

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Analysis of Superfamily Catalytic Modules for Applications in Protein and Pathway Engineering

Permalink

<https://escholarship.org/uc/item/1046b4tp>

Author

Nagatani, Ray Jr. Anthony

Publication Date

2008-03-27

Peer reviewed|Thesis/dissertation

Analysis of Superfamily Catalytic Modules for Applications in Protein and Pathway
Engineering

by

Ray Anthony Nagatani Jr.

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Science and Pharmacogenomics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

UMI Number: 3297791

Copyright 2008 by
Nagatani, Ray Anthony, Jr.

All rights reserved.

UMI[®]

UMI Microform 3297791

Copyright 2008 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Copyright 2008
by
Ray Anthony Nagatani Jr.

For my family,

*And all others who have sacrificed so that I might
benefit*

ACKNOWLEDGEMENTS

Rather than an individual effort, I view the work contained within this thesis as the manifestation of a number of people's effort and guidance, of which I am the beneficiary. To begin, I owe a great deal to the instruction and mentorship of Dr. James Berger and Dr. Scott Classen of the University of California, Berkeley. In my undergraduate study, both men worked with great diligence to help me learn basic techniques in molecular biology that I have relied upon throughout my graduate studies. Most importantly, they helped shape my thinking. I doubt I believed in the need for experimental controls, sterile technique or multiple trials prior to research under their direction. Through their instruction, I learned well the value of doing something correctly once, rather than incorrectly multiple times.

Upon arriving at UCSF, I learned a great deal in my rotations with Dr. Alan Verkman, Dr. Francis Szoka and Dr. Brian Shoichet. From my first rotation with Alan, I quickly learned the value of both self-reliance and collaboration, and I appreciate that he showed me where the bar was set if I wanted to succeed in graduate school. In my rotation with Frank, I experienced the camaraderie and good will of his research group, which I think is a direct reflection of its leader. And working with Brian, I learned to appreciate hard and diligent work as it became my equalizer to address my (many) shortcomings as a scientist.

Over the course of my graduate career, I have been blessed to work with many fantastic scientists in various collaborative efforts. Although extremely busy with their own research groups, Dr. Brian Shoichet and Dr. Linda Brinen provided extraordinary mentorship in teaching me thermodynamics and crystallography for our collaboration in the work described in Chapter 2, published originally in the journal *Biochemistry*:

Nagatani RA, Gonzalez A, Shoichet BK, Brinen LS, Babbitt PC. "Stability for function trade-offs in the enolase superfamily "catalytic module". *Biochemistry*, 2007 Jun 12;46(23):6688-95

I have also benefited from the expertise and patience of my computational collaborators Dr. Tanja Kortemme, Dr. Matt Jacobson, Dr. Chakrapani Kalyanaraman, and Dan Mandell, who went to great lengths to provide computational guidance for the engineering of novel enzyme function. I regret that I returned so little on their investment, and apologize that I could not do more.

Outside of UCSF, I was also fortunate to have the opportunity to travel to the Weizmann Institute in Rehovot, Israel to work in the research group of Dr. Dan Tawfik. In the three months that I spent at the institute, I had several years' worth of memorable experiences in and out of lab. What I learned about protein engineering and the Middle East was roughly equivalent to the entire store of knowledge I had accumulated in several years of prior graduate study. But most importantly, Danny and his group were wonderful hosts to me as I adjusted to a very different environment far from home, and the work in Chapters 3 and 4 would not have been possible without their help. Toda raba.

In addition, I would like to acknowledge the help of Dr. John Gerlt and Ayano Sakai of the University of Illinois, Urbana-Champaign, whose work characterizing the kinetics of a promiscuous activity contributed to the work described in Chapter 4.

Of course, the most credit is due to my advisor, Patsy Babbitt. Many scientists who are faculty at UCSF are brilliant, and they can understand scientific phenomena that are quite complex and convoluted. However, I have never met a scientist who can understand both science and the human condition to the capacity that Patsy can. Among her many qualities (intelligence, eloquence, understanding, pragmatism), I think I benefited the most from her personable nature and optimism. While I frequently have felt that I would not be successful in my experiments, graduating, or any sort of future vocation, Patsy has always believed in good outcomes for both experiments and people. Faced with her relentless optimism and sound opinion, I had no choice but to gain self-confidence and to succeed.

I would be remiss if I were to not mention the support I received from my friends

while in graduate school. I would like to thank the Lim, Narlikar and McKerrow labs, as well as the SparkleMotion Biophysics Basketball Team for scientific and social support. I would also like to thank my colleagues in the Babbitt Lab, with special thanks to Dr. Walter Novak, Dr. Rey Banatao, Dr. Sunil Ojha, Grant Shackelford, and Michael Hicks for their company, instruction, and for being willing accomplices at great risk to their esteemed reputations.

Finally, I would like to thank my family for their unwavering support during a particularly self-centered time in my life. My parents were especially supportive, and I remember calling them several times during fits of self-doubt and self-pity. While they never claimed to understand exactly what challenges I was facing or what I was going through, they always reminded me that they believed in me no matter what I chose to do, and would be supportive of me even if it all fell apart.

Investigating Superfamily Catalytic Modules to Guide Protein and Pathway Engineering

Ray Anthony Nagatani Jr.

Abstract

Efforts in changing protein or pathway function would be greatly aided by design principles that inform how protein sequence and structure dictate function. To elucidate these design principles, we have characterized mechanistically diverse enzyme superfamilies; groups of enzymes that share a conserved architecture of catalytic residues (or “**catalytic module**”) that provide a conserved chemical capability used in each member’s overall function. The work described in this thesis focuses on (1) how to use information from enzymes superfamilies to guide protein and pathway engineering and (2) how to detect new protein or pathway functions using selections for function.

As protein engineering is frequently dependent on the stability of mutant variants, we also wanted to characterize the stability-for-function trade-off for conserved catalytic module residues in order to estimate the stability cost of exporting the module for novel function. Chapter 2 describes the functional, stability and structural characterization of alanine-substitution mutants of *o*-succinylbenzoate synthase (OSBS), a member of the enolase superfamily. In this work, we show that residues that are highly conserved for function across the superfamily also contribute the most to destabilization of the protein, in agreement with past work by other groups.

Although the design and engineering of novel enzyme variants or pathways depends on sound practice in the selection of starting templates or components, high-throughput methods for functional detection are also necessary in order to identify successful outcomes. To this end, Chapter 3 describes the characterization of a selection for glutamate racemase activity. In this work, we describe methods for maximizing recovery of weak variants (typical of initial engineering mutant variants) and describe how the enrichment and recovery of mutants changes as stringency is altered. This

selection was used as a foundation for the work in Chapter 4, in which superfamily functions that are conserved for pathway function in nature were paired to produce a specific output and develop a novel selection for function.

TABLE OF CONTENTS

PREFACE

Copyright	ii
Dedication	iii
Acknowledgements	iv
Abstract	vii
Table of Contents	ix
List of Tables	xi
List of Figures and Schemes	xii
CHAPTER 1. Challenges and Opportunities in Engineering Novel Function	1
Introduction	2
MECHANISTICALLY DIVERSE ENZYME SUPERFAMILIES:	
Functionally Robust Catalytic Modules	3
OPTIMIZING SENSITIVITY IN HIGH-THROUGHPUT:	
Characterizing a Selection for Function	4
FOLLOWING NATURE’S EXAMPLE:	
Linking Superfamily Function to Create a Selectable-Output Pathway	5
References	6
CHAPTER 2: Stability for Function Trade-offs in the Enolase Superfamily	
“Catalytic Module”	8
Abstract	10
Introduction	12
Materials and Methods	15
Results	19

Discussion	20
Conclusions	26
Acknowledgements	26
References	27
Figure Legends	34
CHAPTER 3: Optimizing Enrichment and Recovery for <i>in vivo</i> Enzyme Selections	38
Abstract	40
Introduction	41
Materials and Methods	43
Results	47
Discussion	50
Acknowledgements	53
References	57
CHAPTER 4: Following Nature's Example for Pathway Evolution	62
Abstract	63
Introduction	64
Methods and Materials	65
Results and Discussion	67
Conclusions	70
References	75

LIST OF TABLES

CHAPTER 2

Table 1. List of melting temp, ΔT_M , ΔH , and $\Delta\Delta G_u$ for OSBS wild-type and mutants.....	32
Table 2. Statistics of data collection and structural refinement	33

CHAPTER 3

Table I: Kinetic and Complementation Data for Wild-Type and Single Mutation Glutamate Racemases	54
Table II: Recovery and Enrichment for D10N/E152Q mutant	55
Table III: Enrichment and Recovery for Wild-type and D10N Mutant Glutamate Racemase	56

CHAPTER 4

Table I: Survival rates for AEEs from <i>E. coli</i> , <i>B. subtilis</i> , and BT1313, NAAAR from <i>Amycolatopsis</i> , and pASK 7C plasmid (empty vector)	73
Table II: Kinetic Constants for Hyrdolysis of L-Ala-D-Glu by <i>A. xylosoxydans</i> N-acetyl-D-Glutamate Amidohydrolase.....	74

LIST OF FIGURES AND SCHEMES

CHAPTER 2

Figure 1. Superposition of several enolase superfamily members and reaction scheme	35
Figure 2. Structural superposition of wild-type OSBS and E190A mutant	36
Figure 3. van't Hoff Plots	37

CHAPTER 3

Supplementary Information A. SDS-PAGE Protein gel of wild-type and D10N/E152Q mutant glutamate racemase	59
Supplementary Information B. Kinetic data for wild-type glutamate racemase ...	60
Supplementary Information C. Kinetic data for D10N/E152Q mutant glutamate racemase	61

CHAPTER 4

Figure 1. Diagram showing reaction pathways of paired enolase and amidohydrolase superfamily members to convert N-modified D-Phe to L-Phe and L-Ala-L-Glu to D-Glu	72
--	----

Chapter 1:

Challenges and Opportunities in Engineering Novel Function

INTRODUCTION:

Challenges and Opportunities in Engineering Novel Function

Within organisms, enzymes catalyze the numerous chemical reactions that support development, metabolism, reproduction and apoptosis. Although the existence of enzymes has been known since the 19th century, efforts to reprogram and utilize the functions of enzymes have only recently met with success. Using both computational design and combinatorial methods, enzymes have been engineered to do new chemical reactions (1, 2), or have been altered to recognize novel substrates (3-6). Using enzymes with desired functionalities, one might envision reprogramming microorganisms to carry out useful applications for biotechnology, biodefense or biodegradation.

However, despite the aforementioned successes, a number of challenges prevent the engineering of novel enzyme function from being facile. Firstly, ample evidence suggests that mutation to promote novel function impairs the stability of enzymes, limiting their evolvability (7-11). This instability may alter the ability of enzymes to adopt catalytically competent conformations, or it may hasten unfolding and proteolysis.

Additionally, because the mapping of function to structure is incomplete, the end goal for design is not always clear. Although experimentalists may pursue specific architectures in their design, these architectures may not be robust enough to deliver function in novel contexts. Moreover, actual functional variants may be lost due to inadequate sensitivity within functional detection methods. When combined with the mentioned challenges due to instability, the problem of enzyme design grows more difficult.

To address these challenges, functional robustness is necessary in both the enzymes that are produced by design and engineering as well as the detection methods used to identify successful variants. To address the former necessity, high conservation

of functional architectures in nature may suggest robustness, meaning that these architectures can be used as foundations for engineering new function. To address the latter necessity, the characterization of functional screening methods can suggest methodologies suited for projects demanding high-throughput and high-sensitivity.

MECHANISTICALLY DIVERSE ENZYME SUPERFAMILIES:

Functionally Robust Catalytic Modules

In mechanistically diverse enzymes superfamilies, hundreds of enzymes utilize a conserved chemical capability to facilitate their diverse overall reactions (12). Often, this conserved capability is tied to a highly conserved architecture of catalytic residues (or “catalytic module”). Many superfamilies have been characterized, facilitating the annotation of function from sequence information (13) or in conjunction with other methods (14).

Given that all members of superfamilies utilize the same catalytic module and chemical capability, members are often well suited for engineering. Within the enolase superfamily, members abstract protons alpha to carboxylic acids groups in order to facilitate overall reactions ranging from dehydration to cyclization (12). Within this superfamily, members have been shown to be promiscuous for different substrates (15) that are compatible with substrate specificity regions outside the catalytic module that are unique to individual members. Creating variation within these substrate specificity regions through mutation has been shown to introduce novel functions in individual members (2).

The strategy of mutating non-conserved portions of superfamily members has been used in other superfamilies to engineer novel functions. In the amidohydrolase superfamily, members have a catalytic module of histidine residues used in water-mediated hydrolysis. In this superfamily, Park *et. al* changed the non-conserved loops

surrounding the catalytic module of an enzyme that performs carbon-sulfur bond hydrolysis to engineer the ability to cleave carbon-nitrogen bonds (1). From these examples, the catalytic module of superfamilies appears robust for providing conserved functionalities in a variety of different contexts.

At a higher organizational level of conservation for specific function, superfamily members may also be paired in pathways to provide vital metabolic functions. Sakai *et al.* demonstrated that enolase and amidohydrolase superfamily members are often conserved in operons to convert modified D-amino acids to L-amino acids for cellular use (16). In this context, the robust functions of superfamily members are paired to enhance larger functional diversity, rather than through the engineering of individual enzyme functions.

As stability is a critical variable to optimize in engineering function, one may wonder whether catalytic modules are conserved for some inherent pre-organized stability or if they follow the general trend of destabilizing proteins in exchange for advantageous functionality. Chapter 2 addresses this question by examining the stability for function trade-offs in the enolase superfamily catalytic module.

OPTIMIZING SENSITIVITY IN HIGH-THROUGHPUT:

Characterizing a Selection for Function

Recent successes in the engineering of novel protein function have relied upon both computational and combinatorial methods to identify and test sequence variants based on good starting points. To identify active variants within these large libraries, high-throughput methods for functional detection are necessary to recover successful designs from successive rounds of optimization. One effective tool for high-throughput functional detection is to use selections for function, which link desired function to cell viability(17). Using selections for function, libraries ranging from 10^6 - 10^9 can be screened (18).

