. J. Biochem. 261*, 25-32.

17. Cload, S. T., Liu, D. R., Pastor, R. M., and Schultz, P. G. (1996) Mutagenesis study of active site residues in chorismate mutase from *Bacillus subtilis*, *J. Am. Chem. Soc. 118*, 1787-1788.

18. Liu, D. R., Cload, S. T., Pastor, R. M., and Schultz, P. G. (1996) Analysis of active site residues in *Escherichia coli* chorismate mutase by site-directed mutagenesis, *J. Am. Chem. Soc. 118*, 1789-1790.

19. Kast, P., Hartgerink, J. D., Asif-Ullah, M., and Hilvert, D. (1996) Electrostatic catalysis of the Claisen rearrangment: Probing the role of Glu78 in *Bacillus Subtilis* chorismate mutase by genetic selection, *J. Am. Chem. Soc. 118*, 3069-3070.

20. Kast, P., Asif-Ullah, M., Jiang, N., and Hilvert, D. (1996) Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: The importance of electrostatic catalysis, *Proc. Natl. Acad. Sci. USA 93*, 5043-5048.

21. Zhang, S., Kongsaeree, P., Clardy, J., Wilson, D. B., and Ganem, B. (1996) Site-directed mutagenesis of monofunctional chorismate mutase engineered from the *E. coli* P-protein, *Bioorg. Med. Chem. 4*, 1015-1020.

22. Kienhöfer, A., Kast, P., and Hilvert, D. (2003) Selective stabilization of the chorismate mutase transition state by a positively charged hydrogen bond donor, *J. Am. Chem. Soc. 125*, 3206-3207.

23. Warshel, A. (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites, *J. Biol. Chem. 273*, 27035-27038.

24. Shan, SO, and Herschlag, D. (1996) The change in hydrogen bond strength accompanying charge rearrangement: Implications for enzymatic catalysis, *Proc. Natl. Acad. Sci. USA 93*, 14474-14479.

25. Zhi, W., Srere, P. A., and Evans, C. T. (1991) Conformational stability of pig citrate synthase and some active-site mutants, *Biochemistry 30*, 9281-9286.

26. Meiering, E. M., Serrano, L., and Fersht, A. R. (1992) Effect of active site residues in barnase on activity and stability, *J. Mol. Biol. 225*, 585-589.

27. Shoichet, B. K., Baase, W. A., Kuroki, R., and Matthews, B. W. (1995) A relationship between protein stability and protein function, *Proc. Natl. Acad. Sci. USA 92*, 452-456.

28. Beadle, B. M., and Shoichet, B. K. (2002) Structural bases of stability-function tradeoffs in enzymes, *J. Mol. Biol. 321*, 285-296.

29. Mukaiyama, A., Haruki, M., Ota, M., Koga, Y., Takano, K., and Kanaya, S. (2006) A hyperthermophilic protein acquires function at the cost of stability, *Biochemistry 45*, 12673-12679.

30. Kast, P., Asif-Ullah, M., and Hilvert, D. (1996) Is chorismate mutase a prototypic entropy trap? – Activation parameters for the *Bacillus subtilis* enzyme, *Tet. Lett. 37*, 2691-2694.

31. Lassila, J. K., Keeffe, J. R., Oelschlaeger, P., and Mayo, S. L. (2005) Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity, *Prot. Eng. Des. Sel. 18*, 161-163.

32. Gamper, M., Hilvert, D., and Kast, P. (2000) Probing the role of the C-terminus of *Bacillus subtilis* chorismate mutase by a novel random protein-termination strategy, *Biochemistry 39*, 14087-14094.

33. Zhang, S., Wilson, D. B., and Ganem, B. (2003) An engineered chorismate mutase with allosteric regulation, *Bioorg. Med. Chem. 11*, 3109-3114.

34. Becktel, W. J., and Schellman, J. A. (1987) Protein stability curves, *Biopolymers 26*, 1859-1877.

35. Kast, P., Tewari, Y. B., Wiest, O., Hilvert, D., Houk, K. N., and Goldberg, R. N. (1997) Thermodynamics of the conversion of chorismate to prephenate: Experimental results and theoretical predictions, *J. Phys. Chem. B 101*, 10976-10982.

36. Shortle, D., and Lin, B. (1985) Genetic analysis of staphylococcal nuclease: Identification of three intragenic “global” supressors of nuclease-minus mutations, *Genetics 110*, 539-555.

37. Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E., and Hutchison, C. A. (1989) Complete mutagenesis of the HIV-1 protease, *Nature 340*, 397-400.

38. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., and Sauer, R. T. (1990) Deciphering the message in protein sequences: Tolerance to amino acid substitutions, *Science 247*, 1306-1310.

39. Rennell, D., Bouvier, S. E., Hardy, L. W., and Poteete, A. R. (1991) Systematic mutation of bacteriophage T4 lysozyme, *J. Mol. Biol. 222*, 67-87.

40. Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., and Miller, J. H. (1994) Genetic studies of the lac repressor XIV: Analysis of 4000 altered *Escherichia coli* lac repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence, *J. Mol. Biol. 240*, 421-433.

41. Guo, H. H., Choe, J., and Loeb, L. A. (2004) Protein tolerance to random amino acid change, *Proc. Natl. Acad. Sci. USA 101*, 9205-9210.

42. Liu, L., and Santi, D. V. (1993) Asparagine 229 in thymidylate synthase contributes to, but is not essential for, catalysis, *Proc. Natl. Acad. Sci. USA 90*, 8604-8608.

43. Warren, M. S., Marolewski, A. E., and Benkovic, S. J. A rapid screen of active site mutants in glycinamide ribonucleotide transformylase, *Biochemistry 35*, 8855-8862.

44. Axe, D. D., Foster, N. W., and Fersht, A. R. (1998) A search for single substitutions that eliminate enzymatic function in a bacterial ribonuclease, *Biochemistry 37*, 7157-7166.

45. Altschul, S. F., Madden, T. F., Schaffer, A, A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389-3402.

46. Pruitt K. D., Tatusova, T., Maglott D. R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, *Nucleic Acids Res. 33*, D501-D504.

47. Taylor, S. V., Walter, K. U., Kast, P., and Hilvert, D. (2001) Searching sequence space for protein catalysts, *Proc. Natl. Acad. Sci. USA 98*, 10596-10601.

48. MacBeath, G., Kast, P., and Hilvert, D. (1998) Redesigning enzyme topology by directed evolution, *Science 279*, 1958-1961.

49. Vamvaca, K., Vogeli, B., Kast, P., Pervushin, K., and Hilvert, D. (2004) An enzymatic molten globule: Efficient coupling of folding and catalysis, *Proc. Natl. Acad. Sci. USA 101*, 12860-12864.

50. Lyne, P. D., Mulholland, A. J., and Richards, W. G. (1995) Insights into chorismate mutase catalysis from a combined QM/MM simulation of the enzyme reaction, *J. Am. Chem. Soc. 117*, 11345-11350.

51. Davidson, M. M., Gould, I. R., And Hillier, I. H. (1996) The mechanism of the catalysis of the Claisen rearrangement of chorismate to prephenate by the chorismate mutase from *Bacillus subtilis*. A molecular mechanics and hybrid quantum mechanical/molecular mechanical study, *J. Chem. Soc., Perkin Trans. 2*, 525-532.

52. Hall, R. J., Hindle, S. A., Burton, N. A., and Hillier, I. H. (2000) Aspects of hybrid QM/MM calculations: The treatment of the QM/MM interface regions and geometry optimization with an application to chorismate mutase, *J. Comp. Chem. 21*, 1433-1441.

53. Worthington, S. E., Roitberg, A. E., and Krauss, M. (2001) An MD/QM study of the chorismate mutase-catayzed Claisen rearrangement reaction, *J. Phys. Chem. B 105,* 7087-7095.

54. Guo, H., Cui, Q., Lipscomb, W. N., and Karplus, M. (2001) Substrate conformational transitions in the active site of chorismate mutase: Their role in the catalytic mechanism, *Proc. Natl. Acad. Sci. USA 98*, 9032-9037.

55. Marti, S., Andres, J., Moliner, V., Silla, E., Tunon, I., Bertran, J., and Field, M. J. (2001) A hybrid potential reaction path and free energy study of the chorismate mutase reaction, *J. Am. Chem. Soc. 123*, 1709-1712.

56. Lee, Y. S., Worthington, S. E., Krauss, M., and Brooks, B. R. (2002) Reaction mechanism of chorismate mutase studied by the combined potentials of quantum mechanics and molecular mechanics, *J. Phys. Chem. B 106*, 12059-12065.

57. Guimaraes, C. R. W., Repasky, M. P., Chandrasekhar, J., Tirado-Rives, J., and Jorgensen, W. L. (2003) Contributions of conformational compression and preferential transition state stabilization to the rate enhancement by chorismate mutase, *J. Am. Chem. Soc. 125*, 6892-6899.

58. Strajbl, M., Shurki, A., Kato, M., and Warshel, A. (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization, *J. Am. Chem. Soc. 125*, 10228-10237.

59. Hur, S., and Bruice, T. C. (2003) The near attack conformation approach to the study of the chorismate to prephenate reaction, *Proc. Natl. Acad. Sci. USA 100*, 12015-12020.

60. Crespo, A., Scherlis, D. A., Marti, M. A., Ordejon, P., Roitberg, A. E., and Estrin, D. A. (2003) A DFT-based QM-MM approach designed for the treatment of large molecular systems: Application to chorismate mutase *J. Phys. Chem. B 107*, 13728-13736.

61. Guo, H., Cui, Q., Lipscomb, W. N., and Karplus, M. (2003) Understanding the role of active-site residues in chorismate mutase catalysis from molecular-dynamics simulations, *Angew. Chem. Int. Ed. 42*, 1508-1511.

62. Ranaghan, K. E., and Mulholland, A. J. (2004) Conformational effects in enzyme catalysis: QM/MM free energy calculations of the 'NAC' contribution in chorismate mutase, *Chem. Commun.*, 1238-1239.