Although selections are very useful, their output is binary (life or death), meaning that it is difficult to identify stronger or weaker variants among enzymes that are active enough to complement. Although work has been done to optimize the dynamic range of selections to select for variants that are more efficient than wild-type enzymes (18-22), little has been done to optimize selections for the issue of optimizing the sensitivity of selections to identify very weak variants, a common feature of designed enzymes. To address this issue, Chapter 3 describes work done to characterize the sensitivity of a selection and to optimize its stringency to recover weakly active variants.

FOLLOWING NATURE'S EXAMPLE:

Linking Superfamily Function to Create a Selectable-Output Pathway

In addition to engineering function at the level of individual enzymes, desired overall functions can also be engineered via the organization of enzymes into pathways. For this purpose, the robust properties of superfamily enzymes can still be utilized as starting points for contributing function. As mentioned, examples in nature suggest that enolase and amiohydrolase superfamily members are conserved within the operons of several organisms to provide specific metabolic functions.

Following this example, the work in Chapter 4 used superfamily information to guide the engineering of a pathway that utilized specific enolase and amidohydrolase superfamily members to support a novel pathway function. By relying upon the robust functions of enolase and amidohydrolase superfamily members to provide the component functions, a selectable output was produced that relied upon a novel promiscuity discovered within one of the pathway components.

- (1) Park, H. S., Nam, S. H., Lee, J. K., Yoon, C. N., Mannervik, B., Benkovic, S. J., and Kim, H. S. (2006) Design and evolution of new catalytic activity with an existing protein scaffold. *Science* 311, 535-538.
- (2) Schmidt, D. M., Mundorff, E. C., Dojka, M., Bermudez, E., Ness, J. E., Govindarajan, S., Babbitt, P. C., Minshull, J., and Gerlt, J. A. (2003) Evolutionary potential of (beta/alpha)₈-barrels: functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry* 42, 8387-8393.
- (3) Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q. G. S., Roodveldt, C., and Tawfik, D. S. (2005) The 'evolvability' of promiscuous protein functions. *Nat Genet* 37, 73-76.
- (4) Varadarajan, N., Gam, J., Olsen, M. J., Georgiou, G., and Iverson, B. L. (2005) Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc Natl Acad Sci U S A* 102, 6855-6860.
- (5) Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J., and Stoddard, B. L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell* 10, 895-905.
- (6) Jurgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M., and Sterner, R. (2000) Directed evolution of a (beta alpha)₈-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A* 97, 9925-9930.
- (7) Wang, X., Minasov, G., and Shoichet, B. K. (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* 320, 85-95.
- (8) Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D. S. (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444, 929-932.
- (9) Bloom, J. D., Labthavikul, S. T., Otey, C. R., and Arnold, F. H. (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* 103, 5869-5874.
- (10) Bloom, J. D., Wilke, C. O., Arnold, F. H., and Adami, C. (2004) Stability and the evolvability of function in a model protein. *Biophys J* 86, 2758-2764.
- (11) Besenmatter, W., Kast, P., and Hilvert, D. (2007) Relative tolerance of mesostable and thermostable protein homologs to extensive mutation. *Proteins* 66, 500-506.
- (12) 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.
- (13) Pegg, S. C., Brown, S. D., Ojha, S., Seffernick, J., Meng, E. C., Morris, J. H., Chang, P. J., Huang, C. C., Ferrin, T. E., and Babbitt, P. C. (2006) Leveraging enzyme structure-function relationships for functional inference and experimental design: the structure-function linkage database. *Biochemistry* 45, 2545-2555.
- (14) Song, L., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Brown, S., Imker, H. J., Babbitt, P. C., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2007) Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat Chem Biol* 3, 486-491.
- (15) Taylor Ringia, E. A., Garrett, J. B., Thoden, J. B., Holden, H. M., Rayment, I.,

- and Gerlt, J. A. (2004) Evolution of enzymatic activity in the enolase superfamily: functional studies of the promiscuous o-succinylbenzoate synthase from *Amycolatopsis*. *Biochemistry* 43, 224-229.
- (16) Sakai, A., Xiang, D. F., Xu, C., Song, L., Yew, W. S., Raushel, F. M., and Gerlt, J. A. (2006) Evolution of enzymatic activities in the enolase superfamily: N-succinylamino acid racemase and a new pathway for the irreversible conversion of D- to L-amino acids. *Biochemistry* 45, 4455-4462.
- (17) Taylor, S. V., Kast, P., and Hilvert, D. (2001) Investigating and Engineering Enzymes by Genetic Selection. *Angew Chem Int Ed Engl* 40, 3310-3335.
- (18) van Sint Fiet, S., van Beilen, J. B., and Witholt, B. (2006) Selection of biocatalysts for chemical synthesis. *Proc Natl Acad Sci U S A* 103, 1693-1698.
- (19) Kleeb, A. C., Edalat, M. H., Gamper, M., Haugstetter, J., Giger, L., Neuenschwander, M., Kast, P., and Hilvert, D. (2007) Metabolic engineering of a genetic selection system with tunable stringency. *Proc Natl Acad Sci U S A* 104, 13907-13912.
- (20) Lin, H., Tao, H., and Cornish, V. W. (2004) Directed evolution of a glycosynthase via chemical complementation. *J Am Chem Soc* 126, 15051-15059.
- (21) Bernath, K., Magdassi, S., and Tawfik, D. S. (2005) Directed evolution of protein inhibitors of DNA-nucleases by in vitro compartmentalization (IVC) and nano-droplet delivery. *J Mol Biol* 345, 1015-1026.
- (22) Neuenschwander, M., Butz, M., Heintz, C., Kast, P., and Hilvert, D. (2007) A simple selection strategy for evolving highly efficient enzymes. *Nat Biotechnol* 25, 1145-1147.

Chapter 2:

Stability for Function Trade-offs in the
Enolase Superfamily “Catalytic Module”

Reproduced with permission from

Nagatani RA, Gonzalez A, Shoichet BK, Brinen LS, Babbitt PC. “Stability for function trade-offs in the enolase superfamily ‘catalytic module’.” *Biochemistry*. 2007 Jun 12;46(23):6688-95. Epub 2007 May 16.

Copyright 2007 American Chemical Society.

ABSTRACT

Enzyme catalysis reflects a dynamic interplay between charged and polar active site residues that facilitate function, stabilize transition states and maintain overall protein stability. Previous studies show that substituting neutral for charged residues in the active site often significantly stabilizes a protein, suggesting a stability trade-off for functionality. In the enolase superfamily, a set of conserved active site residues (the “catalytic module”) has repeatedly been used in nature to evolve many different enzymes to perform unique overall reactions involving a chemically diverse set of substrates. This catalytic module provides a robust solution for catalysis that delivers the common underlying partial reaction that supports all of the different overall chemical reactions of the superfamily. As this module has been so broadly conserved in the evolution of new functions, we sought to investigate the extent to which it follows the stability-function trade-off. Alanine substitutions were made for individual residues, groups of residues, and the entire catalytic module of *o*-succinylbenzoate synthase (OSBS), an enolase superfamily member from *Escherichia coli*. Of six individual residue substitutions, four (K131A, D161A, E190A, and D213A) substantially increased protein stability (by 0.46 to 4.23 kcal•mol⁻¹), broadly consistent with prediction of a stability-activity trade-off. The residue most conserved across the superfamily, E190, is by far the most destabilizing. When the individual substitutions were combined into groups, (as they are structurally and functionally organized) non-additive stability effects emerged, supporting previous observations that residues within the module interact as two functional groups within a larger catalytic system. Thus, whereas the multiple mutant enzymes D161A/E190A/D213A and K131A/K133A/D161A/E190A/D213A/K235A (referred to as “3KDED”) are stabilized relative to the wild-type enzyme (by 1.77 and 3.68 kcal•mol⁻¹, respectively) the net stabilization achieved in both cases is much less than would be predicted if their stability contributions were additive. Organization of the catalytic module into systems that mitigate the expected stability cost due to the presence of highly charged active site residues may help to explain its repeated use for the evolution

of many different functions.

KEYWORDS- Enolase superfamily, o-succinylbenzoate synthase, stability, stability-function trade-off

INTRODUCTION

The catalytic function of enzymes often depends upon a specific architecture of charged or polar residues in close proximity to each other. By adopting pre-organized geometries that are energetically unfavorable, these residues are able to orient substrates through binding interactions, stabilize high-energy transition states and provide the flexibility necessary for substrate and product turnover (1, 2). Thus, enzyme catalysis reflects a dynamic interplay between protein stability and protein function. While these high-energy conformations of active site residues often destabilize the global stability of the enzyme through electrostatic or steric repulsion, results from several investigations have shown that this interplay is an essential element of enzyme catalysis. Experiments with the enzyme lysozyme showed that substitution of charged catalytic residues with neutral non-polar residues increases stability, though catalytic function is lost (3). Antibiotic resistant mutants of beta-lactamase that have evolved in nature are destabilized relative to the wild-type enzyme, yet have a selective advantage because of their increased activity on cephalosporin substrates (4). Substituting residues in either the positively charged active site of the enzyme barnase or the negatively charged binding site of its inhibitor barstar has been shown to improve stability at the expense of reduced activity, partly due to reducing electrostatic repulsion between catalytic residues (5, 6).

The stability-activity tradeoff has consequences in the molecular evolution of enzyme structures (7-9). In part to compensate for destabilization incurred from charged catalytic residues or local structural variations between divergent enzymes, systems of catalytic residues that can mitigate the stability cost may be conserved in enzyme families. For example, serine proteases use a conserved catalytic triad of residues to hydrolyze peptide amide bonds (10). The geometry of this catalytic system is conserved even across different protein folds and is the foundation for the catalytic mechanism, with specificity for individual substrates conferred by unique binding elements that

distinguish each enzyme. Similarly, enzymes within mechanistically-diverse enzyme superfamilies (11) use highly conserved active site architectures (or “catalytic modules”) that have retained the ability to catalyze a fundamental partial reaction (or other chemical capability) that leads to the stabilization of a common type of intermediate in all of the divergent enzymes in the superfamily.

Residues within such catalytic modules generally maintain a conserved geometry in many different structural contexts (i.e., divergent enzymes in a family, different enzyme folds) despite the significant predicted destabilization caused by their inclusion (12-14). Does the catalytic module design offer a robust solution for pre-organizing charged residues in a conformation that mitigates the stability cost for providing function? To answer this question, we have investigated the stability-function trade-off within the context of catalytic module residues that are conserved across all members of the mechanistically diverse enolase superfamily.

The enolase superfamily is a well-characterized superfamily of hundreds of enzymes that catalyze a chemically diverse range of different overall reactions (15). To date, fourteen reactions performed by superfamily members have been characterized, and many more candidates have been identified that are likely to perform additional functions. Although these overall reactions range from dehydration to cycloisomerization, all members use a catalytic module of oppositely charged residues to abstract a proton from a common substrate substructure and stabilize the resulting enediolate high-energy intermediate (16) (Figure 1C). In the catalytic module, the conserved residues can be viewed as forming two systems that facilitate catalytic function: one system is involved in substrate orientation and transition state stabilization. The other system abstracts the substrate proton to form the transition state. The first system is comprised of three acidic residues that coordinate a divalent metal ion that plays a role in orienting the carboxylic acid substructure common to all superfamily ligands. These residues and the metal ion are present in all members of the superfamily and exhibit few substitutions across the

divergent set. The negatively charged residues of this system are sandwiched between two sets of weakly basic residues that comprise the second system (Fig 1A). In this system, the weakly basic residues abstract a proton alpha to the conserved carboxylic acid substructure, forming an enolate high-energy intermediate that is stabilized through interaction with the metal ion of the first system. This second system shows more structural and chemical variation across the superfamily than the first, distinguishing at least three different subgroups within the superfamily (17).

Electrostatic calculations have suggested that the juxtaposition of these charged residues in the enolase superfamily catalytic module leads to significant electrostatic destabilization (13, 14). Based on these conclusions, it is somewhat surprising that the geometry and chemical characteristics of this catalytic module have been used repeatedly by nature in the evolution of different functions, especially since several other architectures that perform similar proton-abstraction chemistry have been observed in other biological systems (18, 19). Indeed, even other superfamilies from different fold classes unrelated to the enolase superfamily have utilized similar catalytic architectures and strategies to perform this same fundamental partial reaction (20, 21). The widespread conservation of the enolase superfamily catalytic module across the superfamily (and in other superfamilies) suggests instead that the module is a robust solution that has been commonly used by evolution to obtain catalytic functionality at an acceptable cost in destabilization.

To investigate quantitatively the structure-function tradeoff for the catalytic module of the enolase superfamily, we performed alanine substitution of the component residues of *o*-succinylbenzoate synthase (OSBS) from *Escherichia coli*. The OSBS from *E. coli* catalyzes the dehydration of 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) to *o*-succinylbenzoate (OSB) (Figure 1C), a reaction that is essential for anaerobic respiration (22). The OSBS from *E. coli* is a monomeric protein for which a wild-type structure is available (23). It is a useful system for addressing

stability-function trade-offs in the enolase superfamily because its overall reaction mainly relies upon the proton abstraction step provided by the catalytic module to proceed to completion (Figure 1C), without the complications introduced by additional partial reactions required to perform the different overall reactions catalyzed by other members of the superfamily. Here, we investigate the thermostability of wild-type OSBS, and of alanine-substituted mutants of the catalytic module residues, both as point mutants and as combinations that cover the entire catalytic module and the two systems of catalytic residues within the module. To gain additional insight regarding the structural mechanisms of stabilization, we determined the crystal structure of the most thermostable mutant.

This is the first analysis of stability-function trade-offs applied to an entire active site system conserved across hundreds of different enzymes. The results provide information about the robustness of such systems for natural enzyme evolution and for engineering new catalysts *in vitro*.

Materials and Methods

Site-Directed Mutagenesis

The 0.96 kb *menC* gene from *E. coli* MG1655, cloned into a pET15b expression vector (Novagen, San Diego, CA), a generous gift from Dr. John Gerlt, was used for all site-directed mutagenesis experiments. The QuickChange site-directed mutagenesis method (Stratagene, La Jolla, CA) was used to introduce point mutations. Each mutation was confirmed by DNA sequencing.