63. Ranaghan, K. E., Ridder, L., Szefczyk, B., Sokalski, W. A., Hermann, J. C., and Mulholland, A. J. (2004) Transition state stabilization and substrate strain in enzyme catalysis: *Ab initio* QM/MM modeling of the chorismate mutase reaction, *Org. Biomol. Chem. 2*, 968-980.

64. Szefczyk, B., Mulholland, A. J., Ranaghan, K. E., and Sokalski, W. A. (2004) Differential transition-state stabilization in enzyme catalysis: Quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field, *J. Am. Chem. Soc. 126*, 16148-16159.

65. Guimaraes, C. R. W., Udier-Blagovic, M., Tubert-Brohman, I., and Jorgensen, W. L. (2005) Effects of Arg90 neutralization on the enzyme-catalyzed rearrangement of chorismate to prephenate, *J. Chem. Theory Comput. 1*, 617-625.

66. Claeyssens, F., Ranaghan, K. E., Manby, F. R., Harvey, J. N., and Mulholland, A. J. (2005) Multiple high-level QM/MM reaction paths demonstrate transition-state stabilization in chorismate mutase: Correlation of barrier height with transition-state stabilization, *Chem. Commun.*, 5068-5070.

67. Crespo, A., Marti, M. A., Estrin, D. A., and Roitberg, A. E. (2005) Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase, *J. Am. Chem. Soc. 127*, 6940-6941.

68. Ishida, T., Fedorov, D. G., and Kitaura, K. (2006) All electron quantum chemical calculation of the entire enzyme system confirms a collective catalytic device in the chorismate mutase reaction, *J. Phys. Chem B 110*, 1457-1463.

69. Hur, S., and Bruice, T. C. (2002) The mechanism of catalysis of the chorismate to prephenate reaction by the *Escherichia coli* mutase enzyme, *Proc. Natl. Acad. Sci. USA 99*, 1176-1181.

70. Hur, S., and Bruice, T. C. (2003) Comparison of formation of reactive conformers (NACs) for the Claisen rearrangment of chorismate to prephenate in water and in the *E. coli* mutase: The efficiency of the enzyme catalysis, *J. Am. Chem. Soc. 125*, 5964-5972.

71. Zhang, X., Zhang, X., and Bruice, T. C. (2005) A definitive mechanism for chorismate mutase, *Biochemistry 44*, 10444-10448.

72. Khanjin, N. A., Snyder, J. P., and Menger, F. M. (1999) Mechanism of chorismate mutase: Contribution of conformational restriction to catalysis in the Claisen rearrangement, *J. Am. Chem. Soc. 121*, 11831-11846.

73. Lassila, J. K., Privett, H. K., Allen, B. D., and Mayo, S. L. (2006) Combinatorial methods for small-molecule placement in computational enzyme design, *Proc. Natl. Acad. Sci. USA 103*, 16710-16715.

74. Bone, R., Silen, J. L., and Agard, D. A. (1989) Structural plasticity broadens the specificity of an engineered protease, *Nature 339*, 191-195.

75. Gerlt, J. A., Babbitt, P. C., and Rayment, I. (2005) Divergent evolution in the enolase superfamily: The interplay of mechanism and specificity, *Arch. Biochem. Biophys. 433*, 59-70.

76. Aharoni, A., Gaidukiov, L., Kheronsky, O., Gould, S. M., Roodveldt, C., and Tawfik, D. S. (2005) The 'evolvability' of promiscuous protein functions, *Nature Genet. 37*, 73-76.

77. Yoshikuni, Y., Ferrin, T. E., and Keasling, J. D. (2006) Designed divergent evolution of enzyme function, *Nature 440*, 1078-1082.

Chapter 5

Importance of a Helix N-Capping Residue for Stability and Catalytic Activity of *Escherichia Coli* Chorismate Mutase

This work was performed in the Mayo laboratory. Jennifer Keeffe contributed to the purification and analysis of proteins.

Abstract

A helix N-capping residue in the monofunctional *E. coli* chorismate mutase enzyme has been investigated to evaluate the importance of this residue to catalytic activity. This residue, Asp48 in the wild-type enzyme, simultaneously interacts with the helix N-terminus and an arginine side chain that is critical for catalytic activity. We investigated the effect of all 19 mutations in this position on *in vivo* activity. Four variants were found to have biological activity. *In vitro* kinetics and thermal stabilities were measured for these four variants. All showed reduced activity and thermostability. Because computational design tests suggest that modeling in this position is not reflective of the observed requirements for catalytic activity, we are further investigating the role of this residue in catalysis and substrate binding.

The N- and C-termini of α-helices are flanked by residues with distinct preferences for side chains that can satisfy unpaired helical hydrogen bond donors and acceptors.[1-3] These helix-capping interactions are reflected in protein database surveys, protein stability measurements, and model peptide studies (reviewed in ref. 4). At N-termini, N-cap residues can affect helix stability by as much as 4 kcal/mol[5] and are found to be important considerations in protein design.[6]

We report an investigation of an N-capping residue in the active site of an important model enzyme and find that the residue is essential for efficient catalytic efficiency.

Chorismate mutases catalyze the rearrangement of chorismate to prephenate (Fig. 1), a key step in the biosynthesis of aromatic products including tyrosine and phenylalanine. The reaction is a Claisen rearrangement, a concerted [3,3]-sigmatropic process, and proceeds through a chairlike transition state in enzyme and solution.[7-11] Chorismate mutases from *Escherichia coli* are fused to prephenate dehydratase or prephenate dehydrogenase to form bifunctional enzymes,[12] but truncated, monofunctional mutase constructs have been used for mechanistic and structural studies.[11,13,14]

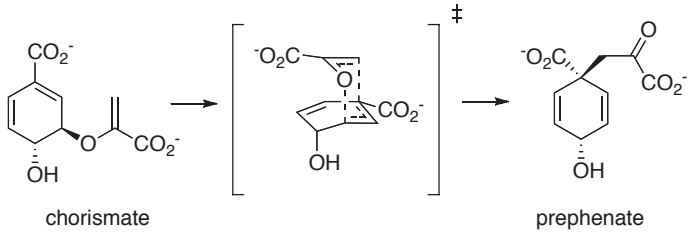


Figure 1. The Claisen rearrangement of chorismate to prephenate.

The chorismate mutase domain of *E. coli* chorismate mutase-prephenate dehydratase (EcCM) is a homodimeric helix bundle with residues from both chains forming each active site (Fig. 2A).[14] The N-capping residue Asp48 hydrogen bonds with both the helix N-terminus and Arg11, a critical residue that contacts the transition state analog in crystal structures (Fig. 2B). Mutational studies have shown the importance of Arg11 to catalysis; the Arg11Lys and Arg11Ala mutants lead to 10^3-fold and 10^4-fold reductions in catalytic efficiency, respectively.[15]

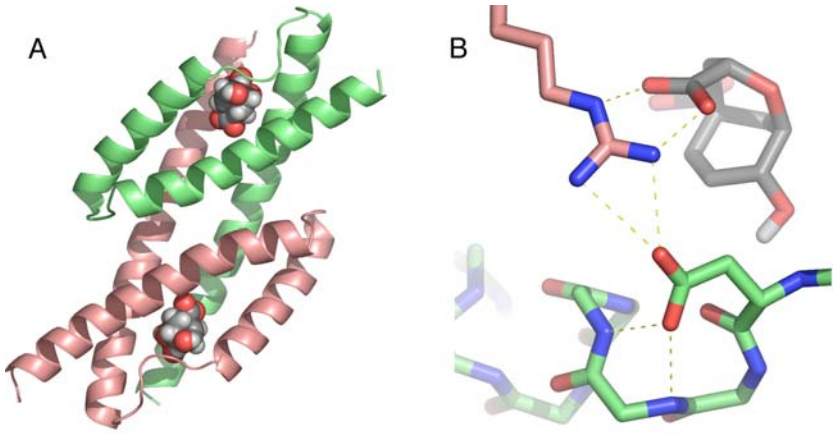


Figure 2. The N-terminal 109 residues of *E. coli* chorismate mutase-prephenate dehydratase, from pdb code 1ecm (*A*) The two chains in the homodimer are colored differently and the crystallographic transition state analog is shown with spheres. (*B*) Interactions of Asp48 with helix N-terminus and Arg11, shown contacting the transition state analog. Side chains, protons, and other elements are neglected for clarity.

We tested the effect of all 20 amino acids in position 48. Each variant was assessed for its ability to complement the *E. coli* chorismate mutase deletion strain of Kast, Hilvert, and co-workers,[16,17] which requires chorismate mutase activity for growth on media lacking tyrosine and phenylalanine. Variants showing *in vivo* mutase activity were expressed and purified, and their *in vitro* kinetic parameters and thermal stabilities were evaluated. Table 1 shows the results of complementation tests, kinetics, and thermal denaturation.

All but four position 48 mutants (Leu, Asn, Gln, and Cys) lost complementation capability, and these four variants showed significant reduction in both T_m and k_{cat}/K_M. All four mutants showed significant increases in K_M such that k_{cat} and K_M could not be determined accurately using the assay procedure. While the T_m values reported in this study should be viewed as approximate measures of stability because thermal denaturations were not reversible, the four T_m values were reduced 4-8 °C relative to wild type.

Table 1. Results from *in vivo* complementation, thermal denaturation, and kinetic assays[a]

pos 48		k_{cat}	K_M	k_{cat}/K_M	T_m
ALA	-				
ARG	-				
ASN	+		> 750	0.5	56.5
ASP	**+**	**2330 ± 310**	**304 ± 52**	**7.8**	**61.9**
CYS	+		> 750	1.1	53.8
GLN	+		> 750	0.6	55.5
GLU	-				
GLY	-				
HIS	-				
ILE	-				
LEU	+		> 750	0.4	57.7
LYS	-				
MET	-				
PHE	-				
PRO	-				
SER	-				
THR	-				
TRP	-				
TYR	-				
VAL	-				

[a] k_{cat} in μM/min; K_M in μM; k_{cat}/K_M in min^{-1} μM^{-1}. T_m in °C.

The fact that all position 48 mutants lost stability coincides with previous studies placing Asp as one of the most favored N-capping residues.[3,18,19] However, the ordering of stabilities of the five biologically active variants, Asp > Leu > Asn > Gln > Cys, differs somewhat from the observed trends, which would rank Leu lower in stability than Asn or than Cys, depending on the study.

The interaction of Asp48 with Arg11 in the EcCM crystal structure (Figure 2B) suggests that the Asp48 side chain occupies an ideal geometry for simultaneous stabilization of the helix and the essential catalytic residue. The simplest explanation for the loss of activity is that the salt bridge is required to stabilize the arginine in a geometry that provides for optimal interaction with the reacting molecule. However, more complex processes could be involved. The Asp48-Arg11 salt bridge spans the interface between two helices proposed to offer the site of entry of substrate into the binding pocket. Structural studies on related enzymes[20] suggest that the two helices may be able to move relative to each other, opening an entry pathway. If this model is correct, it suggests that the Asp48-Arg11 interaction could impact the process of substrate entry into the binding pocket.

As discussed further in chapter 6, tests of computational enzyme design methods suggest that the effects of position 48 on catalytic activity are not being modeled accurately. Thus, we are further investigating the role of this position with the hope of determining how the interactions of position 48 affect substrate binding and catalytic activity.

Materials and Methods

Mutagenesis and in vivo *evaluation of mutants*

The truncated monofunctional *E. coli* chorismate mutase[13] was previously inserted into a self-cleavable affinity expression system from the IMPACT-CN kit (New England Biolabs).[21] All amino acid mutations in position 48 were constructed using the Quikchange multi-site mutagenesis kit (Stratagene). Both degenerate codon and non-degenrate oligos were used to speed the process of identifying mutants. Each mutant was tested for viability in chorismate mutase deletion system KA12 pKIMP-UAUC[16,17] using M9c minimal agar plates as described[22] except that 20 μg/mL L-Phe was included in the selective plates (non-selective plates contained both L-Phe and L-Tyr). Small-scale expression and SDS-PAGE for all variants verified that appropriately sized constructs accumulated following induction with IPTG.

Protein expression and purification

Expression and purification followed previously described procedures[21]. After overexpression in *E. coli* BL21 DE3, affinity purification and self-cleavage according to the IMPACT-CN protocol was followed by gel filtration in 50 mM Tris pH 7.8, 100 mM NaCl. Protein concentration was determined by Bradford assay using BSA as a standard for consistency with previous experiments.[21,23,24]

Thermal denaturation

Denaturation was monitored with an Aviv 62DS circular dichroism spectrometer equipped with thermoelectric temperature control. CD was followed at 222 nm over a range of 15-99°C. Temperature steps of 1°C were used with averaging times of 30 sec.

and equilibration times of 90 sec. Samples contained 20 μM protein in 50 mM Tris pH 7.8, 100 mM NaCl. Midpoints of thermal denaturation (T_m) were obtained by fitting data to a two-state model.[25] T_m values should be construed as approximate measures of thermostability, as the denaturation was not a reversible process.

Kinetic assays

The absorbance of chorismate was followed at 275 nm (ε = 2630 M^{-1} cm^{-1}).[7] Assays were performed at 37°C. Protein concentrations of 20 nM, 40 nM, or 100 nM were used in 50 mM Tris pH 7.8, 2.5 mM EDTA, 20 mM β-mercaptoethanol, and 0.01% BSA. Where k_{cat} and K_M are reported, substrate ranges of at least 0.3-3 times K_M could be tested, and three trials were performed with 9 or more points each. Kinetic parameters were determined by non-linear fitting to the Michaelis-Menten equation.

Acknowledgements

We thank Dr. Peter Kast and Dr. Donald Hilvert for generously sharing *E. coli* strain KA12 and plasmid pKIMP-UAUC. This work was supported by the Howard Hughes Medical Institute, the Defense Advanced Research Projects Agency, and the Institute for Collaborative Biotechnologies (ARO).

References

1. Argos, P. and Palau, J. (1982) Amino-acid distribution in protein secondary structures. *Int. J. Pept. Protein Res. 19*, 380-393.

2. Presta, L. G. and Rose, G. D. (1988) Helix signals in proteins. *Science 240*, 1632-1641.

3. Richardson, J. S. and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of α helices. *Science 240*, 1648-1652.

4. Aurora, R. and Rose, G. D. (1998) Helix capping. *Protein Sci.* 7, 21-38.

5. Doig, A. J. and Baldwin, R. L. (1995) N- and C-capping preferences for all 20 amino acids in α-helical peptides. *Protein Sci. 4*, 1325-1336.

6. Marshall, S. A., Morgan, C. S., and Mayo, S. L. (2002) Electrostatics significantly affect the stability of designed homeodomain variants. *J. Mol. Biol. 316*, 189-199.

7. Addadi, L., Jaffe, E. K., and Knowles, J. R. (1983) Secondary tritium isotope effects as probes of the enzymic and nonenzymic conversion of chorismate to prephenate. *Biochemistry 22*, 4494-4501.

8. Sogo, S. G., Widlanski, T. S., Hoare, J. H., Grimshaw, C. E., Berchtold, G. A., and Knowles, J.R. (1984) Stereochemistry of the rearrangement of chorismate to prephenate: Chorismate mutase involves a chair transition state. *J. Am. Chem. Soc. 106*, 2701-2703.

9. Copley, S. D., and Knowles, J. R. (1985) The uncatalyzed Claisen rearrangement of chorismate to prephenate prefers a transition state of chairlike geometry. *J. Am. Chem. Soc. 107*, 5306-5308.

10. Gustin, D. J., Mattei, P., Kast, P., Wiest, O., Lee, L., Cleland, W. W., and Hilvert, D. (1999) Heavy atom isotope effects reveal a highly polarized transition state for chorismate mutase. *J. Am. Chem. Soc. 121*, 1756-1757.

11. Wright, S. K., DeClue, M. S., Mandal, A., Lee, L., Wiest, O., Cleland, W. W., and Hilvert, D. (2005) Isotope effects on the enzymatic and nonenzymatic reactions of chorismate. *J. Am. Chem. Soc. 127*, 12957-12964.

12. Cotton, G. H. and Gibson, F. (1965) The biosynthesis of phenylalanine and tyrosine: Enzymes converting chorismic acid into prephenic acid and their relationships to prephenate dehydratase and prephenate dehydrogenase. *Biochim. Biophys. Acta 100*, 76-88.

13. Stewart, J., Wilson, D. B., and Ganem, B. (1990) A genetically engineered monofunctional chorismate mutase. *J. Am. Chem. Soc. 112*, 4582-4584.

14. Lee, A. Y., Karplus, P. A., Ganem, B., and Clardy, J. (1995) Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase. *J. Am. Chem. Soc. 117*, 3627-3628.

15. Liu, D. R., Cload, S. T., Pastor, R. M., and Schultz, P. G. (1996) Analysis of active site residues in *Escherichia coli* chorismate mutase by site-directed mutagenesis. *J. Am. Chem. Soc. 118*, 1789-1790.

16. Kast, P., Asif-Ullah, M., and Hilvert, D. (1996) Is chorismate mutase a prototypic entropy trap? – Activation parameters for the *Bacillus subtilis* enzyme. *Tet. Lett. 37*, 2691-2694.

17. Kast, P., Asif-Ullah, M., Jiang, N., and Hilvert, D. (1996) Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: The importance of electrostatic catalysis. *Proc. Natl. Acad. Sci. USA 93*, 5043-5048.

18. Serrano, L. and Fersht, A. R. (1989) Capping and α-helix stability. *Nature 342*, 296-299.

19. Doig, A. J., MacArthur, M. W., Stapley, B. J., and Thornton, J. M. (1997) Structures of N-termini of helices in proteins. *Protein Science 6*, 147-155.

20. Övkist, M., Dey, R., Sasso, S., Grahn, E., Kast, P, and Krengel, U. (2006) 1.6 Å crystal structure of the secreted chorismate mutase from *Mycobacterium tuberculosis*: Novel fold topology revealed. *J. Mol. Biol. 357*, 1483-1499.

21. Lassila, J. K., Keeffe, J. R., Oelschlaeger, P., and Mayo, S. L. (2005) Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity. *Prot. Eng. Des. Sel. 18*, 161-163.

22. Gamper, M., Hilvert, D., and Kast, P. (2000) Probing the role of the C-terminus of *Bacillus subtilis* chorismate mutase by a novel random protein-termination strategy. *Biochemistry 39*, 14087-14094.

23. Zhang, S., Kongsaeree, P., Clardy, J., Wilson, D. B., and Ganem, B. (1996) Site-directed mutagenesis of monofunctional chorismate mutase engineered from the *E. coli* P-protein. *Bioorg. Med. Chem. 4*, 1015-1020.

24. Zhang, S., Wilson, D. B., and Ganem, B. (2003) An engineered chorismate mutase with allosteric regulation. *Bioorg. Med. Chem. 11*, 3109-3114.

25. Becktel, W. J. and Schellman, J. A. (1987) Protein stability curves. *Biopolymers 26*, 1859-1877.

Chapter 6

Perspectives on Computational Enzyme Design from Mutational Studies and Attempted Design of a New Chorismate Mutase

The major objective of this thesis research was to design a completely new enzyme to catalyze the Claisen rearrangement of chorismate to prephenate. Following the same process as described in Chapter 2, computational protein design methods were used to optimize side chains around an *ab initio* calculated transition state structure[1] for the chorismate-prephenate rearrangement. Instead of performing designs on a natural enzyme as in the experiments described so far, a series of protein scaffolds were selected that had no existing chorismate mutase activity. Roughly eighteen protein scaffolds were investigated, a large number of individual protein variants were purified and characterized, and variant protein libraries were constructed through error-prone PCR, gene synthesis, and degenerate codon mutagenesis. Both *in vitro* assays and the *in vivo* selection system of Kast, Hilvert, and co-workers[2,3] were used to test for activity. Ultimately, no designed protein catalyzed the rearrangement. What follows is not a complete analysis of the efforts, but simply an anecdotal discussion of some of the lessons that were learned from both the *de novo* design effort and from the mutational studies described in Chapters 3, 4 and 5.