Biosynthesis of SHCHC

SHCHC was synthesized as described previously from chorismic acid (24), using the coupled actions of isochorismate mutase and SHCHC synthase. The compound was then purified as described previously (25, 26).

Spectrophotometric Assay

OSBS activity was assayed at pH 8 as previously described (25) using 250 nM

SHCHC in an assay buffer containing 50 mM Tris-HCL, 5 mM MgCl₂, pH 8.

Isolation of wild-type and alanine-mutant OSBS from Escherichia coli.

E. coli BL21 (DE3) cells were transformed with the plasmid and grown at 37 °C in LB medium for 18 hours without induction. Cells were resuspended in 30 mLs of binding buffer (10 mM Tris-HCl, pH 8 containing 5 mM imidazole, 0.5 M NaCl and 5 mM MgCl₂), then disrupted by sonication. Following centrifugation, the supernatant was loaded onto a HisTrap 5 mL HP column (GE Biosciences, Piscataway, NJ) that had been equilibrated with binding buffer. Following a wash step with 5 column volumes of buffer, the OSBS wild-type and mutants were eluted using a linear imidazole gradient (0 to 1 M imidazole in 10 mM Tris-HCl, pH 8, containing 5 mM MgCl₂ and 0.5 M NaCl). Fractions containing OSBS wild-type or mutant proteins were then dialyzed against 1 L of phosphate buffered saline. The His-tag was removed by incubation with 1 U thrombin/mg OSBS for 72 hours at 4 °C. The dialyzed protein was then loaded onto a DEAE anion exchange column (GE Biosciences, Piscataway, NJ) and eluted with a linear NaCl gradient (0 to 0.5 M NaCl in 10 mM Tris-HCl, pH 8.0 containing 5 mM MgCl₂). Fractions containing homogenous OSBS wild-type or mutants were pooled and concentrated using an Amicon 30 KDa MW-cutoff centrifugal filter. The purity of the samples was verified using SDS-PAGE, which showed a single band (data not shown).

Enzyme Stability analysis by Circular Dichroism (CD)

CD experiments were conducted using a Jasco J-715 spectropolarimeter with a Jasco PCT-348WI peltier-effect temperature controller. Quartz cells with a 1 cm path length (Hellma, Inc., Plainview, NY) were used for all measurements. Temperature was monitored and controlled using an in-cell thermometer and stir bar.

Thermal denaturation of OSBS wild-type and all mutant proteins was carried out in a 50 mM KPO₄, 200 mM KCl, 20% ethylene glycol, pH 8 buffer. For the experiment monitoring the tertiary structure, a buffer containing 50 mM KPi, 200 mM KCl, 45% ethylene glycol, pH 8.5 was used. Denaturation was monitored between 293 K and 343

K, with a heating rate of 2 degrees/min.

All proteins were monitored for helical content by CD in the far-UV region (223 nm) at an enzyme concentration of 0.02 mg/mL. The wild-type protein was also monitored for tertiary structure by CD in the near-UV region (285 nm) at an enzyme concentration of 0.1 mg/mL.

Thermal denaturation was analyzed using the program EXAM (27) to calculate all melting temperatures (T_M) and van't Hoff enthalpy of unfolding (ΔH_{VH}) values. Melting curve baselines were fixed based on initial predictions made by EXAM. With the exception of mutants E190A and 3KDED, the experimentally derived change in heat capacity (ΔCp) for wild-type OSBS (3.2 kcal/mol•K) was used as an input parameter for the curve-fitting program in EXAM. For E190A and 3KDED, each mutant's experimentally derived ΔCp was used as an input parameter. Each melting temperature reported is the average of at least three independent experiments.

The change in free energy of unfolding ($\Delta\Delta G_u$) was calculated using the method of Schellman (28), where $\Delta\Delta G_u = \Delta T_M \Delta S_{u,WT}$. For the E190A and 3KDED mutants, the change in free energy of unfolding was also calculated using the equation $\Delta G = \Delta H_{wt} - T\Delta S_{wt} + \Delta Cp (T - T_{wt} - T \ln(T/T_{wt}))$, which takes differences in ΔCp into account. Increases in melting temperature indicate increased stability, resulting in a positive $\Delta\Delta G_u$. The ΔCp of the wild-type and selected OSBS mutant proteins was determined from van't Hoff plots in the presence of varying amounts of freshly prepared urea as described below.

Calculation of ΔCp

Wild-type OSBS and mutants that were highly stabilized (E190A and 3KDED) were reversibly denatured by temperature in a 50 mM KPi, 200 mM KCl, 20% ethylene glycol, pH 8 buffer with ultra-pure urea (Fluka, St. Louis, MO) added to a total concentration of up to 2 M. Samples in each buffer were then analyzed using far-UV CD. Thermal denaturation was analyzed using the program EXAM to calculate all melting temperatures (T_M) and van't Hoff enthalpy of unfolding (ΔH_{VH}) values. Melting curve

baselines were fixed based on initial predictions made by EXAM. Because the ΔC_p was unknown prior to the experiment, an estimated ΔC_p value of 1.7 kcal/(mol•K) was used as an input parameter for the curve-fitting program in EXAM. This parameter value was found to consistently provide robust curve-fitting of the data across the different denaturant concentrations. The average T_M of each protein was plotted against the average ΔH_{VH} at each urea concentration to generate a van't Hoff plot, where ΔC_p is represented by the slope (Figure 3). For each van't Hoff plot, every point reflects at least three melts in each urea concentration.

Crystal growth and structure determination

Crystals of the E190A mutant were grown by vapor diffusion in sitting drops equilibrated over 0.1 M MES (pH 5.8), 75 mM Sodium Molybdate and 25% (v/v) PEG 4000. The initial concentration of protein in the drop was 6 mg/mL. Crystals appeared over a few days after equilibration at 18 °C. Before data collection, crystals were immersed in a cryoprotectant solution of 20% (v/v) ethylene glycol, 0.1 M MES (pH 5.8), 75 mM Sodium Molybdate and 25% (v/v) PEG 4000.

Diffraction data were collected on cryo-cooled crystals on BL 9-1 at the Stanford Synchrotron Radiation Laboratory. The data set was measured from a single crystal. Reflections were indexed, integrated and scaled using MOSFLM (29)(29) and SCALA (30). For the E190A structure, the space group was $P2_1$; the unit cell was refined to $a=69.47 \text{ \AA}$, $b=77.39 \text{ \AA}$, $c=110.68 \text{ \AA}$, $\beta=90.08^\circ$. The data set statistics are summarized in Table 2.

The structure was determined by the Molecular Replacement method, using the program MOLREP (30, 31). Four molecules were located in the asymmetric unit. Refinement of the structure was carried out with REFMAC5 (30, 32). Independent rigid body refinement of each of the four molecules was followed by alternating cycles of restrained refinement and water molecular search using ARP/wARP (33). The model was inspected and rebuilt with the program COOT (34). TLS refinement (35) was carried out

at the end of the refinement process.

Results

Effects of alanine-substitution on activity

All alanine-substituted OSBS mutants were assayed for the ability to catalyze the dehydration of SHCHC by monitoring the loss of substrate. None of the mutants displayed detectable activity.

Reversible Denaturation and Two-State Behavior of wild-type OSBS

OSBS wild-type enzyme was reversibly denatured by temperature and analyzed by far-UV and near-UV CD. The loss of helical content and tertiary structure was fit to the data using a two-state analysis program (27) and each method gave a similar T_M of 50.8 °C (Table 1) and 52.7 °C, respectively (data not shown). Upon cooling of the sample, 80-90% of the original CD signal (for completely folded enzyme) was recovered, and refolding curves showed a transition temperature that was similar to the melting temperature (data not shown). In addition, the T_M was found to be robust and consistent among different rates of sample heating. The ΔH_{VH} of the wild-type enzyme was calculated to be 86.1 kcal, which is consistent with other reported values for $(\beta/\alpha)_8$ -barrel proteins (36).

Thermal Denaturation of OSBS mutants

OSBS alanine-substituted mutants were reversibly denatured by temperature and analyzed by far-UV CD. The loss of helical content was fit as described for the wild-type enzyme. Upon cooling of the sample, 80-90% of the original CD signal was recovered. The calculated T_M , $\Delta\Delta G_u$ and ΔH_{VH} of each alanine-substituted mutant are listed in Table 1. Although most substitutions either caused a negligible change in melting temperature or stabilized the protein, the K235A mutant exhibited a lower melting temperature, indicating that this substitution is destabilizing.

Determination of ΔC_p for OSBS wild-type, E190A and 3KDED mutants

The non-additive stabilization effects in the 3KDED mutant and the magnitude

of difference in melting temperature between the E190A and 3KDED mutants and the wild-type enzyme raise the question whether the thermostability changes we observed are consistent over a wide temperature range. To address this question, the ΔC_p was determined for wild-type OSBS, the E190A mutant, and the 3KDED mutant by performing reversible thermal denaturations in the presence of varying concentrations of urea (Figure 3). The calculated ΔC_p of wild-type OSBS is 3.2 kcal/(mol•K), in accord with the predicted ΔC_p for a protein of this size (37). In comparison, the calculated ΔC_p of E190A mutant was 3.4 kcal/(mol•K). The calculated ΔC_p for the 3KDED mutant was 3.2 kcal/(mol•K).

X-ray Crystallographic Structure Determination

The E190A mutant structure refinement statistics are given in Table 2. The model is not complete, with several side chains and two external loops (residues 13-24 and 111-127) showing very poor or no electron density at all, presumably due to disorder. No bound metal was observed in the active site. The mutant site was clearly visible in the map, with density truncated at the C β atom and backbone stereochemistry consistent with alanine rather than glutamate in that position. An attempt to fit the *apo* model in the mutant map resulted in strong negative density around the E190 side chain and distorted torsion angles for this residue (to avoid clashes with neighboring glutamate side chains). Nevertheless, the structure superimposed well with the wild-type *apo* structure, retaining the overall (β/α)₈-barrel fold with an overall RMSD of 0.55 Å over 292 α -carbon atom pairs. In this superposition, no significant distortion of the other residues in the catalytic module were observed.

Discussion

Stabilized mutants

As predicted by theory, the majority of the alanine substitutions in the OSBS catalytic module resulted in proteins with higher melting temperature. Many of these melting temperature changes were similar in magnitude to those reported in other studies

of function/stability trade-offs. Shoichet et al. report increases in melting temperature of up to 5.1 degrees ($\Delta\Delta G_u$ of 2.0 kcal•mol⁻¹) upon substitution of single catalytic residues in lysozyme (3). Kanzaki et. al report an increased melting temperature of 4 degrees upon substituting a phenylalanine for a catalytic glutamic acid in the core of tryptophan synthase, another (β/α)₈-barrel protein (36). Point mutants K131A, D161A, and D213A all show increases in melting temperature and $\Delta\Delta G_u$ within this range (Table 1).

The circular dichroism experiments were performed in the absence of metal added to the buffer. The results suggest that the presence of metal was unlikely as circular dichroism experiments with OSBS in the presence of metal led to irreversible denaturation (data not shown). Although metal cofactors are required for catalysis by members of the enolase superfamily, substantial evidence from calorimetry experiments and structural characterization of other members of the superfamily suggest that these enzymes are stable and show little distortion of their active sites in the absence of metal (38-40). Similarly, as noted below in the comparison between a structure of wild-type OSBS (solved with metal present) and our structure of the E190A mutant, absence of metal does not have a significant effect on the geometry of the active site or the overall structure.

Notably, the point mutant E190A has a melting temperature that is 15.9 °C degrees higher than wild-type OSBS. This stabilization ($\Delta\Delta G_u = 4.23$ kcal•mol⁻¹) is far greater than that seen in other alanine-substituted OSBS mutants, and is also extremely high relative to previously reported single-mutant values in other systems. Other increases in melting temperature of comparable magnitude have required multiple mutations. For example, Pantoliano *et al.* report an increase of 14.3 °C ($\Delta\Delta G_u = 3.80$ kcal•mol⁻¹) for a six-mutation subtilisin mutant (38). Additionally, Matsumura *et al.* describe a lysozyme mutant that required three engineered disulfide bridges to increase the melting temperature by 23.4 °C (39).

Inspection of the wild-type OSBS structure suggests that this stabilization may

result from the relief of electrostatic repulsion. E190 is located between D161 and D213, the two other negatively charged metal-coordinating residues, and presumably causes electrostatic repulsion with both (Figure 2). While negatively charged residues D161 and D213 are close to the positively charged residues K133 and K235 respectively (Figure 2), E190 is not close to any positively charged residues that could mitigate this effect through stabilizing interactions. In addition, analysis of the wild-type OSBS structure suggests that the side chains of E190 and D213 clash in several rotameric conformations. Given that the mutation to alanine truncates the side chain, these clashes are not predicted to occur in the E190A mutant structure. Thus, it is not surprising that removing the uncompensated negative charge from E190 greatly stabilizes the protein.

Analysis of the E190A crystal structure suggests that the substitution to alanine relieves some strain in the active site, as the carboxylic side chain of D213 has shifted 0.5 Å closer to the space once occupied by E190 in the wild-type structure (Figure 2). However, it is noteworthy that there is otherwise little structural change in the overall architecture of the active site between the wild-type and stabilized mutant, suggesting that the structure of the barrel is robust against destabilization.

If stability and function trade-offs in OSBS were directly correlated, one may wonder whether E190 is the most catalytically important residue since its mutation to alanine makes it the most stable of the mutants we analyzed. We also note that E190 is the single most highly conserved residue in the enolase superfamily, highlighting its contribution to the catalytic module (40). Although these experiments cannot determine whether E190 is the “most” catalytically important residue for function in OSBS (as all residues are necessary for function), its obligate conservation suggests that it may play a role in the module that is the least accommodating of change across the members of the superfamily. Viewing our results for OSBS within the context of the superfamily suggests a relationship between global enzyme destabilization and functional necessity associated with this catalytic module residue in all members of the superfamily.