Computational design methods performed well with many natural enzyme residues

As described in detail in Chapters 2 and 3, the computational design process was able to recover the geometrical configuration of the wild-type *E. coli* chorismate mutase active site and the critical side chains known to be required for catalysis. This is not surprising, given that the calculations were designed to look for active site configurations with geometries similar to those seen in the wild type. Yet the reproduction of the natural active site represents a necessary first step in being able to identify enzyme-like configurations within other scaffolds.

The test calculation described in Chapter III yielded six mutations in secondary active site positions—residues that are not known to be essential for catalysis. When these mutations are compared to the results from exhaustive mutagenesis performed in Chapters 4 and 5, it becomes clear that some residues were modeled quite successfully, even though the wild-type residue was not recovered. Table 1 shows a summary of these findings. Mutations in four positions led to increases in activity and/or stability relative to wild type, with three of the predicted mutations found to be the most stable or most active of the possible substitutions. Mutations to two positions led to reduced activity. Most of the changes were relatively small in magnitude, corresponding to changes in energy of activation of less than 1 kcal/mol. Only position 48 showed a loss of activity of multiple orders of magnitude. These results are encouraging, suggesting that out of 18 residues designed, only one position presented serious problems in modeling the wild-type enzyme.

Table 1. Summary of positions mutated in EcCM computational design experiment

Variant	Change in Stability	Change in Activity
Leu7Ile	4.4° C more stable than wild type; the most stable variant in that position.	Activity maintained within error
Ala32Ser	Stability maintained within error	59% increase in catalytic efficiency; the most active variant in that position.
Val35Ile	5.2° C more stable than wild type	13% increase in catalytic efficiency; the most active variant in that position
Asp48Ile	n.d.	Reduced activity significantly
Ile81Leu	Reduced stability 3.7° C	Reduced activity 40%
Val85Ile	1.5° C more stable than wild type	40% increase in k_{cat}; efficiency reduced

Inadequacies of the model may lead to significant problems in some positions

As described above and in Chapters 3 and 5, the Asp48Ile mutation led to a dramatic loss of catalytic activity. While the exact reduction in activity was not quantified, results from the assay procedure suggest that the reduction was greater than 10^3-fold. Asp48 simultaneously caps the helix N-terminus and interacts with an arginine side chain that is critical for catalysis (Figure 1). While it is disturbing that the energy function would choose isoleucine in a position that clearly requires an N-capping interaction, this discussion will focus specifically on implications for catalysis. A loss in stability would have been expected, but the impact of loss of the Arg11-Asp48 salt bridge on catalytic activity is perhaps less obvious. A predicted structure of the Asp48Ile variant suggested that the interactions of solvent and neighboring side chains could continue to stabilize the desired conformation of Arg11 and the Ile48 side chain would not likely destabilize the reacting molecule. As demonstrated in Chapter 5, however, all mutations in position 48 led to reduced catalytic activity and increased K_M.

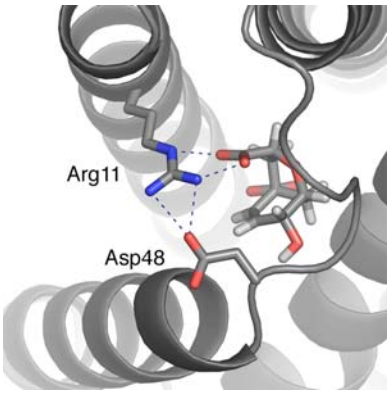


Figure 1. Asp48 caps the helix N-terminus while forming a salt bridge with catalytically essential Arg11 (shown contacting the crystallographic transition state analog).

As discussed in Chapter 5, the loss of the Asp48-Arg11 salt bridge may affect activity by leading to a loss of precision in the positioning of Arg11 for interaction with the reacting molecule or by allowing the Arg11 side chain to reorient towards solvent. Alternatively, it could impair the substrate entry process and cause substrate binding to impact the rate-limiting step. The available data do not allow us to distinguish between these or other possibilities. However, one thing is certain—a protein design model that optimizes for global stability in a single state does not account for these complex effects.

The inadequacy of the model for dealing with position 48 is illustrated in Figure 2, where calculated energies for all 20 position 48 variants are plotted against experimental catalytic efficiencies. These calculations were performed using a standard energy function and the methods outlined in Chapter 2. An *ab initio* calculated transition state structure was superimposed with the crystallographic transition state analog and then allowed to translate and rotate within the binding pocket. Regardless of whether the total energy of the system or the energy of the transition state structure is plotted, the majority of the variants are assigned lower energies than Asp48. While a linear correlation would not necessarily be expected given known limitations in protein design methods, the fact that the only efficient enzyme, the wild type, was ranked as one of the highest-energy amino acids is certainly cause for concern. Repulsion of the charged Asp side chain with the carboxylate of the reacting molecule results in unfavorable electrostatic energy. Perhaps more importantly, the model is not sophisticated enough to register any favorable effect that the Asp side chain might have on the reacting molecule indirectly through its interaction with the arginine. Relevant elements that are completely neglected in the calculation include the side-chain entropy of the arginine, polarization effects, orientations of the substrate and side chains in the unbound state, dynamics of the substrate entry process, and changes in backbone structure resulting from loss of the N-capping interaction.

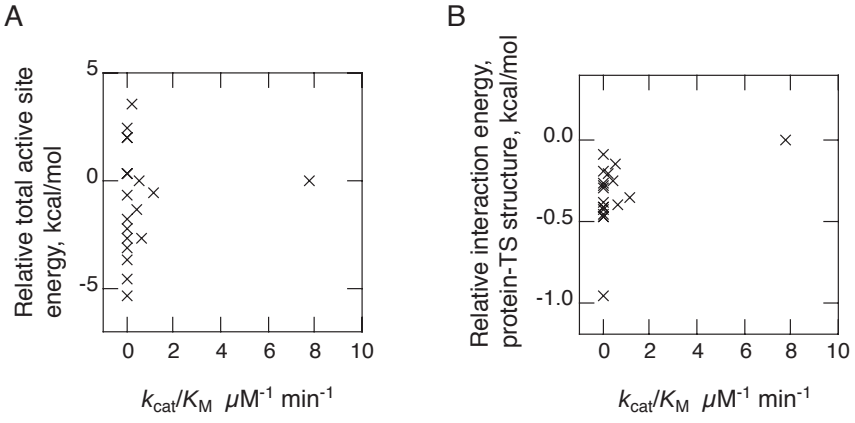


Figure 2. Calculated energies do not correlate with activity in position 48 variants. Calculated energies (using standard design force field) are plotted against experimental catalytic efficiencies for all 20 position 48 variants. Catalytic efficiencies of biologically inactive variants are set to zero. Wild-type Asp48 is the single data point with high catalytic efficiency (k_{cat}/K_M). (A) Total calculated active site energies. Includes contributions from all side chain, backbone, and TS structure interactions within the active site. Normalized to the wild-type energy. (B) Interaction energies of TS structure with side chain and backbone elements in the active site only. Normalized to the wild-type energy.

Selection of robust starting scaffolds is critical for design

At one point during the efforts toward attempted design of a new chorismate mutase, a structure-based approach to scaffold selection was tested. A computational procedure was developed to search the Protein Data Bank for protein scaffolds containing binding sites similar in volume and shape to that of *E. coli* chorismate mutase. One of the structures identified was an enzyme with a TIM barrel fold that had some similarities in active site geometry and electrostatics to the natural chorismate mutase active sites. Because design calculations yielded favorable results, designed proteins based on this scaffold were constructed and experimentally evaluated.

The wild-type protein was purified in good yield from the soluble cell fraction and showed all evidence of normal folding behavior by circular dichroism. However, a designed active site variant of the scaffold, a 10-fold mutant, was found primarily in the insoluble cell fraction following lysis and centrifugation. Changes in the expression temperature and protocol yielded no improvement of solubility, and refolding attempts under a variety of conditions failed. To investigate how seriously the design mutations affected the expression properties of the protein, a minimal design with 5 mutations was constructed. This variant was found to express in the insoluble fraction, as were single and double mutations of the wild type.

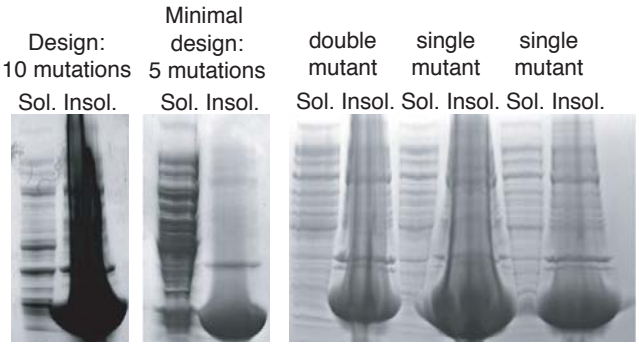


Figure 3. Although the wild-type TIM barrel expressed in the soluble fraction with good yield and appeared well folded by circular dichroism, both computationally designed variants and minimal versions thereof were found in the insoluble cell fraction. SDS-PAGE gels are shown, with cell extracts loaded as indicated. *Sol.*, soluble cell fraction; *Insol*, insoluble cell fraction.

The scaffold was originally selected because of similarity with natural chorismate mutase active sites in hopes that conservation of existing active site elements would minimize disruption. However, the scaffold ultimately was not able to accommodate even single mutations to the active site. Future efforts utilized a monomeric TIM barrel from a hyperthermophilic organism, which proved to be much more tolerant to mutation. This experience underscores the importance of identifying a robust starting scaffold for active site redesign, even at the cost of potentially making more mutations.

Enzyme design is a difficult task

At least a million variants were screened in the present effort to design a new chorismate mutase and no active enzyme was found. A graduate student who continued to construct individual designed molecules at the rate that they were made in the early years of this thesis research might be better off waiting for natural evolution to produce the new enzyme. The probability of a single design having catalytic activity will always be less than one, and in many cases it will be less than one in a million. Accordingly, it makes sense to construct and screen large libraries of designed variants rather than testing individual designs. Enzyme design efforts during the later years of this thesis research took this approach, combining the strengths of computational design and directed evolution methods. Powerful approaches have been developed for construction and evaluation of gene libraries[4-8] and the combination of computational design and directed evolution[9-11] will likely yield many successful new enzymes in the future.

Even with the combined computational design and directed evolution approach, an unfortunate consequence of work on this difficult problem is that a considerable amount of effort has been expended and yet relatively little understanding has been gained. A large number of computationally designed proteins have been constructed and evaluated for their ability to catalyze a reaction. In no case was any measurable catalytic activity identified in a designed protein. Because the absence of catalytic activity provides no feedback about energy functions, conformational sampling, substrate binding, or transition state stabilization, we are left without any understanding of what might have gone wrong and no pathway for improvement.

In planning future projects, those attempting computational enzyme design would benefit from taking the time to develop systems that can provide experimental feedback through intermediate steps. Clearly, the binding of a substrate or transition state analog represents a first step to enzyme development and could be monitored. In the case of the chorismate-prephenate rearrangement, binding of Bartlett's transition state analog[12] has been measured with isothermal titration calorimetry.[13] If a designed protein does not demonstrate some affinity for this inhibitor, it has little hope of becoming a catalyst for the rearrangement, even with further optimization by directed evolution. In a new system, however, a higher-throughput method would be preferable in order to screen designed protein libraries for variants with binding affinity. One would ideally develop creative strategies to systematically identify candidates for further optimization based on binding and turnover events. The directed evolution literature contains elegant examples of techniques applied for these purposes.

Another important intermediate step in new design efforts may simply be to perform design calculations on natural enzymes that catalyze reactions similar to the target reaction. Experiments like those of Chapters II and III could help to clarify appropriate ways to model the transition state structure and catalytic side chains in the new system and could help to identify major obstacles to accurate modeling. In general, more applications of enzyme design to naturally existing enzymes would help to provide reality checks for the design field, indicating systematic weaknesses in the methods and providing direction for improvement.

References

1. Wiest, O. and Houk, K. N. (1994) On the transition state of the chorismate-prephenate rearrangement. *J. Org. Chem. 59*, 7582-7584.

2. Kast, P., Asif-Ullah, M., and Hilvert, D. (1996) Is chorismate mutase a prototypic entropy trap? Activation parameters for the *Bacillus subtilis* enzyme, *Tet. Lett. 37*, 2691-2694.

3. Kast, P., Asif-Ullah, M., Jiang, N., and Hilvert, D. (1996) Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: The importance of electrostatic catalysis, *Proc. Natl. Acad. Sci. USA 93*, 5043-5048.

4. Arnold, F. H. and Volkov, A. A. (1999) Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol. 3*, 54-59.

5. Reetz, M. T. (2001) Combinatorial and evolution-based methods in the creation of enantioselective catalysis. *Angew. Chem. Int. Ed. 40*, 284-310.

6. Woychechowsky, K. J. and Hilvert, D. (2004) Deciphering enzymes: Genetic selection as a probe of structure and mechanism. *Eur. J. Biochem. 271*, 1630-1637,

7. Bloom, J. D., Meyer, M. M., Meinhold, P, Otey, C. R., MacMillan, D., and Arnold, F. H. (2005) Evolving strategies for enzyme engineering, *Curr. Opin. Struct. Biol. 25,* 447-452.

8. Aharoni, A., Griffiths, A. D., and Tawfik, D. S. (2005) High-throughput screens and selections of enzyme-encoding genes, *Curr. Opin. Chem. Biol. 9*, 210-216.

9. Voigt, C. A., Mayo, S. L., Arnold, F. H., and Wang, Z. G. (2001) Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci. USA 98*, 3778-3783.

10. Bolon, D. N., Voigt, C. A., and Mayo, S. L. (2002) *De novo* design of biocatalysts, *Curr. Opin. Chem. Biol. 6*, 125-129.

11. Hayes, R. J., Bentzien, J., Ary, M. L., Hwang, M. Y., Jacinto, J. M., Vielmetter, J., Kundu, A., and Dahiyat, B. I. (2002) Combining computational and experimental screening for rapid optimization of protein properties. *Proc. Natl. Acad. Sci. USA 99*, 15926-15931.

12. Bartlett, P. A., and Johnson, C. R. (1985) An inhibitor of chorismate mutase resembling the transition-state conformation, *J. Am. Chem. Soc. 107*, 7791-7793.

13. Lee, A. Y., Zhang, S., Kongsaeree, P., Clardy, J., Ganem, B., Erickson, J. W., and Xie, D. (1998) Thermodynamics of a transition state analog inhibitor binding to *Escherichia coli* chorismate mutase: Probing the charge state of an active site residue and its role in inhibitor binding and catalysis, *Biochemistry 37*, 9052-9057.

Appendix A

Evaluation of the Energetic Contribution of an Ionic Network to Beta-sheet Stability

The text of this chapter was adapted from a published manuscript coauthored with Deepshikha Datta and Stephen L. Mayo.

Lassila, K. S., Datta, D., and Mayo, S. L. (2002) Evaluation of the energetic contribution of an ionic network to beta-sheet stability. *Protein Science 11*, 688-690.

Abstract

We have evaluated the interaction energy of a three residue ionic network constructed on the beta-sheet surface of protein G using double mutant cycles. Although the two individual ion pairs were each stabilizing by around 0.6 kcal/mol, the excess gain in stability for the triad was small (0.06 kcal/mol).

The β-sheet surface of the protein G immunoglobulin-binding domain B1 (GB1) has been used as a model system for evaluating the β-sheet forming propensities of amino acids.[1,2] These studies, in combination with statistical surveys of known structures and theoretical models of β-sheet propensity provide some general guidelines for amino acid selection in β-sheet design.[3,4]

An important next step in understanding β-sheet stability is to define the role of side chain interactions such as hydrogen bonding and ionic interactions. In particular, the energetic effects of surface ionic interactions have been debated. Solvent exposed ion pairs have been found to stabilize folded proteins in a number of cases.[5-9] In the context of the β-sheet surface environment, ion pairs have been reported to stabilize folded proteins by 0.4–1.0 kcal/mol.[10-12] However, some surface ion pairs exhibit neutral or destabilizing effects.[13,14] The high dielectric of the aqueous environment and the loss of side chain conformational freedom have been invoked to explain the marginal stabilizing effects of some pairwise electrostatic interactions.

Networks of charged surface residues have been observed in hyperthermophile proteins and have been proposed to offer an energetic advantage over single ion pairs due to the reduced entropic cost of fixing a third residue.[14,15] Indeed, two analyses of solvent exposed ionic triads in α-helical regions have shown that three-residue networks offer a stabilizing effect greater than would be observed for the sum of the two individual pairwise interactions.[6,7]

To test the effect of an ionic triad in the context of the β-sheet surface, we have evaluated the energetic contribution of a three-residue triad constructed on the β-sheet surface of GB1. The network consists of Arg 6, Glu 53, and Arg 44, residues which lie on three adjacent strands of the β-sheet surface (Figure 1). Double mutant cycle analysis was used to isolate the interaction energy of the triad.[16] Eight GB1 variants were constructed which represent all permutations of Arg or Ile at position 6, Glu or Ala at position 53, and Arg or Ala at position 44. In this three-residue thermodynamic cycle, the interaction energy of the ionic network is calculated as in equation 1.

$$\Delta\Delta\Delta G^{RER}_{interaction} = \{(\Delta G^{RER} - \Delta G^{RAA}) - [(\Delta G^{REA} - \Delta G^{RAA}) + (\Delta G^{RAR} - \Delta G^{RAA})]\} - \{(\Delta G^{IER} - \Delta G^{IAA}) - [(\Delta G^{IEA} - \Delta G^{IAA}) + (\Delta G^{IAR} - \Delta G^{IAA})]\} \quad [\mathbf{1}]$$

ΔG^{XYZ} is the free energy of unfolding for the GB1 mutant with amino acids X, Y, and Z at positions 6, 53, and 44, respectively. The interaction energy of an Arg-Glu ion pair in the presence of another residue X is calculated as in equation 2.

$$\Delta\Delta G^{XER}_{interaction} = (\Delta G^{XER} - \Delta G^{XAA}) - [(\Delta G^{XEA} - \Delta G^{XAA}) + (\Delta G^{XAR} - \Delta G^{XAA})] \quad [\mathbf{2}]$$

Free energies of unfolding (ΔG) were evaluated by two-state analysis of thermal denaturation curves monitored by circular dichroism (CD) (Figure 2).

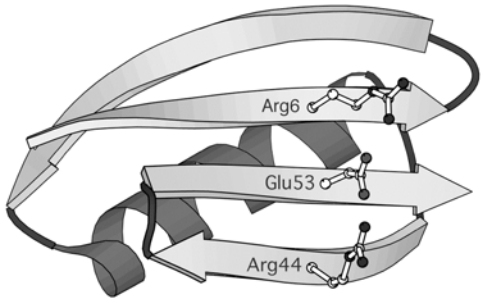


Figure 1. The β-sheet surface of GB1 showing possible orientations for side chains Arg 6, Glu 53, and Arg 44. In the positions shown, nitrogen-oxygen distances are 2.92 Å and 2.85 Å for residue pairs 6-53 and 44-53, respectively. Side chains were positioned with a dead-end elimination algorithm[19] and the figure was created with MOLSCRIPT[20].