Analysis of mutants in which groups of catalytic module residues are removed (representing the two catalytic subsystems described above) show that the substitution of additional catalytic residues does not increase stability in an additive manner. This is not surprising, as only substituting residues that are not interacting with each other would likely lead to additive effects (41). In sum, this observation supports prior reports that the catalytic module residues interact as two systems with non-additive effects on stability.

To obtain an estimate of the impact of the entire catalytic module on global enzyme stability, all six residues of the module were mutated to alanine in the mutant, 3KDED. 3KDED melts at a temperature that is far greater than that of the wild-type enzyme (Table 1). The stability of this catalytically “empty” $(\beta/\alpha)_8$ -barrel mutant is similar to that of a computationally designed active site-less $(\beta/\alpha)_8$ -barrel protein which was designed to fold stably (but without a function) (64.6 °C for the 3KDED mutant as compared to 65 °C for the designed protein) (42). This comparison between the similar catalytically “empty” $(\beta/\alpha)_8$ -barrels and the wild-type OSBS suggests the stability cost of the catalytic module.

Notably, the temperature stabilization of the 3KDED mutant is less than that of the E190A mutant (ΔT_M 's of 13.8 °C vs. 15.9 °C and $\Delta \Delta G_u$'s of 3.68 and 4.23 kcal•mol⁻¹ relative to wild-type OSBS, respectively). As removing other components of the system besides E190 does not yield an additive gain in stability, this may indicate that the additional components of the system do not incur a greater cost in destabilization than that incurred by the presence of E190 alone. Thus, these results suggest one possible reason why the catalytic module has been so widely used in the evolution of new function; the organization of the module perhaps supports the insertion of the several functionally important residues conserved in the active site without concomitant additive losses in stability. Through such a strategy, variations in the active site can be achieved while retaining a minimum level of stability necessary for folding or to accommodate additional functional mutations that may be required for specific reactions in the

superfamily.

Destabilized mutants

Although the majority of the alanine-substitution mutations resulted in proteins with higher melting temperatures, some mutations actually destabilize the protein. Most notably, both mutants that contain a lysine to alanine mutation at position 235 (K235A, K131A/K133A/K235A) are less stable than the wild-type enzyme. Interestingly, when paired with a stabilized alanine substitution mutant (K131A/K133A, $\Delta T_M = 3.6$ degrees, $\Delta\Delta G_u = 0.95$ kcal•mol⁻¹), the K235A mutation dominates the stability profile, leading to a mutant that is destabilized relative to the wild-type ($\Delta\Delta G_u = -0.77$ kcal•mol⁻¹).

K235 has been deemed critical for proton abstraction in OSBS because mutation to alanine, serine or arginine obviates both exchange of the alpha proton with solvent deuterium and the subsequent dehydration step (23). While mechanistic roles for structural homologs of K235 have been characterized in some superfamily members (43, 44), the role of this residue has not been conclusively determined in the catalytic mechanism of OSBS (23). Because K235 is a charged residue in the enzyme core (13), we would predict that it would be destabilizing and is conserved because of its contribution to function. Thus, it is surprising to find that removal of the residue's charge (through mutation to alanine) is destabilizing.

Inspection of the crystal structure of wild-type OSBS suggests several reasons why the K235A mutant may be destabilized. Although past work by Stites *et al.* has indicated that the methylene groups in the lysine side chain can pack within a hydrophobic core, this packing is dependent on the lysine being deprotonated (45). Given that continuum electrostatics methods have shown that the structurally analogous lysine in related family members has a high upward pKa shift (13), it is unlikely that K235 is deprotonated and able to make significant packing interactions in OSBS. Instead, the destabilization is likely due to the removal of a salt bridge between the positively charged K235 and a conserved aspartate residue at position 213 (D213) (the distance between the

side chains of K235 and D213 is approximately 3 Å). Upon removal of the stabilizing salt bridge in the K235A mutant, residue D213 may then shift closer to its metal-coordinating partner glutamate 190 (E190). This shift may cause all three acidic metal coordinating residues to be in closer proximity, thereby destabilizing the global structure (Fig 2).

In our experimental system, these observations suggest that the conservation of K235 contributes to global stability in the context of the catalytic module, thus rationalizing its requirement for functional activity and suggesting an additional dimension to the relationship between structure and function in this catalytic module; some residues may be catalytically essential because they stabilize other catalytic residues.

Applications for protein design and engineering

These results have potentially important implications with regard to protein design and engineering. Since protein stability is predicted to be a limiting factor in the evolution of novel function (7, 9, 46), both the E190A and 3KDED mutants could potentially be used as platforms for protein engineering in $(\beta/\alpha)_8$ -barrels. The overall stability and higher melting temperature of the 3KDED mutant gives a foundation for estimating the upper limit of the underlying stability of an $(\beta/\alpha)_8$ -barrel fold after an active site is removed. This limit agrees well with that defined for an $(\beta/\alpha)_8$ -barrel designed without an active site (42). This suggests that the biophysical measurements derived from comparing the stabilities of the “empty” 3KDED protein and the functional wild-type OSBS may be useful in predicting the level of destabilization that can be tolerated by an “empty” (no active site) $(\beta/\alpha)_8$ -barrel protein. Additionally, the crystal structure of the E190A mutant verifies that removal of the most destabilizing element does not change the architecture around the active site, indicating that perhaps other catalytic modules could be placed into the robust OSBS framework to engineer new function. In sum, this study suggests that the OSBS scaffold may provide a useful and general system for the design of new active sites within $(\beta/\alpha)_8$ -barrel protein scaffolds.

Conclusions

Investigation of the single and combinatorial contributions of the enolase superfamily “catalytic module” residues to structural stability and enzymatic function extends previous observations of the trade-offs between enzyme stability and function. Principally, this allows us to view the problem within the context of a critical set of residues required for catalysis among all members of a large and mechanistically diverse superfamily. Our results show a non-additive increase in stability upon altering the components of the catalytic module which supports prior work describing the function of the module as an interacting system (17). Within this system, we have identified the components that make the greatest contributions to destabilization, as well as functionally essential residues that actually stabilize the global structure, suggesting refinement to simple theories for explaining the stability/function trade off in enzyme active sites. To the extent that our results for the *E. coli* OSBS can be generalized to the entire enolase superfamily, we can propose a new explanation for nature’s repeated use of this catalytic module in the evolution of a large number of different enzymatic reactions.

ACKNOWLEDGMENTS- R.N. would like to thank Clement Chu, Christian Cunningham, Chris Waddling, David Savage, Kerim Babaoglu, Matthew Good, Pascal Egea, Sam Pfaff and Yu Chen for assistance and advice pertaining to x-ray crystallography. The authors would also like to thank Dr. John Gerlt for his insightful suggestions.

REFERENCES

- (1) Beadle, B. M., and Shoichet, B. K. (2002) Structural bases of stability-function tradeoffs in enzymes. *J Mol Biol* 321, 285-296.
- (2) Benkovic, S. J., and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis. *Science* 301, 1196-1202.
- (3) 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 U S A* 92, 452-456.
- (4) Wang, X., Minasov, G., and Shoichet, B. K. (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* 320, 85-95.
- (5) 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.
- (6) Schreiber, G., Buckle, A. M., and Fersht, A. R. (1994) Stability and function: two constraints in the evolution of barstar and other proteins. *Structure* 2, 945-951.
- (7) Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D. S. (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444, 929-932.
- (8) Besenmatter, W., Kast, P., and Hilvert, D. (2004) New enzymes from combinatorial library modules. *Methods Enzymol* 388, 91-102.
- (9) Bloom, J. D., Labthavikul, S. T., Otey, C. R., and Arnold, F. H. (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* 103, 5869-5874.
- (10) Dodson, G., and Wlodawer, A. (1998) Catalytic triads and their relatives. *Trends Biochem Sci* 23, 347-352.
- (11) Gerlt, J. A., and Babbitt, P. C. (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem* 70, 209-246.

- (12) Elcock, A. H. (2001) Prediction of functionally important residues based solely on the computed energetics of protein structure. *J Mol Biol* 312, 885-896.
- (13) Livesay, D. R., Jambeck, P., Rojnuckarin, A., and Subramaniam, S. (2003) Conservation of electrostatic properties within enzyme families and superfamilies. *Biochemistry* 42, 3464-3473.
- (14) Livesay, D. R., and La, D. (2005) The evolutionary origins and catalytic importance of conserved electrostatic networks within TIM-barrel proteins. *Protein Sci* 14, 1158-1170.
- (15) 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.
- (16) Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, R., Ringe, D., Petsko, G. A., Kenyon, G. L., and Gerlt, J. A. (1995) A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids. *Science* 267, 1159-1161.
- (17) Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids. *Biochemistry* 35, 16489-16501.
- (18) Esaki, N., and Walsh, C. T. (1986) Biosynthetic alanine racemase of *Salmonella typhimurium*: purification and characterization of the enzyme encoded by the *alr* gene. *Biochemistry* 25, 3261-3267.
- (19) Glavas, S., and Tanner, M. E. (1999) Catalytic acid/base residues of glutamate racemase. *Biochemistry* 38, 4106-4113.
- (20) Herron, S. R., Benen, J. A., Scavetta, R. D., Visser, J., and Journak, F. (2000) Structure and function of pectic enzymes: virulence factors of plant pathogens. *Proc Natl Acad Sci U S A* 97, 8762-8769.

- (21) Scavetta, R. D., Herron, S. R., Hotchkiss, A. T., Kita, N., Keen, N. T., Benen, J. A., Kester, H. C., Visser, J., and Jurnak, F. (1999) Structure of a plant cell wall fragment complexed to pectate lyase C. *Plant Cell* 11, 1081-1092.
- (22) Sharma, V., Meganathan, R., and Hudspeth, M. E. (1993) Menaquinone (vitamin K2) biosynthesis: cloning, nucleotide sequence, and expression of the menC gene from *Escherichia coli*. *J Bacteriol* 175, 4917-4921.
- (23) Klenchin, V. A., Taylor Ringia, E. A., Gerlt, J. A., and Rayment, I. (2003) Evolution of enzymatic activity in the enolase superfamily: structural and mutagenic studies of the mechanism of the reaction catalyzed by o-succinylbenzoate synthase from *Escherichia coli*. *Biochemistry* 42, 14427-14433.
- (24) Rieger, C. E., and Turnbull, J. L. (1996) Small scale biosynthesis and purification of gram quantities of chorismic acid. *Prep Biochem Biotechnol* 26, 67-76.
- (25) Palmer, D. R., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. (1999) Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase. *Biochemistry* 38, 4252-4258.
- (26) Popp, J. L., Berliner, C., and Bentley, R. (1989) Vitamin K (menaquinone) biosynthesis in bacteria: high-performance liquid chromatographic assay of the overall synthesis of o-succinylbenzoic acid and of 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylic acid synthase. *Anal Biochem* 178, 306-310.
- (27) Kirchhoff, W. (1993) in *National Institute of Standards and Technology Technical Note 1401* (Technology, N. I. o. S. a., Ed.), US Department of Commerce Technology Administration, Washington DC.
- (28) Becktel, W. J., and Schellman, J. A. (1987) Protein stability curves. *Biopolymers* 26, 1859-1877.
- (29) Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein*

Crystallography 26.

- (30) Collaborative Computational Project, N. (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst. D50*, 760-763.
- (31) Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* 30, 1022-1025.
- (32) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53, 240-255.
- (33) Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6, 458-463.
- (34) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
- (35) Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* 57, 122-133.
- (36) Kanzaki, H., McPhie, P., and Miles, E. W. (1991) Effect of single amino acid substitutions at positions 49 and 60 on the thermal unfolding of the tryptophan synthase alpha subunit from *Salmonella typhimurium*. *Arch Biochem Biophys* 284, 174-180.
- (37) Dill, K. A., Alonso, D. O., and Hutchinson, K. (1989) Thermal stabilities of globular proteins. *Biochemistry* 28, 5439-5449.
- (38) Pantoliano, M. W., Whitlow, M., Wood, J. F., Dodd, S. W., Hardman, K. D., Rollence, M. L., and Bryan, P. N. (1989) Large increases in general stability for subtilisin BPN' through incremental changes in the free energy of unfolding. *Biochemistry* 28, 7205-7213.
- (39) Matsumura, M., Signor, G., and Matthews, B. W. (1989) Substantial increase of protein stability by multiple disulphide bonds. *Nature* 342, 291-293.

- (40) Pegg, S. C., Brown, S. D., Ojha, S., Seffernick, J., Meng, E. C., Morris, J. H., Chang, P. J., Huang, C. C., Ferrin, T. E., and Babbitt, P. C. (2006) Leveraging enzyme structure-function relationships for functional inference and experimental design: the structure-function linkage database. *Biochemistry* 45, 2545-2555.
- (41) Zhang, X. J., Baase, W. A., Shoichet, B. K., Wilson, K. P., and Matthews, B. W. (1995) Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. *Protein Eng* 8, 1017-1022.
- (42) Offredi, F., Dubail, F., Kischel, P., Sarinski, K., Stern, A. S., Van de Weerd, C., Hoch, J. C., Prosperi, C., Francois, J. M., Mayo, S. L., and Martial, J. A. (2003) De novo backbone and sequence design of an idealized alpha/beta-barrel protein: evidence of stable tertiary structure. *J Mol Biol* 325, 163-174.
- (43) Taylor Ringia, E. A., Garrett, J. B., Thoden, J. B., Holden, H. M., Rayment, I., and Gerlt, J. A. (2004) Evolution of enzymatic activity in the enolase superfamily: functional studies of the promiscuous o-succinylbenzoate synthase from *Amycolatopsis*. *Biochemistry* 43, 224-229.
- (44) Landro, J. A., Kallarakal, A. T., Ransom, S. C., Gerlt, J. A., Kozarich, J. W., Neidhart, D. J., and Kenyon, G. L. (1991) Mechanism of the reaction catalyzed by mandelate racemase. 3. Asymmetry in reactions catalyzed by the H297N mutant. *Biochemistry* 30, 9274-9281.
- (45) Stites, W. E., Gittis, A. G., Lattman, E. E., and Shortle, D. (1991) In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core. *J Mol Biol* 221, 7-14.
- (46) Besenmatter, W., Kast, P., and Hilvert, D. (2007) Relative tolerance of mesostable and thermostable protein homologs to extensive mutation. *Proteins* 66, 500-506.
- (47) Petterson, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem* 25, 1605-1612.