The variant containing both a single ion pair and isoleucine, $I_6E_{53}R_{44}$, had the highest T_m and ΔG of unfolding (Table 1) while the variant with the ionic triad, $R_6E_{53}R_{44}$ was only slightly less stable. It is interesting to note that the addition of the third charged residue almost fully compensates for the loss of the β-branched (and therefore β–sheet stabilizing) amino acid.

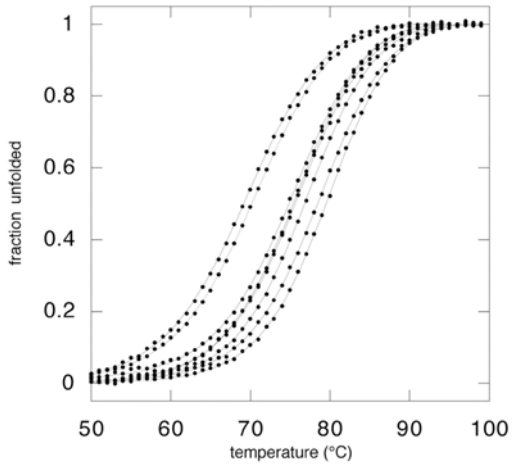


Figure 2. Thermal denaturation curves for GB1 variants. From left to right (at 50% unfolded): $R_6A_{53}R_{44}$, $R_6A_{53}A_{44}$, $I_6A_{53}R_{44}$, $I_6A_{53}A_{44}$, $R_6E_{53}A_{44}$, $I_6E_{53}A_{44}$, $R_6E_{53}R_{44}$, and $I_6E_{53}R_{44}$.

Table 1. Stability data for GB1 variants

Variant	Tm[a] (°C)	ΔH_{Tm}[b] (kcal/mol)	ΔG (75 °C)[c] (kcal/mol)
$R_6E_{53}R_{44}$	79.1 ± 0.5	53.5 ± 2.0	0.61
$R_6E_{53}A_{44}$	76.1 ± 0.4	52.3 ± 2.0	0.16
$R_6A_{53}R_{44}$	69.6 ± 0.3	45.0 ± 1.6	-0.74
$R_6A_{53}A_{44}$	70.6 ± 0.3	46.6 ± 1.5	-0.62
$I_6E_{53}R_{44}$	80.1 ± 0.6	56.1 ± 2.5	0.79
$I_6E_{53}A_{44}$	77.2 ± 0.4	54.1 ± 1.9	0.34
$I_6A_{53}R_{44}$	75.5 ± 0.3	50.1 ± 1.5	0.08
$I_6A_{53}A_{44}$	76.0 ± 0.3	51.4 ± 1.6	0.14

[a] T_m, midpoint of thermal denaturation transition
[b] ΔH_{Tm}, enthalpy of unfolding at T_m
[c] ΔG (75 °C), free energy of unfolding calculated at 75 °C

Interaction energies of the Arg6-Glu53 pair were 0.58 kcal/mol in the presence of Ala44 and 0.64 kcal/mol in the presence of Arg44 (Table 2). The Arg44-Glu53 pair had interaction energies of 0.51 kcal/mol (in the presence of Ile6) and 0.57 kcal/mol (with Arg6). This level of stabilization is consistent with other surface electrostatic interactions studied by double mutant cycles.[7,9,10]

Although the pairwise electrostatic interactions are clearly favorable, the ionic network does not appear to significantly enhance GB1 stability any more than the simple sum of the individual pairs. As shown in Table 2, the interaction energy of unfolding for the ionic network, $\Delta\Delta\Delta G_{interaction}$, determined at 75 °C (approximately the average T_m for the eight variants) was 0.06 kcal/mol. This very low interaction energy suggests that the contributions of the ion pairs are additive; there is no additional stabilization of one ion pair in the presence of a third charged residue. In contrast, previous studies of charged networks on α-helices using the double mutant cycle method showed stabilizing interaction energies of 0.77 kcal/mol for an Asp-Arg-Asp triad[6] and 0.65 kcal/mol for an Arg-Glu-Arg triad.[7]

Table 2. Interaction energies for ion pairs and the three-residue network

Interaction	ΔΔG (75 °C)[a] Kcal/mol	ΔΔΔG (75 °C)[b] Kcal/mol
R_6E_{53} (A_{44})	0.58	
R_6E_{53} (R_{44})	0.64	
$E_{53}R_{44}$ (I_6)	0.51	
$E_{53}R_{44}$ (R_6)	0.57	
$R_6E_{53}R_{44}$		0.06

[a]ΔΔG (75 °C) interaction energy (calculated at 75 °C) of the ion pair in the presence of the residue indicated in parentheses

[b]ΔΔΔG (75 °C) interaction energy (calculated at 75 °C) of the triad as described in the text

The lack of a significant stabilizing interaction energy of the Arg6-Glu53-Arg44 triad may be due to a variety of factors. Previously reported factors such as desolvation, side chain entropy loss, and conformational strain may counteract the electrostatic benefits of the network. However, the local environment of the triad, including secondary structure and neighboring residues, may also affect the magnitude of the interaction energy of the triad. Further studies on β-sheet surface electrostatic interactions may help to clarify whether or not secondary structure influences the stabilizing effect of ionic networks.

Materials and Methods

Mutagenesis and protein expression

GB1 variants were constructed by inverse PCR mutagenesis and expressed using the T7 promoter system as previously described.[17] Purification of 57-residue GB1 variants containing an N-terminal methionine was accomplished by reverse phase HPLC and verified by mass spectrometry.

Thermal denaturation

The increase in CD signal at 218 nm was followed during thermal unfolding from 1 °C to 99 °C using 50 μM protein in 50 mM sodium phosphate, pH 5.5. The midpoint of the thermal denaturation (T_m) and the enthalpy of unfolding (ΔH) were determined from a two-state analysis of each denaturation curve.[1,2] The change in heat capacity upon unfolding (ΔC_p) was held constant at 0.621 kcal/K•mol, a value previously reported for wild-type GB1.[18] ΔG values were assigned using the Gibbs-Helmholtz relation with ΔC_p = 0.621 kcal/K•mol.[1,2] The average error in calculating ΔG (as determined from curve fitting) was 0.06 kcal/mol.

References

1. Minor, D. L. and Kim, P. S. (1994) Measurement of the β-sheet-forming propensities of amino acids. *Nature 367*, 660-663.

2. Smith, C. K., Withka, J. M., and Regan, L. (1994) A thermodynamic scale for the β-sheet forming propensities of the amino acids. *Biochemistry 33*, 5510-5517.

3. Munoz, V. and Serrano, L. (1994) Intrinsic secondary structure propensities of the amino-acids, using statistical phi-psi matrices—comparison with experimental scales. *Proteins 20*, 301-311.

4. Street, A. G. and Mayo, S. L. (1999) Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc. Natl. Acad. Sci. USA 96*, 9074-9076.

5. Lyu, P. C. C., Gans, P. J., and Kallenbach, N. R. (1992) Energetic contribution of solvent-exposed ion-pairs to alpha-helix structure. *J. Mol. Biol. 223*, 343-350.

6. Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990) Strength and cooperativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol. 216*, 1031-1044.

7. Spek, E. J., Bui, A. H., Lu, M., and Kallenbach, N. R. (1998) Surface salt bridges stabilize the GCN4 leucine zipper. *Protein Sci. 7*, 2431-2437.

8. Takano, K., Tsuchimori, K., Yamagata, Y., and Yutani, K. (2000) Contribution of salt bridges near the surface of a protein to the conformational stability. *Biochemistry 39*, 12375-12381.

9. Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A. R. (1990) Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry 29*, 9343-9352.

10. Merkel, J. S., Sturtevant, J. M., and Regan, L. (1999) Sidechain interactions in parallel beta sheets: the energetics of cross-strand pairings. *Structure 7*, 1333-1343.

11. Blasie, C. A. and Berg, J. M. (1997) Electrostatic interactions across a beta-sheet. *Biochemistry 36*, 6218-6222.

12. Smith, C. K. and Regan, L. (1995) Guidelines for protein design- the energetics of β sheet side-chain interactions. *Science 270*, 980-982.

13. Strop, P. and Mayo, S. L. (2000) Contribution of surface salt bridges to protein stability. *Biochemistry 39*, 1251-1255.

14. Dao-pin, S., Sauer, U., Nicholson, H., and Matthews, B. W. (1991) Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry 30*, 7142-7153.

15. Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R., and Rice, D. W. (1995) The structure of *Pyrococcus-furiosis* glutamate-dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure 3*, 1147-1158.

16. Horovitz, A. and Fersht, A. R. (1990) Strategy for analyzing the cooperativity of intramolecular interactions in peptides and proteins. *J. Mol. Biol. 214*, 613-617.

17. Su, A. and Mayo, S. L. (1997) Coupling backbone flexibility and amino acid sequence selection in protein design. *Protein Sci. 6*, 1701-1707.

18. Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) Thermodynamic analysis of the folding of the streptococcal protein-G IgG-binding domains B1 and B2- why small proteins tend to have high denaturation temperatures. *Biochemistry 31*, 3597-3603.

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

20. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallography 24*, 946-950.

Appendix B

Dioxane Contributes to the Altered Conformation and Oligomerization State of a Designed Engrailed Homeodomain Variant

The text of this chapter was adapted from a published manuscript coauthored with Geoffrey K. Hom, Leonard M. Thomas, and Stephen L. Mayo.

Hom, G. K., Lassila, J. K., Thomas, L. M., and Mayo, S. L. (2005) Dioxane contributes to the altered conformation and oligomerization state of a designed engrailed homeodomain variant. *Protein Science 14*, 1115-1119.