Table 1: List of melting temp, ΔT_M , ΔH , and $\Delta\Delta G_u$ for OSBS wild-type and mutants.

	Melting Temp. (°C)	ΔT_M	ΔH (kcal/ mol)	$\Delta\Delta G_u$ (kcal/mol)
wt	50.8	0.0	86.1	0.00
Proton Abstraction Mutants				
K131A	55.4	4.7	84.0	1.24
K133A	50.6	-0.2	97.4	-0.05
K235A	48.5	-2.2	64.5	-0.59
K131A/K133A	54.3	3.6	83.4	0.95
K131A/K133A/K235A	47.9	-2.9	73.0	-0.77
Metal Coordination Mutants				
D161A	53.9	3.2	114.7	0.85
E190A	66.7	15.9	100.6	4.23
D213A	52.5	1.7	90.5	0.46
D161A/E190A/D213A	57.4	6.7	72.2	1.77
Complete Module Mutant				
K131A/K133A/D161A	64.6	13.8	87.5	3.68
E190A/D213A/K235A ("3KDED")				

[§]In addition to using the method of Schellman (28) for calculating the $\Delta\Delta G_u$ of these mutants, the method of Gibbs and Helmholtz was also used to find the change in free energy of unfolding using the equation

$$\Delta G = \Delta H_{wt} - T\Delta S_{wt} + \Delta C_p(T - T_{wt} - T \ln(T/T_{wt})).$$

Both methods yielded similar values.

Table 2: Statistics of data collection and structural refinement

Number of reflections	48414
Data cutoff	0
Resolution range (Å)	70 – 2.3
R_{work}/R_{free}	0.22/0.29
$I/\sigma(I)$	8.4 (1)
Completeness (%)	97.46
Redundancy	3.7
R_{merge} (%)	13
Atoms per asymmetric unit	9608
Average B -factor, protein atoms (Å ²)	41.1
Average B -factor, solvent atoms (Å ²)	37.6

Figure Legends:

Figure 1: (A) A superposition of several members of the enolase superfamily members, the active site residues of each protein are color-coded as: o-succinylbenzoate synthase (OSBS, pdb ID: 1fhv) black; Mandelate Racemase (MR, pdb ID: 2mnr), green; Muconate Lactonizing enzyme (MLE, pdb ID: 1muc), cyan; Enolase (pdb ID: 1ebh), gold; L-Ala-D/L-Glu epimerase (AEE, pdb ID: 1jpm), red. Active site residues are labeled using numbering from OSBS (B) Reactions performed by MR, MLE, Enolase and AEE. (C) Reaction performed by OSBS. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) (47)

Figure 2: Structural superposition of the OSBS wild-type structure (light gray) and the E190A mutant structure (dark gray.)

Figure 3: van't Hoff Plots for OSBS wild-type (black triangle), E190A (diamond) and 3KDED mutants (square).

Figure 1:

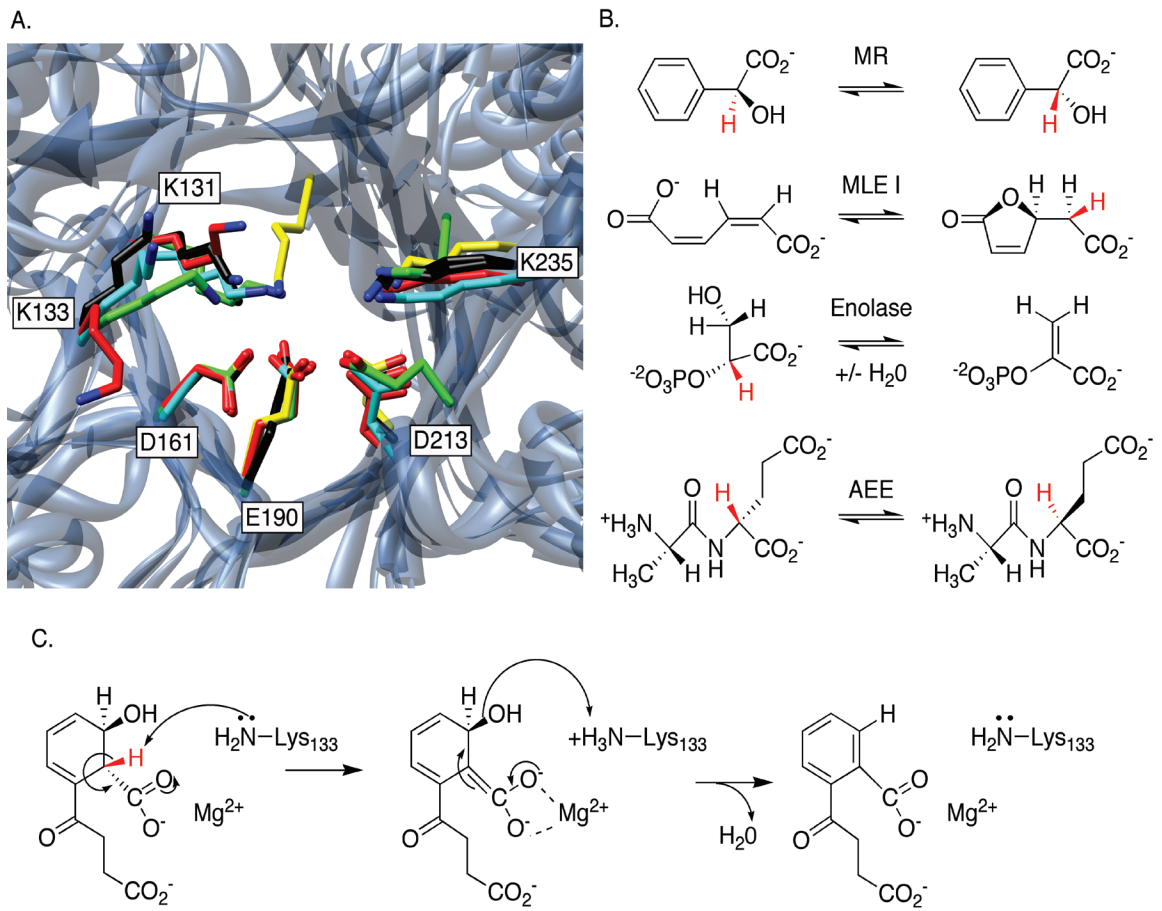


Figure 2:

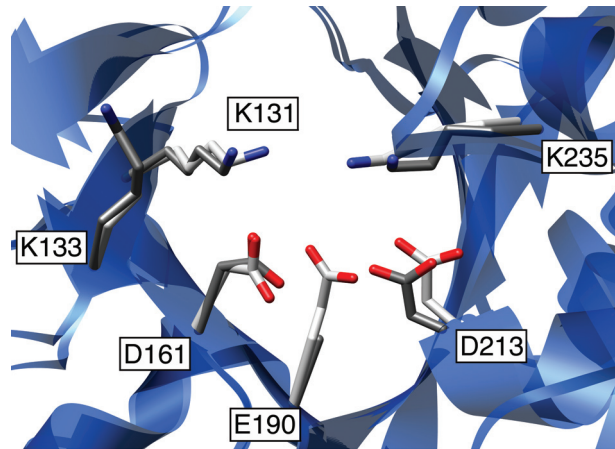
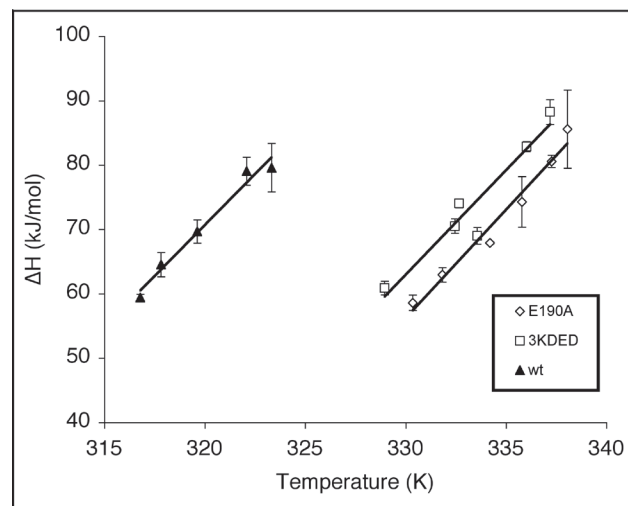


Figure 3:



Chapter 3:

Optimizing Enrichment and Recovery for *in vivo* Enzyme Selections

The following has been submitted for publication as follows:

Nagatani RA, Tawfik DS, Babbitt PC. “Optimizing Enrichment and Recovery for in vivo Enzyme Selections: The Glutamate Racemase Model.” *Protein Engineering Design and Selection*, 2008, *Submitted*.

ABSTRACT

Selection for function through the complementation of a knockout strain enables the screening of millions of candidate protein variants for the ability to provide a biologically significant level of an activity of interest. While high-throughput, survival in a selection is binary, generally providing little information about the level of activity that is detected. Yet, knowing the lower limit of detection for a selection is important for determining whether it is sufficiently sensitive for screening engineered libraries and isolating rare and weakly active variants. Moreover, selection conditions are often optimized for maximum enrichment of highly active variants at the expense of useful variants with much lower activity (the enrichment-recovery tradeoff). Using glutamate racemase as a model system, we attempted to address both of these issues using a series of mutants of this enzyme to estimate the minimum level of activity required to complement a knockout strain. The results show that this approach can detect and enrich mutants that are ~100,000 fold less active than the wild-type enzyme. The mutant with the lowest activity was used to calibrate the stringency of the selection and maximize the recovery of weak variants that comprise a very small proportion of the screened library. A gradual increase of the selection stringency enables maximal enrichment, as well as recovery of weakly active mutants.

INTRODUCTION

As the successes in engineering novel and optimized enzyme functions have grown in recent years, so too have the methods for detecting novel function. High-throughput screening (HTS) methods can now quickly screen thousands of constructs for changes in such properties as spectrophotometric absorbance, fluorescence, or color that reflect higher rates of engineered enzyme activity. Using selection based techniques like phage display (1) or *in vitro* compartmentalization (2), it is possible to dramatically increase the number of variant constructs that can be assayed to millions ($\sim 10^6$ to 10^{10}).

Another approach to selection for function, complementation in auxotrophic strains, allows identification of mutations that confer survival, the most stringent possible test for improvement of activity. Selection for function provides a powerful method for investigating how structural changes affect enzyme function and *in vivo* complementation, illustrated by many examples. These include work using mutants of beta-lactamase (3-5), triosephosphate isomerase (6, 7), DNA polymerase (8), imidazoleglycerol phosphate synthase (9), and prephenate dehydratase (10). Selection for function is also valuable for the directed evolution of specific enzyme activities, as demonstrated in the engineering of new reactions for enzymes previously unable to perform the β -galactosidase (11), phosphoribosylanthranilate isomerase (12), *o*-succinylbenzoate synthase (13, 14), and beta-lactamase (15), reactions. Additionally, selections for function have produced a variant of altered specificity in an aminotransferase (16), a monomeric chorismate mutase (17), and fold variants of methyltransferase (18).

Despite their value, the read-outs obtained from selection via complementation are usually binary (life vs. death), and provide little information about the level of activity that can be detected. Yet, if the goal of selection is to generate a novel function or to detect the presence of catalytic promiscuity, knowing the sensitivity of a selection is important because engineered and promiscuous enzymes often exhibit a level of activity

that is far lower than that of wild-type enzymes evolved to perform vital processes (13, 15). Moreover, without adequate sensitivity, weak variants that might be optimized further in subsequent rounds of mutation or shuffling are lost. Although past work has shown that selective pressure can be tuned to increase the dynamic range of a selection for function (10, 16, 19-21), determining the lower limit of sensitivity for a selection, and its efficiency of recovery for different levels of activity, is still not a matter of routine.

In the work reported here, we attempted to address these issues using as a model system a selection for glutamate racemase activity. Mutants of the glutamate racemase enzyme from *Lactobacillus fermenti* that exhibit compromised activity relative to the wild-type protein were tested to determine whether they could complement a glutamate racemase knockout strain (WM335). Selection was based on the strain's requirement for D-glutamate for cell wall synthesis, which is normally produced from L-glutamate by glutamate racemase (22). We hypothesized that selections using mutants compromised in the ability to racemize L-glutamate to D-glutamate would lead to lowered concentrations of D-glutamate available for cell wall synthesis, thereby resulting in observable changes in phenotype or growth rates.

To quantify the lower limit of activity required to complement the knockout strain, we tested both single site mutants of compromised activity that had been previously characterized (23), and double mutants we designed to exhibit even lower glutamate racemase activity. Using the most severely compromised double mutant that could still complement the knockout strain, we were able to calibrate the stringency of the glutamate racemase selection to maximize the enrichment (namely reduce the fraction of false positives) as well as the recovery of weak variants. This approach to calibrating selection for function may be generally useful in other systems to quantify the levels of activity required for selection in auxotrophic strains and to identify engineered or promiscuous enzymes that have very low levels of function.

MATERIALS AND METHODS

Strains

Escherichia coli strain WM335 (22), which has no endogenous glutamate racemase activity, was a generous gift from Dr. Masaaki Wachi (Tokyo Institute of Technology) and Dr. Makoto Ashiuchi (Kochi University).

Site-Directed Mutagenesis

L. fermenti glutamate racemase mutants D10N, D36N, E152Q and H186N cloned into a puc18 vector (Fermentas, Ontario, Canada), a generous gift from Dr. Martin Tanner (University of British Columbia), were used for site-directed mutagenesis experiments. The QuickChange site-directed mutagenesis method (Stratagene, La Jolla, CA) was used to introduce additional point mutations. The mutants were confirmed by DNA sequencing.