Abstract

Our goal was to compute a stable, full-sequence design of the *Drosophila melanogaster* engrailed homeodomain. Thermal and chemical denaturation data indicated the design was significantly more stable than the wild-type protein. The data were also nearly identical to those for a similar, later full-sequence design, which was shown by NMR to adopt the homeodomain fold: a three-helix, globular monomer. However, a 1.65 Å crystal structure of the design described here turned out to be of a completely different fold: a four-helix, rodlike tetramer. The crystallization conditions included ~25% dioxane, and subsequent experiments by circular dichroism and sedimentation velocity analytical ultracentrifugation indicated that dioxane increases the helicity and oligomerization state of the designed protein. We attribute at least part of the discrepancy between the target fold and the crystal structure to the presence of a high concentration of dioxane.

Introduction

The original purpose of this project was to computationally design an amino acid sequence that stably adopts the homeodomain fold. The target fold was the same as for previous homeodomain designs from our lab,[1,2] a 51-residue, crystallographically well-defined fragment of the *Drosophila melanogaster* engrailed homeodomain[3] (Figure 1). This fragment is a globular, three-helix monomer. As in our previous homeodomain designs, we did not consider DNA binding but rather focused on protein stability.

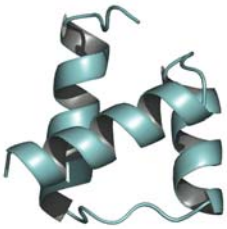


Figure 1. Target homeodomain fold for UMC.

We designed two sequences, UMC and UVF[4]. UMC was obtained via a Monte Carlo algorithm.[5] UVF has a slightly lower computed energy and could be obtained via either the Vegas[6] or the FASTER[7] algorithm.

UMC and UVF have 79% sequence identity and also have nearly identical thermal and chemical denaturation profiles. For both proteins, the melting temperature is $> 99°C$ and $\Delta G_{unfolding}$ is 4.2 kcal/mol. The 1D 1H NMR spectra of the proteins display the characteristics expected of a monomer, and the NMR-determined structure of UVF matches the homeodomain fold.[4]

The above evidence indicated that UMC also adopts the homeodomain fold. However, a crystal structure of UMC would give direct confirmation of the overall fold and allow for a detailed comparison of crystallographic and computed side-chain conformations, which would provide critical data for improving our protein design algorithm.[8]

Here we report a 1.65 Å crystal structure of UMC. The structure is a rodlike, four-helix tetramer (Figure 2), not the expected globular, three-helix monomer. This discrepancy could be due to a lack of explicit negative design in our design algorithm; however, because of the similarity of UMC to the successful UVF design, we

investigated if the crystallographic conditions could be responsible for the discrepancy. In particular, the role of dioxane was examined.

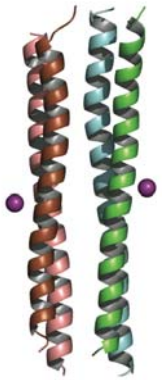


Figure 2. Ribbon diagram of the UMC crystal structure. The coloring is cyan (chain A), green (chain B), brown (chain C), salmon (chain D), and purple (cadmium).

Results

The crystal structure of UMC was determined by using single wavelength anomalous diffraction (SAD). Crystallographic statistics are shown in Table 1. The asymmetric unit contains four UMC molecules forming an anti-parallel helical bundle with one UMC molecule per helix. Main-chain and side-chain density could not be interpreted for some terminal residues (residues 1-3 of chain A; 1-4, 51-52 of chain B; 1-2, 47-52 of chain C; and 1-4, 51-52 of chain D). The asymmetric unit also contains two cadmium atoms, one acetate molecule and ten dioxane molecules. The cadmium atoms are each coordinated by four carboxylate anions: one cadmium is coordinated by four glutamate side chains (Figure 3), and the other cadmium is coordinated by three glutamate side chains and an acetate molecule.

Several dioxane molecules mediate helix-helix packing (Figure 4). This observation led us to examine the effect of dioxane on the helicity and oligomerization of UMC in solution.

Helicity was examined by far-UV circular dichroism. Ellipticity was virtually unchanged when UMC was exposed to CdCl2 alone (not shown) or CdCl2 with 10% dioxane (Figure 5). However, exposure of UMC to 20% dioxane lowered the minima at

208 and 222 nm and was thus indicative of an increase in helicity. The increase, while significant, was still less than that for 30% trifluoroethanol (TFE), a helix stabilizer.[9] Higher percentages of dioxane did not further increase the helicity significantly (not shown).

Table 1. X-ray data collection and refinement statistics

	R-Axis IV	SSRL[a]
Unit cell		
a	50.767 Å	50.712 Å
b	52.562 Å	52.646 Å
c	82.147 Å	82.182 Å
Space group	$P2_12_12_1$	$P2_12_12_1$
Wavelength	1.5418 Å	0.8265 Å
Resolution range	81.65 – 1.90 Å	44.32 – 1.65 Å
No. of reflections collected	208991	204302
No. of unique reflections	17953	27060
R_{merge}[b]	5.6% (55.1%)[c]	4.7% (19.6%)
I/σ(I)	10.1 (1.3)	31.8 (8.5)
Completeness	99.9% (99.5%)	99.8% (100.0%)
Final refinement		
R_{cryst}		18.7%
R_{free}[d]		22.7%
Figure of merit		0.863
No. of residues		368
No. of water molecules		168
No. of non protein molecules		11
Mean B value		28.1 Å^2
RMSD from standard stereochemistry		
Bond length		0.017 Å
Bond angle		1.527°
Ramachandran plot statistics		
Most favored regions		99.4%
Additional allowed regions		0.6%
Generously allowed regions		0.0%
Disallowed regions		0.0%

[a] Stanford Synchrotron Radiation Laboratory.
[b] $R_{merge} = \Sigma \left| I - \langle I \rangle \right| / \Sigma (I)$, where I is the observed intensity and <I> is the average intensity.
[c] Numbers in parentheses represent values in the highest resolution shell (1.90-1.99 Å for the R–Axis IV data and 1.652-1.695 Å for the SSRL data).
[d] R_{free} was calculated for 5% of randomly selected reflections excluded from refinement.

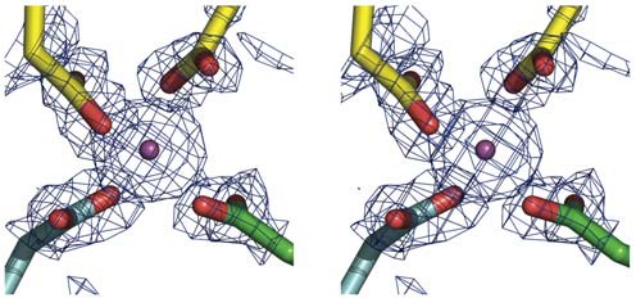


Figure 3. Coordination of one of the two cadmium atoms by four glutamates. The coloring is cyan (chain A), green (chain B), yellow (chain C), purple (cadmium), and red (oxygen). The σA-weighted density map is contoured at 2σ, up to 3.5 Å from the cadmium. Chain C is from a symmetry-related molecule.

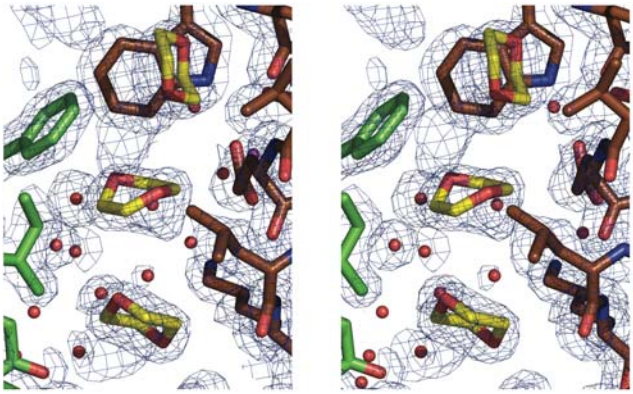


Figure 4. Dioxane molecules mediating helix-helix packing. The coloring is green (chain B), brown (chain C), yellow (dioxane), red (oxygen), and blue (nitrogen). The σA-weighted density map is contoured at 1σ, up to 3 Å from the dioxane.

Oligomerization was examined by sedimentation velocity analytical ultracentrifugation. Exposure to 20% dioxane significantly decreased the percentage of monomeric UMC, from 81.4% to 62.8%, and concomitantly increased the percentage of dimeric UMC, from 14.8% to 36.3% (Figure 6). The frictional ratio, which describes the shape of the sedimenting species, also increased. A sphere has a ratio of ~1.2, whereas rodlike shapes have higher ratios. The frictional ratio increased from 1.22 to 1.42 in the presence of dioxane.

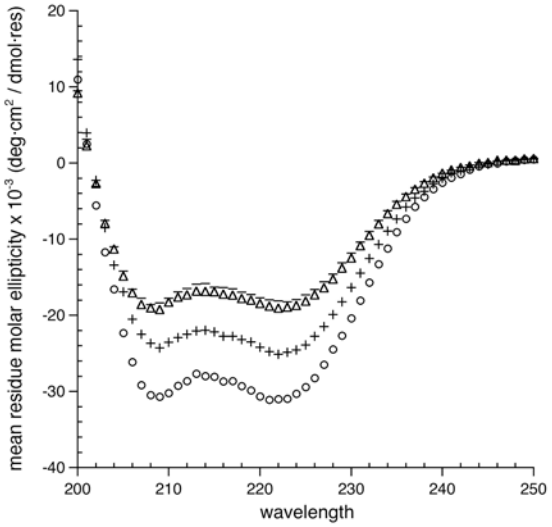


Figure 5. Far-UV circular dichroism analysis of UMC. The spectra are UMC (–); UMC in 5 mM CdCl2 and 10% dioxane (triangles); UMC in 20% dioxane (+); and UMC in 30% TFE (circles).