Selection for Complementation

Plasmids encoding single and double mutants of *L. fermenti* glutamate racemase were transformed into cells of *E. coli* strain WM335 by electroporation (1800 V) using a Eppendorf Electroporator 2510 system (Eppendorf, Hamburg, Germany). Cells were allowed to recover in Luria Broth (LB) media containing D-glutamate (0.005% v/v) for 1 hour at 37°C with gentle shaking. Experiments in which recovery was performed in the absence of D-glutamate prior to plating yielded similar results. For complementation experiments in the presence of varying concentrations of D-glutamate, a portion of the cells were plated on LB-agar plates containing ampicillin (50 ug/mL) and between 0 and 0.0025% (v/v) D-glutamate, then incubated at 37° C.

Library by Error-prone PCR

The non-complementing double mutant D10N/D36N was used as a template for biased error-prone PCR. PCR mixtures contained 10 ng of D10N/D36N plasmid DNA, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 4 mM MgCl₂, 0.2 mM dCTP, 0.2 mM dATP, 1 mM dTTP, 1 mM dGTP, 2.5 U Taq polymerase, 125 µM MnCl₂ and primers specific for

the glutamate racemase gene to a final concentration of 0.4 μM . Following amplification, PCR products were digested with EcoRI and HindIII, then ligated into puc18 plasmid. Based on sequence data, the mutation rate was ~ 0.6 mutations per gene. ElectroTen Blue competent cells were transformed with the ligation products via electroporation, then plated on LB-agar plates + carbenicillin (50 $\mu\text{g}/\text{mL}$). Following incubation, the resulting colonies were collected, and the plasmid library was extracted using the Qiagen Mini-Prep kit (Qiagen, La Jolla, CA). The library was subsequently subjected to two rounds of selection by transformation into WM335 cells and plating on LB-agar supplemented with carbenicillin (50 $\mu\text{g}/\text{mL}$) and 0.0008% (m/v) D-glutamate. Following each selection, cells were scraped off each plate and DNA was extracted. After two rounds of 0.0008% D-glutamate selection, the resultant DNA library was used to transform WM335 cells that were plated on LB-agar + carbenicillin (50 $\mu\text{g}/\text{mL}$). DNA was extracted from resultant colonies and sequenced.

Enrichment and Recovery

WM335 cells were transformed with ~ 100 ng of plasmid DNA mixtures containing 0.5 ng, 0.05 ng, 0.005 ng or 0.0005 ng of the puc18 plasmid carrying wild-type glutamate racemase (or its D10N, or D10N/E152Q, mutants) and ~ 100 ng of the puc19 control plasmids. Following transformation, cells were allowed to recover in LB media with D-glutamate (0.005%, v/v) for one hour, then were plated on LB-agar containing between 0 and 0.0025% (v/v) D-glutamate. Plates were incubated at 37° until colonies appeared (overnight for wild-type and D10N mutant, approximately 2 days for the D10N/E152Q mutant). In the absence of selective pressure (0.0025% D-glutamate), the average number of colonies was $\sim 1.6 \times 10^6$.

To estimate the number of colonies that contained the wild-type or mutant glutamate racemase gene (true positive) and the number that contained no cloned gene (false positive), colony PCR analysis was performed using Taq polymerase (BioTaq, Maryland) and primers that flank the multiple cloning site of both the puc18 plasmid that

carried the glutamate racemase gene (~1,900 bp PCR product) and the puc19 control plasmid (456 bp PCR product). PCR products were evaluated by electrophoresis.

To determine enrichment after successive rounds of selection, plates of colonies were scraped and resuspended in LB media. Plasmid DNA from a portion of the suspension was isolated using the Qiagen Mini-Prep kit (Qiagen, La Jolla, CA). The recovered plasmid DNA (150ng) was then used to transform WM335 cells as previously described, and colonies were analyzed using the previously mentioned colony PCR method.

To determine how the expression levels of mutant glutamate racemase protein affected the enrichment and recovery of the weak D10N/E152Q mutant, results from enrichment and recovery experiments using uninduced cells were compared to cells that had been plated on 0.5 mM IPTG.

Purification of wild-type L. fermenti glutamate racemase

E. coli BL21 (DE3) cells were transformed with the plasmid and grown at 37 °C in LB medium for 18 hours without induction. Cells were resuspended in 30 mLs of binding buffer (30 mM TEA, 1 mM (D,L) glutamate, 10% glycerol (v/v), 2 mM DTT, pH 7.5), then disrupted by passage through a microfluidizer. Following centrifugation, the supernatant was loaded onto a DEAE 5 mL HP column (GE Biosciences, Piscataway, NJ) that had been equilibrated with binding buffer. Following a wash step with 5 column volumes of buffer, the wild-type protein was eluted using a linear NaCl gradient (0 to 0.5 M NaCl in 30 mM TEA, 1 mM (D,L) glutamate, 10% glycerol (v/v), 2 mM DTT, pH 7.5). The active fractions were diluted to a total volume of 50 mL in a buffer comprised of 50 mM TEA, 1 mM (D,L) glutamate, 10% glycerol, 0.3 mM DTT, pH 7.0, and loaded onto a DEAE anion exchange column (GE Biosciences, Piscataway, NJ), then eluted with a linear NaCl gradient (0 to 0.5 M NaCl in 50 mM TEA, 1 mM (D,L) glutamate, 10% glycerol, 0.3 mM DTT, pH 7.0). Active fractions were pooled and frozen at -80 °C.

Frozen fractions were thawed and diluted 10 fold in a buffer containing 10 mM

potassium phosphate, 1 mM (D,L) glutamate, 10% glycerol and 0.44 mM DTT, pH 7.6. The diluted protein was concentrated using an Amicon 30 KDa MW-cutoff centrifugal filter to 400 ul and loaded onto a Sephadex 75 size exclusion column (GE Biosciences, Piscataway, NJ). Active fractions were collected and the purity of the samples was verified using SDS-PAGE, which showed a single band (Supplementary Information, Figure A).

Purification of the D10N/E152Q glutamate racemase mutant

E. coli BL21 (DE3) cells were transformed with the plasmid and grown at 37 °C in LB medium for 18 hours without induction. Cells were then induced with IPTG to a total concentration of 1 mM and grown for 3 hours. The cells were then centrifuged and frozen at - 80°C.

The frozen pellet was resuspended in 30 mLs of binding buffer (30 mM TEA, 1 mM (D,L) glutamate, 10% glycerol (v/v), 2 mM DTT, pH 8), then disrupted by passage through a microfluidizer. Following centrifugation, the supernatant was loaded onto a DEAE 5 mL HP column (GE Biosciences, Piscataway, NJ) that had been equilibrated with binding buffer. Following a wash step with 5 column volumes of buffer, the mutant protein was eluted using a linear NaCl gradient (0 to 0.5 M NaCl in 30 mM TEA, 1 mM (D,L) glutamate, 10% glycerol (v/v), 2 mM DTT, pH 8). Fractions were analyzed using SDS-PAGE electrophoresis, and those with bands corresponding to the expected size for glutamate racemase were pooled and diluted to a total volume of 50 mL in a buffer comprised of 50 mM TEA, 1 mM (D,L) glutamate, 10% glycerol, 0.3 mM DTT, pH 7.2, and again loaded onto a DEAE anion exchange column (GE Biosciences, Piscataway, NJ), then eluted with a linear NaCl gradient (0 to 0.5 M NaCl in 50 mM TEA, 1 mM (D,L) glutamate, 10% glycerol, 0.3 mM DTT, pH 7.2). Fractions were analyzed using SDS-PAGE electrophoresis, and those with bands corresponding to the expected molecular weight of glutamate racemase were pooled, diluted to a total volume of 50 mL in a buffer containing 50 mM TEA, 1 mM (D,L) glutamate, 10% glycerol, 0.3 mM

DTT, pH 7.2, and loaded onto a Q 5 mL HP column (GE Biosciences, Piscataway, NJ). Fractions were again analyzed using SDS-PAGE, and the fraction containing the mutant protein was diluted into a buffer containing 10 mM Potassium phosphate, 1 mM (D,L) glutamate, 10% glycerol and 0.44 mM DTT, pH 7.6 and frozen at -80°C.

The frozen protein was concentrated to a total volume of 200 ul and loaded onto a Sephadex 75 column. Fractions were analyzed via SDS-PAGE, and the fraction containing a single band (Supplementary Information, Figure A) at the expected molecular weight for the mutant protein was frozen at -80°C.

Coupled Enzyme Assay for Glutamate Racemase Activity

The D10N/E152Q mutant and wild-type protein were assayed for activity using the coupled enzyme assay for glutamate racemase activity described by Gallo *et al.*(24). Protein concentrations in the assay were 1.03 µM for the D10N/E152Q mutant and 12 nM for the wild-type protein.

RESULTS

Complementation of the knockout strain WM335 by Mutants of L. fermenti Glutamate Racemase

Single site mutants of *L. fermenti* glutamate racemase that had been previously characterized by Glavas and Tanner (23) were able to complement the *E. coli* glutamate racemase knockout strain WM335 (Table I). Following transformation, the cells grew in a manner that was indistinguishable from cells transformed with the wild-type gene.

To create a mutant that might result in a more compromised phenotype, mutations were made in all combinations of the single site mutants to generate the mutants D10N/D36N, D10N/E152Q, D10N/H186N, D36N/E152Q, D36N/H186N and E152Q/H186N. Cells transformed with plasmids containing the double mutants D36N/E152Q and E152Q/H186N grew at the same rate observed for cells transformed with the wild-type plasmid, both resulting in a similar number of colonies following overnight growth. Cells

transformed with plasmids containing the double mutants D10N/D36N, D10N/H186N and D36N/H186N were monitored for 1 week with no evidence of growth. Colonies took ~2 days to appear on plates of cells transformed with plasmid encoding the double mutant D10N/E152Q, and these plates consistently had fewer colonies compared to cells transformed with the wild-type plasmid, suggesting that this mutant produces a selectable phenotype of further compromised glutamate racemase activity relative to both wild type and the activity-compromised single site mutants we tested. This double mutant was used to calibrate the selection further.

Enzyme Kinetics

For the wild-type protein, the derived k_{cat} and K_M were 21 s^{-1} and 0.23 mM , respectively (Table I), which are similar to previously published values of 68 s^{-1} and 0.26 mM (24). Kinetic constants for the D10N/E152Q mutant (Table I) indicate that the catalytic efficiency of this enzyme is approximately 10^5 lower than the previously reported efficiency of the wild-type enzyme (See also Supplementary Information, Figure B).

Enrichment and Recovery Experiments

To estimate how well the selection would perform in an engineering experiment aimed at recovering gene variants exhibiting very low levels of a selectable activity, we used the D10N/E152Q as a proxy for weakly active constructs that might be expected in initial rounds of engineering and design. Experiments were performed using this mutant to determine the rates of enrichment that could be obtained under different conditions and to calibrate the stringency of the complementation selection required to isolate such very weak variants. We found that enrichment depended on the initial percentage of true positives in a background of false positives (total library size was $\sim 1.6 \times 10^6$ on average) and on the stringency of selection. Although selective pressure has previously been tuned in other systems by regulating the expression level of the target protein and other selection conditions (10, 16, 19-21), this approach was not used in our case

because induction of protein expression by addition of IPTG was found to reduce mutant recovery. As an alternative, we chose to add varying amounts of reaction product (D-glutamate) to the growth media to supplement the low amount produced by the weak activity of the double mutant, thereby reducing the selection stringency and increasing cell viability (Table II).

As shown in Table IID, high stringency conditions (no D-glutamate supplementation) only retained the weak variant (plasmids encoding the D10N/E152Q mutant) when it constituted a relatively large proportion of the library (0.5%). Conversely, if libraries are initially grown on media fully supplemented with D-glutamate (0.0025%) with no selective pressure, false positives comprise the majority of colonies, “washing out” the true positives (Table IIA). Even with gradual increases in the stringency of selection, libraries that are initially grown with no selective pressure fail to recover the weak variants when they comprise less than 0.05% of the library (Table IIA).

Enrichment of true positives is 50% when the initial DNA pool contains 0.5% D10N/E152Q and the cells are grown on 0.0008% D-glutamate (low stringency). However, this enrichment rises to 100% if the library isolated from the first round is subsequently subjected to a second round of low stringency selection (0.0008% followed by 0.0008% selection, Table IIB). For libraries where the weak variants comprise only 0.05% and 0.005% of the total population, the enrichment rises from 0% (no true positives) in the first round to 100% and 70% true positives in the second round, respectively.

For maximum recovery of weak variants, it appears that successive rounds of low stringency selection lead to the highest recovery. While only one round of low stringency selection yields true positives for the 0.5%, 0.05% and 0.005% libraries (0.0008% followed by 0% selection, Table IIC), the number of recovered true positives increases if two rounds of low stringency selection are performed prior to the high stringency selection (0.0008% to 0.0008% to 0% selection, Table IIB). In contrast, the wild-type and

D10N single mutant, which are both more active than the D10N/E152Q mutant (Table I), were able to complement the auxotrophic WM335 cells, even when constituting only 0.0005% of the total DNA library used for transformation (Table III). This sensitivity for true positives is comparable to results reported for other selections (19).

Library Selection by Complementation

To simulate how the selection might perform in an engineering experiment using a library of variants generated from a gene unable to complement the knockout strain, the double mutant D10N/D36N was used as a template for library generation by error-prone PCR. Following two rounds of low-stringency selection (media supplemented with 0.0008% D-glutamate), the resultant library was used to transform WM335 cells for selection in the absence of D-glutamate supplementation. Over 10,000 colonies were recovered, and sequencing of three of the progeny showed that all contained a N10D mutation, thereby reverting the gene back to the sequence of the single mutant D36N. We recall that this gene, while showing complementation capability equivalent to that of wild-type (as described above), has an enzyme efficiency approximately 350-fold worse than wild-type (Table I).

DISCUSSION

Calibration of the Selection using Activity Compromised Mutants

To determine a starting point for calibrating the glutamate racemase selection, we used activity compromised mutants that had been previously characterized in experiments aimed at identifying the contribution to activity and mechanism of specific residues (23). Although the least active of these mutants exhibited catalytic efficiencies approximately 5,000-fold worse than wild-type glutamate racemase, we found that their phenotypes in the selection were indistinguishable from that of the wild-type. To generate mutants of even further compromised activity, we used the simple strategy of combining all of the single site mutants to generate a set of double mutants. These double mutants

exhibited both nonviable as well as somewhat to severely compromised phenotypes. (In a conceptually related strategy, Sengupta et al. used mutants of β -lactamase that spanned several orders of magnitude in catalytic efficiency to follow gene transcription *in vivo* (25)). Based on our results, we suggest that the activity of the worst of these, the D10N/E152Q mutant, may be close to the lower bound that can be detected by the selection, making it especially useful for calibrating the selection and identifying conditions under which rare and weakly active mutants can be recovered. While the differences in the growth rate and survival *in vivo* between this mutant and the wild-type protein are likely a function of both expression levels of each protein and differences in enzyme activity, these results also suggest that the dramatic drop in catalytic efficiency is likely the primary factor.

Tuning for Enrichment and Recovery of Weakly Active Mutants

Previous work has utilized the tuning of selection stringency to improve the recovery of highly active proteins (10, 16, 19-21), or to improve the rate of recovery of true positives (19). We have focused here on an alternative goal, tuning selection stringency for recovery of enzymes that display only very low levels of a desired activity while minimizing the number of false positives, thereby rescuing weakly active variants that are potentially valuable for subsequent optimization in engineering experiments.

While in theory, the glutamate racemase selection should allow recovery for very weakly active as well as moderately active mutants, we have seen in practice that a normally stringent selection can eliminate weakly active true positives during initial rounds of selection (Table II). The lower survival rates of the D10N/E152Q mutant suggest that at the lower limit of complementation capability, cell viability is a continuum across a population of WM335 competent cells, where healthy cells live and the weaker cells die despite harboring the same weakly active construct. Conversely, we found that subjecting the library to no selective pressure also decreases the recovery of weak variants. Thus, although experiments at low stringency selection showed that

some true positives exist in the initial population of growing cells (Table IIB, IIC), these weak variants could not be enriched and recovered unless some minimal level of selective pressure is applied to prevent growth of the large number of false positives that dominate under these conditions. Taken together, these results suggest that for this system, enrichment of true positives in the extremes of selection conditions can be a daunting challenge, especially when the proportion of true positives may be very small in extremely diverse libraries. Gradual selection (in our case, increasing stringency by decreasing supplement levels over several rounds of selection) provides a powerful method for recovering and enriching true positives. Thus, for the glutamate racemase selection, using an initially low stringency selection eventually leads to 100% enrichment and high recovery of weak variants even when their proportion in the initial library is only 1 in 20,000.

Applications in Enzyme Engineering and Design

It is generally recognized that the activity of designed and evolving enzymes is typically far worse than that of their wild-type counterparts, especially in early rounds of the optimization process. For example, Park *et al.* utilized several rounds of selection conferring the ability to destroy the antibiotic cefotaxime to obtain a construct that was over 4,000 fold worse than the wild-type metal dependent beta-lactamase IMP-1 (15). Schmidt *et al.* used an anaerobic selection to evolve novel *o*-succinylbenzoate synthase (OSBS) function in a dipeptide epimerase that in initial selection experiments was 248,000 fold less efficient than the wild-type OSBS enzyme (13). Consistent with these reports, our results characterizing the glutamate racemase selection show that constructs that are both rare and extremely poor (as much as 100,000 fold worse) in catalytic efficiency relative to the wild-type enzyme can still be recovered. Thus, this level of sensitivity is likely to be more than adequate for many efforts in identifying functional variants for subsequent optimization.

In summary, our results using the glutamate racemase selection indicate that by

calibrating the selection sensitivity and using that information to define an appropriate window for gradual selection, it is possible to recover functional variants that are weakly active and that represent only a very small proportion of an initial library. Previously characterized single site mutants of glutamate racemase (23) provided a simple way to identify and design compromised mutants required for these purposes. We suggest that this strategy may be generally applicable for the engineering or design of enzyme function in other systems.

ACKNOWLEDGMENTS

The authors thank Dr. Martin Tanner (University of British Columbia) for providing the single mutants of *Lactobacillus fermenti* glutamate racemase and Dr. Makoto Ashiuchi (Kochi University) and Dr. Masaaki Wachi (Tokyo Institute of Technology) for providing the WM335 cells. The authors would also like to thank Dr. Sergio Peisajovich (UCSF) and Roy Bekerman (Weizmann Institute) for technical guidance and Matthew Good, Claire Rowe, and Janet Yang for sharing resources and instrumentation. This work was supported by NIH RO1-GM60595 to PCB, the UCSF President's Dissertation Fellowship to RAN, and a Minerva Foundation grant to DST.

Tables and Figures:

Table I: Kinetic and Complementation Data for Wild-Type and Single Mutation

Glutamate Racemases

	k_{cat} (s ⁻¹)	K_m (mM)	k_{cat} / K_m (s ⁻¹ M ⁻¹)	Complement?
wild-type ^a	21	0.23	2.1 x 10 ⁵	++ ^c
D10N ^b	0.068	1.3	5.3 x 10 ¹	++
D36N ^b	20	35	5.7 x 10 ²	++
E152Q ^b	38	5.8	6.5 x 10 ³	++
H186N ^b	0.045	1.1	4.1 x 10 ¹	++
D10N/E152Q	(1.5 ± 0.1) x 10 ⁻³	0.59 ± 0.2	(2.5 ± 0.2)	+

^a. Kinetic data from Gallo and Knowles (1993) reported a k_{cat} of 68 s⁻¹ and a K_M of 0.26 mM with the purified wild-type glutamate racemase.

^b. Kinetic data from Glavas and Tanner (2001).

^c. “++” indicates a high survival and growth rates identical to wild-type (colonies grew overnight). “+” indicates a lower survival rate and slower growth (colonies grew in ~2 days).

Table II: Recovery and Enrichment for D10N/E152Q mutant

A. No Stringency to High-Stringency Gradual Selection (0.0025% to 0.0008% to 0% D-glutamate)^a

True positive DNA (%)	Round 1- 0.0025%		Round 2- 0.0008%		Round 3- 0%	
	Recovery	Enrichment	Recovery	Enrichment	Recovery	Enrichment
0.5	lawn	0	lawn	0	~350	100%
0.05	lawn	0	lawn	0	~250	100%
0.005	lawn	0	lawn	0	0	N/A
0.0005	lawn	0	lawn	0	0	N/A

B. Low-Stringency Successive Selection (0.0008% to 0.0008% to 0% D-glutamate)

True positive DNA (%)	Round 1- 0.0008%		Round 2- 0.0008%		Round 3- 0%	
	Recovery	Enrichment	Recovery	Enrichment	Recovery	Enrichment
0.5	lawn	50%	lawn	100%	lawn	100%
0.05	lawn	0	lawn	100%	lawn	100%
0.005	lawn	0	lawn	70%	lawn	100%
0.0005	lawn	0	lawn	0	0	N/A

C. Low-Stringency to High-Stringency Selection (0.0008% to 0% D-glutamate)

True positive DNA (%)	Round 1- 0.0008%		Round 2- 0%	
	Recovery	Enrichment	Recovery	Enrichment
0.5	lawn	50%	~1200	100%
0.05	lawn	0	~900	100%
0.005	lawn	0	~400	70%
0.0005	lawn	0	0	N/A

D. High-Stringency Selection (0% D-glutamate)

True positive DNA (%)	Round 1- 0%	
	Recovery	Enrichment
0.5	7	100%
0.05	0	N/A
0.005	0	N/A
0.0005	0	N/A

a. Recovery and enrichment results are for the indicated amount of D-glutamate supplementation for each successive round of selection.

Table III: Enrichment and Recovery for Wild-type and D10N Mutant Glutamate

Racemase

A. Enrichment and Recovery of Wild-Type Glutamate Racemase

True positive DNA (%)	0.0025 ^a % D-glutamate		0.0008% D-glutamate		0% D-glutamate	
	Recovery	Enrichment	Recovery	Enrichment	Recovery	Enrichment
0.5	lawn	0	lawn	20%	~3200	100%
0.05	lawn	0	lawn	0	~630	100%
0.005	lawn	0	lawn	0	~120	100%
0.0005	lawn	0	lawn	0	27	100%

B. Enrichment and Recovery of the D10N mutant of Glutamate Racemase

True positive DNA (%)	0.0025% D-glutamate		0.0008% D-glutamate		0% D-glutamate	
	Recovery	Enrichment	Recovery	Enrichment	Recovery	Enrichment
0.5	lawn	0	lawn	50%	~163	100%
0.05	lawn	0	lawn	10%	~78	100%
0.005	lawn	0	lawn	0	4	N/A
0.0005	lawn	0	lawn	0	2	N/A

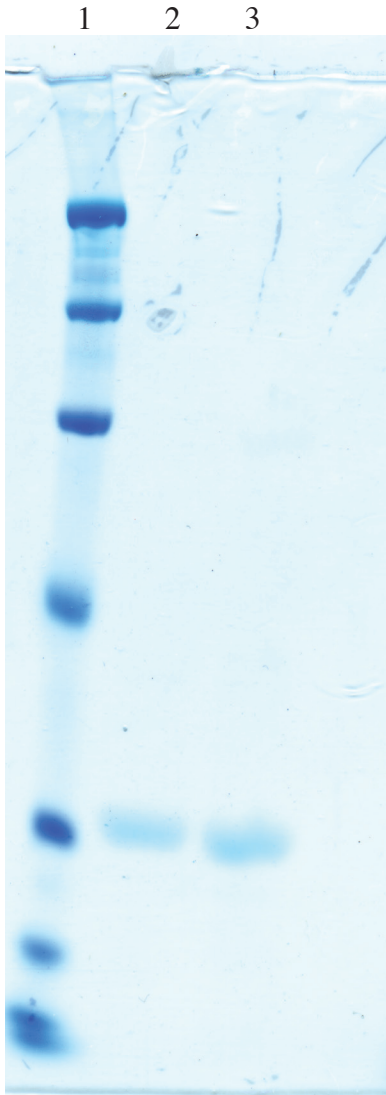
^a No stringency, low stringency, and high stringency conditions are defined as in Table II.

- (1) Olsen, M., Iverson, B., and Georgiou, G. (2000) High-throughput screening of enzyme libraries. *Curr Opin Biotechnol* 11, 331-337.
- (2) Griffiths, A. D., and Tawfik, D. S. (2006) Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol* 24, 395-402.
- (3) Wang, X., Minasov, G., and Shoichet, B. K. (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* 320, 85-95.
- (4) Hecky, J., and Muller, K. M. (2005) Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of beta-lactamase. *Biochemistry* 44, 12640-12654.
- (5) Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D. S. (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444, 929-932.
- (6) Silverman, J. A., Balakrishnan, R., and Harbury, P. B. (2001) Reverse engineering the (beta/alpha)₈ barrel fold. *Proc Natl Acad Sci U S A* 98, 3092-3097.
- (7) Kursula, I., Salin, M., Sun, J., Norledge, B. V., Haapalainen, A. M., Sampson, N. S., and Wierenga, R. K. (2004) Understanding protein lids: structural analysis of active hinge mutants in triosephosphate isomerase. *Protein Eng Des Sel* 17, 375-382.
- (8) Patel, P. H., and Loeb, L. A. (2000) DNA polymerase active site is highly mutable: evolutionary consequences. *Proc Natl Acad Sci U S A* 97, 5095-5100.
- (9) Hocker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A., and Sterner, R. (2001) Dissection of a (beta/alpha)₈-barrel enzyme into two folded halves. *Nat Struct Biol* 8, 32-36.
- (10) Kleeb, A. C., Edalat, M. H., Gamper, M., Haugstetter, J., Giger, L., Neuenschwander, M., Kast, P., and Hilvert, D. (2007) Metabolic engineering of a genetic selection system with tunable stringency. *Proc Natl Acad Sci U S A* 104, 13907-13912.
- (11) Hall, B. G. (1995) Evolutionary potential of the *ebgA* gene. *Mol Biol Evol* 12, 514-517.
- (12) Jurgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M., and Sterner, R. (2000) Directed evolution of a (beta/alpha)₈-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A* 97, 9925-9930.
- (13) Schmidt, D. M., Mundorff, E. C., Dojka, M., Bermudez, E., Ness, J. E., Govindarajan, S., Babbitt, P. C., Minshull, J., and Gerlt, J. A. (2003) Evolutionary potential of (beta/alpha)₈-barrels: functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry* 42, 8387-8393.
- (14) Vick, J. E., and Gerlt, J. A. (2007) Evolutionary Potential of (beta/alpha)₈-Barrels: Stepwise Evolution of a “New” Reaction in the Enolase Superfamily. *Biochemistry* 46, 14589-14597.
- (15) Park, H. S., Nam, S. H., Lee, J. K., Yoon, C. N., Mannervik, B., Benkovic, S. J., and Kim, H. S. (2006) Design and evolution of new catalytic activity with an existing protein scaffold. *Science* 311, 535-538.