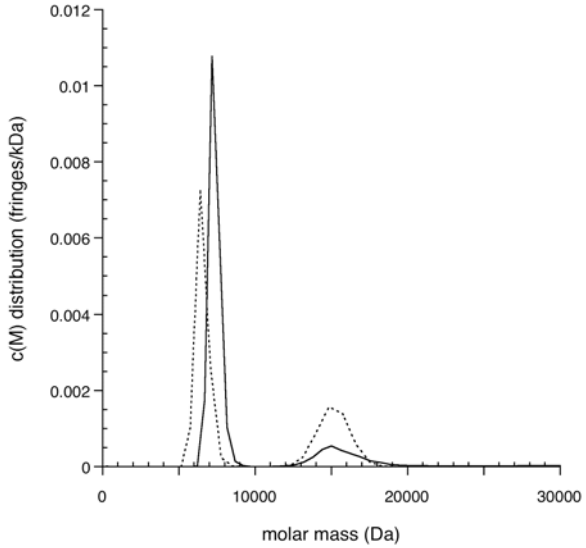


Figure 6. Molar mass distribution of UMC as determined by sedimentation velocity: UMC (solid line); UMC in 20% dioxane (dotted line).

Discussion

Our crystal structure of UMC is quite dissimilar to the target homeodomain fold. Instead of three short helices, each monomer is a single long helix. However, the crystallization conditions, especially the high concentration of dioxane, may induce UMC into a conformation unrepresentative of UMC in solution.

Increased helicity and oligomerization due to dioxane

Dioxane increased the helicity of UMC. While dioxane had a significant effect, the [θ]222 for 20% dioxane (–25,000 deg cm^2 / dmol res) was less negative than for 30% TFE (–31,000 deg cm^2 / dmol res).

The effect of dioxane on increasing helicity has been reported previously[10-12]. The increase in helicity can be explained entropically: non-polar solvent increases the entropic cost of forming protein hydrogen bonds to water and thus decreases the relative cost of forming helical hydrogen bonds. The use of organic solvents may have played a role in the crystallization of a number of short aminoisobutyric acid-containing peptides, which also adopt extended continuous helical structures.[13]

The sedimentation velocity data showed that dioxane increases the oligomerization state and frictional ratio of UMC. While there was a significant increase in the amount of dimer, there was no evidence of a tetramer, as might be expected from the crystal structure. One explanation is that formation of a tetrameric species requires cadmium. Although CdCl2 alone and CdCl2 with 10% dioxane had no effect on the helicity of UMC, low millimolar amounts of CdCl2 (e.g., 2–5 mM) caused essentially all UMC to precipitate out in the presence of > 15% dioxane. The UMC crystals appeared a couple of weeks after precipitate had formed in the well. Perhaps cadmium further increases the dioxane-induced helicity and/or oligomerization of UMC but requires the very slow mixing that occurs in the crystallization well.

Conclusion

Overall, the crystal structure has increased helicity and altered oligomerization compared to the target fold. Both of these differences were inducible by dioxane. We thus attribute at least part of the discrepancy between the target fold and the crystal structure of the designed sequence to the presence of a high concentration of dioxane. Although low concentrations (1%-2%) of dioxane have been reported to improve crystallization of some proteins,[14,15] we suggest that high concentrations of dioxane be used with caution.

Materials and methods

Protein design and purification

The UMC design, construction, expression, and purification were similar to our previous engrailed designs[1,2] and are described in detail elsewhere.[4] A brief summary is below.

The starting model for all engrailed designs was PDB entry 1enh.[3] Because residue 35 has a positive phi angle it was preserved as glycine. The UMC design protocol was identical to the B6 design protocol of Marshall and Mayo,[1] except that in the UMC design all residues were designed simultaneously, and a Monte Carlo simulation[5] was used instead of a dead-end elimination based algorithm[16] to find a low-energy sequence. The protein was expressed in *E. coli* and purified via freeze-thaw[17] followed by HPLC using an acetonitrile/water gradient containing 0.1% TFA. Mass spectrometry indicated UMC has an N-terminal methionine.

Crystallization

Crystals were obtained using a modified sitting drop method that utilizes a "reservoir mimic." The well reservoir is minimized to contain only the volatile reagents and NaCl. The nonvolatile reagents normally in the reservoir are kept in a separate solution (the mimic) that is only added to the crystallization drop.

We also used Fluorinert (Hampton Research), which is expected to be denser than the drop and allow the drop to float. Under our conditions, Fluorinert floated above the drop. However, this serendipitously slowed the otherwise rapid crystal degradation that would happen upon well opening and presumably due to the volatility of dioxane.

The initial crystallization condition was 35% dioxane (Hampton Research Crystal Screen 2). The final crystallization conditions were as follows: the well reservoir contained 500 μL of either 24 or 25% dioxane; the well post contained 1 μL of protein solution (~17 mg/mL UMC, 50 μM sodium citrate at pH 5.5) followed by 1 μL of reservoir mimic solution (0.1 M MES at pH 5.7, 30% PEG 400, 10–15 mM CdCl2); 20 μL of Fluorinert (Hampton Research) was then added on top of each post. Trays were incubated at 20° C; crystals appeared after about two weeks. The largest crystals had dimensions of approximately 150 × 150 × 200 μm.

Structure determination

Data were collected using a Cu source on a Rigaku RU3HR generator with an R–AXIS IV detector at 100 K. Data were processed using the HKL program suite v1.97.9.[18] Initial electron density maps were generated using SAD phasing as implemented in the program suite Elves.[19] The final model was determined by subsequent rounds of building and refinement using O[20] and REFMAC[21] from the CCP4 program suite[22] to an R–factor of 22.2% (Rfree = 27.8%). Final refinement was done with high-resolution data collected at beamline 9.2 at the Stanford Synchrotron Radiation Laboratory and produced a final R–factor of 18.7% (Rfree = 22.7%).

Coordinates and structure factors have been deposited in the PDB under the accession code 1Y66.

Circular dichroism and sedimentation velocity

Circular dichroism data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit. Wavelength scans were done from 190-250 nm at 20° C in a 0.1 mm path-length cell. All samples contained 532 μM UMC and 10 mM sodium citrate, pH 5.5. Protein concentration was determined by absorbance at 280 nm in the presence of 8 M guanidine HCl.

Sedimentation velocity data were collected on a Beckman XL-I analytical ultracentrifuge with interference optics. Samples contained 532 μM UMC and 0.1 M sodium citrate, pH 5.5. Samples were dialyzed for 3 h at room temperature against ~100 mL of the corresponding solution without protein. A 12 mm Epon centerpiece and sapphire windows were used. The rotor, an An-60 Ti, was spun at 55,000 RPM at 25° C. Scans were taken every 5 min for ~15 h. Data were analyzed with SEDFIT.[23]

References

1. Marshall, S. A., and Mayo, S. L. (2001) Achieving stability and conformational specificity in designed proteins via binary patterning. *J. Mol. Biol. 305*, 619-631.

2. Marshall, S. A., Morgan, C. S., and Mayo, S. L. (2002) Electrostatics significantly affect the stability of designed homeodomain variants. *J. Mol. Biol. 316*, 189-199.

3. Clarke, N. D., Kissinger, C. R., Desjarlais, J., Gilliland, G. L., and Pabo, C. O. (1994) Structural studies of the engrailed homeodomain. *Protein Sci. 3*, 1779-1787.

4. Shah, P. S., Hom, G. K., Ross, S. A., and Mayo, S. L. (2004) Thermodynamic and structural characterization of full sequence designs. *In preparation.*

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

6. Shah, P. S., Hom, G. K., and Mayo, S. L. (2004) Preprocessing of rotamers for protein design calculations. *J. Comput. Chem. 25*, 1797-1800.

7. Desmet, J., Spriet, J., and Lasters, I. (2002) Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Proteins 48*, 31-43.

8. Dahiyat, B. I., and Mayo, S. L. (1996) Protein design automation. *Protein Sci. 5*, 895-903.

9. Rohl, C. A., Chakrabartty, A., and Baldwin, R. L. (1996) Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci. 5*, 2623-2637.

10. Tanford, C., De, P. K., and Taggart, V. G. (1960) The role of the α-helix in the structure of proteins. Optical rotatory dispersion of β-lactoglobulin. *J. Am. Chem. Soc. 82*, 6028-6034.

11. Urnes, P., and Doty, P. (1961) Optical rotation and the conformation of polypeptides and proteins. In *Advances in Protein Chemistry*. (eds. C. B. Anfinsen Jr., M. L. Anson, K. Bailey, and J. T. Edsall), pp. 401-543. Academic Press, New York and London.

12. Iizuka, E., and Yang, J.T. (1965) Effect of salts and dioxane on the coiled conformation of poly-L-glutamic acid in aqueous solution. *Biochemistry 4*, 1249-1257.

13. Karle, I.L. (1992) Folding, aggregation and molecular recognition in peptides. *Acta Crystallogr B 48 (Pt 4),* 341-356.

14. Sigler, P.B., Jeffery, B.A., Matthews, B.W., and Blow, D.M. (1966) An x-ray diffraction study of inhibited derivatives of alpha-chymotrypsin. *J. Mol. Biol. 15,* 175-192.

15. Matthews, B.W., Sigler, P.B., Henderson, R., and Blow, D.M. (1967) Three-dimensional structure of tosyl-alpha-chymotrypsin. *Nature 214*, 652-656.

16. Gordon, D.B., Hom, G.K., Mayo, S.L., and Pierce, N.A. (2003) Exact rotamer optimization for protein design. *J. Comput. Chem. 24*, 232-243.

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

18. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods in Enzymology 276, Macromolecular Crystallography, Part A*, 307-326.

19. Holton, J., and Alber, T. (2004) Automated protein crystal structure determination using ELVES. *Proc. Natl. Acad. Sci. U S A 101*, 1537-1542.

20. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A 47 (Pt 2),* 110-119.

21. Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., and Dodson, E.J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr. 55 (Pt 1),* 247-255.

22. Collaborative Computational Project, N. (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr. 50*, 760-763.

23. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J. 78*, 1606-1619.