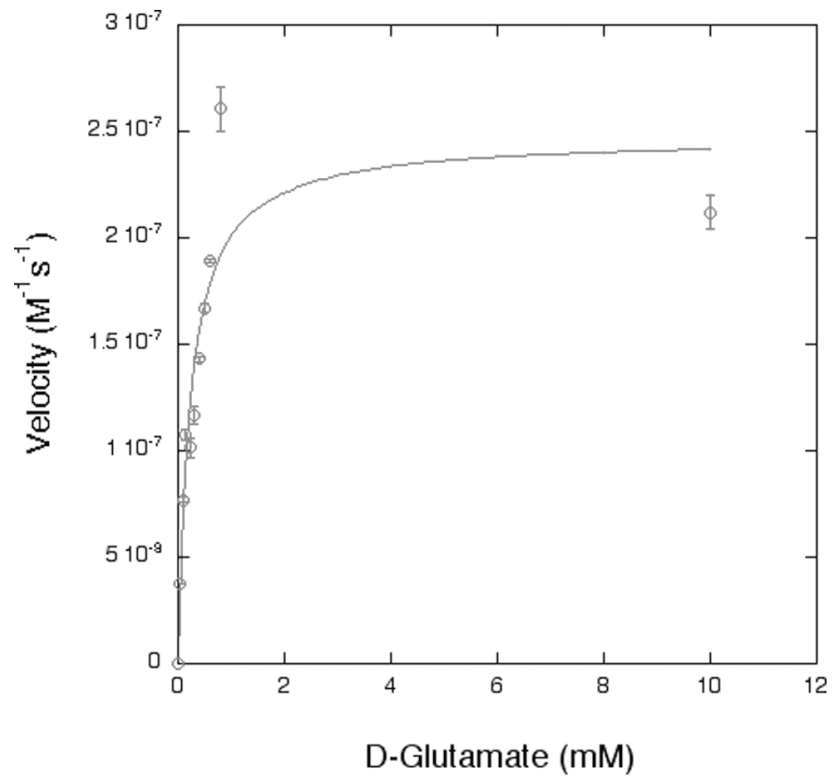
- (16) Yano, T., Oue, S., and Kagamiyama, H. (1998) Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc Natl Acad Sci U S A* 95, 5511-5515.
- (17) MacBeath, G., Kast, P., and Hilvert, D. (1998) Redesigning enzyme topology by directed evolution. *Science* 279, 1958-1961.
- (18) Peisajovich, S. G., Rockah, L., and Tawfik, D. S. (2006) Evolution of new protein topologies through multistep gene rearrangements. *Nat Genet* 38, 168-174.
- (19) van Sint Fiet, S., van Beilen, J. B., and Witholt, B. (2006) Selection of biocatalysts for chemical synthesis. *Proc Natl Acad Sci U S A* 103, 1693-1698.
- (20) Neuenschwander, M., Butz, M., Heintz, C., Kast, P., and Hilvert, D. (2007) A simple selection strategy for evolving highly efficient enzymes. *Nat Biotechnol* 25, 1145-1147.
- (21) Bernath, K., Magdassi, S., and Tawfik, D. S. (2005) Directed evolution of protein inhibitors of DNA-nucleases by in vitro compartmentalization (IVC) and nano-droplet delivery. *J Mol Biol* 345, 1015-1026.
- (22) Dougherty, T. J., Thanassi, J. A., and Pucci, M. J. (1993) The Escherichia coli mutant requiring D-glutamic acid is the result of mutations in two distinct genetic loci. *J Bacteriol* 175, 111-116.
- (23) Glavas, S., and Tanner, M. E. (2001) Active site residues of glutamate racemase. *Biochemistry* 40, 6199-6204.
- (24) Gallo, K. A., and Knowles, J. R. (1993) Purification, cloning, and cofactor independence of glutamate racemase from Lactobacillus. *Biochemistry* 32, 3981-3990.
- (25) Sengupta, D., Lin, H., Goldberg, S. D., Mahal, J. J., and Cornish, V. W. (2004) Correlation between catalytic efficiency and the transcription read-out in chemical complementation: a general assay for enzyme catalysis. *Biochemistry* 43, 3570-3581.

SUPPLEMENTARY INFORMATION

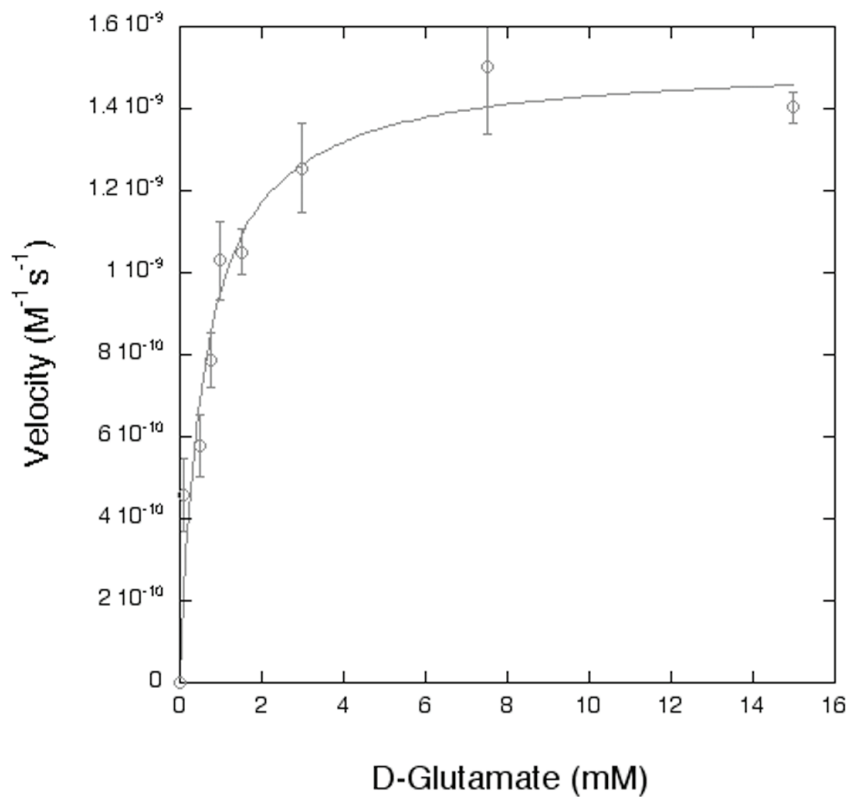
A. SDS-GEL of wild-type, and D10N/E152Q mutant, glutamate racemase from *Lactobacillus fermenti* purified as described. Lane 1: BioRad Broad Range ladder, Lane 2: Wild-type glutamate racemase, Lane 3: D10N/E152Q mutant.



B. Kinetic data for wt. glutamate racemase. Assays were performed using an enzyme concentration of 12 nM. Each data point represents at least two trials (with the exception of the data point for 0.05 mM which used 1 trial). Error bars were calculated based on the standard error.



C.. Kinetic data for the D10N/E152Q mutant of glutamate racemase. Assays were performed with an enzyme concentration of 1.03 μM . Each data point represents at least 3 trials and error bars represent the calculated standard error.



Chapter 4:

Following Nature's Example for Pathway Evolution

ABSTRACT

In nature, organisms recruit and optimize for function robust enzymes to form metabolic pathways and gain selective advantage. In some organisms, members of the enolase and amidohydrolase superfamilies have been paired in operons to build metabolic pathways capable of converting metabolites or antibiotic degradation products to L-amino acids (Sakai). Guided by this example, we have developed an in vitro selection for function that pairs members of the two superfamilies to build a metabolic pathway for production of D-glutamate, which is vital for cell wall synthesis. Using this selection, we can detect L-Ala-L-Glu epimerization in characterized enolase superfamily members from several different organisms and have identified that N-acetyl-D-glutamate amidohydrolase from *Achromobacter xylosoxydans* catalyzes the hydrolysis of L-Ala-D-Glu to produce D-glutamate.

INTRODUCTION

Within the mechanistically diverse enolase superfamily, over 2,000 members share a conserved architecture of catalytic residues used to abstract protons alpha to carboxylic acid groups, leading to stabilization of a common type of enolate anion intermediate (1). Members are able to perform this conserved partial reaction on a variety of substrates, giving rise to different overall reactions that include dehydration, racemization, epimerization, and lactonization reactions that span a broad set of different E.C. (Enzyme Commission) classes. In some cases, members are able to recognize and turn over more than one substrate due to differences in specificity elements that lead to promiscuity (2-5) or that have been created through *in vitro* enzyme engineering (6).

Similarly, members of the amidohydrolase superfamily utilize a common catalytic architecture of residues for which several related constellations have been identified to coordinate metal residues and stabilize a common type of intermediate (7). Over 5,000 sequences belonging to this superfamily have been identified, all of which are presumed to share a similar active site organization and mechanistic attributes. Many members perform their unique overall reactions using a water molecule activated by coordinated metal ion cofactors. Like the enolase superfamily, members perform their characteristic reactions on a variety of substrates (8). In one member of the amidohydrolase superfamily, engineering specificity elements remote from the catalytic module of metal coordinating residues conferred the ability to turn over novel substrates (9).

While individual members of these superfamilies are known to be involved in specific metabolic processes (1, 7), recent work has also shown that enolase and amidohydrolase superfamily members are also found together in operons required for the performance of vital cellular functions. In these operons, enolase superfamily members racemize or epimerize a variety of N-terminally modified amino acids or dipeptides. The amidohydrolases in these operons specifically cleave the N-C bonds of the racemized products, thereby removing N-terminal modifications as a strategy for scavenging amino

acids for other cellular processes (Figure 1A)(4).

Guided by the functional pairing of amidohydrolase and enolase superfamily members in pathways for scavenging modified D-amino acids that we observe in nature (4), our goal was to extend the functional capabilities observed in individual superfamily members to engineering new function at the pathway level. We have used superfamily members with specific functionalities to construct an enzyme pathway that depends on both enolase and amidohydrolase superfamily reactivity to produce a selectable output. Our selection for function detects epimerization of L-Ala-L-Glu to L-Ala-D-Glu (via the enolase superfamily reaction), followed by cleavage of the product to form the constituent amino acids (via the amidohydrolase superfamily reaction) (Figure 1B). Using this selection, we can detect L-Ala-L-Glu epimerase activity in several members of the enolase superfamily and have identified a novel promiscuous function for a characterized member of the amidohydrolase superfamily.

METHODS AND MATERIALS:

Strains

Escherichia coli strain WM335 (10), which has no endogenous glutamate racemase activity, was a generous gift from Dr. Masaaki Wachi (Kochi University) and Dr. Makoto Ashiuchi (Tokyo Institute of Technology). *Alcaligenes xylosoxydans* was purchased from the American Type Culture Collection (ATCC, Manassas, VA).

Cloning of N-acetyl-D-glutamate Amidohydrolase

The N-acetyl-D-glutamate amidohydrolase (NADGA) gene described by Wakayama et al. (11) was amplified from a culture of *Alcaligenes xylosoxydans* using colony PCR. The gene was cloned into the *EcoRI* and *HindIII* restriction enzyme sites of the plasmid puc18 (Fermentas, Ontario, Canada) to construct the pNADGA plasmid.

Selection for L-Ala-L-Glu Epimerase activity

WM335 cells were co-transformed with the libraries encoding variants of the L-

Ala-D/L-Glu epimerase (AEE) from *E. coli* and the pNADGA plasmid by electroporation (1800 V) using a Eppendorf Electroporator 2510 system (Eppendorf, Hamburg, Germany). Cells were allowed to recover in Super Optimal Catabolite Repression Broth (SOC) media containing D-glutamate (0.005% v/v) for 1 hour at 37°C with gentle shaking. Cells were plated on both minimal media (12) and Luria Broth agar plates containing carbenicillin (35 µg/mL) and chloramphenicol (30 µg/mL) in the presence and absence of 50 µg/L anhydrotetracycline (AHT) and incubated at 37° C.

In order to isolate mutants with a higher rate of activity, DNA from the resultant colonies was isolated using the Qiagen Mini-Prep kit (Qiagen, Valencia, CA). 150 ng of this DNA was used to transform WM335 cells, and the cells were plated as previously described. This process was repeated several times until a specific mutant was enriched in the population.

Construction of Mutant Controls

In order to determine whether complementation was due to non-specific effects from expression of AEE, lysines 149, 151 and 247 were mutated to alanine using the Quickchange PCR mutagenesis method (Stratagene, La Jolla, CA). None of the mutants could complement the WM335 cells when co-transformed with NADGA.

Protein Purification

To purify NADGA in adequate amounts for biochemical assay, the gene was subcloned into the *BamHI* and *HindIII* sites of the expression vector pET-15b (Novagen, Madison, WI). The resulting plasmid, pNADGA-His, was co-transformed with plasmid encoding the chaperone GroEL into BL21 DE3 cells (Stratagene, La Jolla, CA). Cells were grown overnight and centrifuged, then resuspended in 30 mLs of binding buffer (10 mM Tris-HCl, pH 8 containing 5 mM imidazole, 0.5 M NaCl and 5 mM MgCl₂), then disrupted by passage through a microfluidizer. Following centrifugation, the supernatant was loaded onto a HisTrap 5 mL HP column (GE Biosciences, Piscataway, NJ) that had been equilibrated with binding buffer. Following a wash step with 5 column

volumes of buffer, the NADGA protein was eluted using a linear imidazole gradient (0 to 1 M imidazole in 10 mM Tris-HCl, pH 8, containing 5 mM MgCl₂ and 0.5 M NaCl). Fractions from the elution peak were exchanged into a buffer containing 10mM Kpi, 10% Glycerol, 2 mM DTT, pH 7 using an Amicon 15 30,000 MWCO filter (Millipore, Billerica, MA). The purity of the protein was verified using SDS-PAGE electrophoresis. The protein activity was also verified using a coupled spectrophotometric assay, in which purified NADGA protein was incubated with glutamate racemase, 3 U of glutamate dehydrogenase, and 0.2 U diaphorase in a buffer containing 20 mM N-acetyl-D/L-glutamate, 50 mM TEA, 5 mM NAD, 2.5 mM ADP, 0.65 mM INT pH 7.8 (13).

The wild-type *E. coli* AEE was purified as previously described (14).

Kinetic Characterization of L-Ala-D-glutamate Amidohydrolase Activity

The L-Ala-D-Glu amidohydrolase activity of NADGA was determined through modification of a coupled enzymatic assay that has been previously described (15). Briefly, NADGA protein was incubated with varying concentrations of L-Ala-D-Glu in a reaction buffer consisting of 50 mM Tris, pH 8.5, 0.67mM CoCl₂, 1.5mM NAD⁺, 1.5mM indonitrotetrazolium chloride (INT), 50 U L-alanine dehydrogenase, and 2 U of diaphorase. Hydrolysis of L-Ala-D-Glu by NADGA amidohydrolase was measured by following INT reduction at 500nm by UV spectrometer.

RESULTS AND DISCUSSION

Complementation of WM335 cells by AEE and NADGA

Co-transformation of *E. coli* AEE variant libraries and N-acetyl-D-glutamate amidohydrolase (NADGA) led to complementation and survival of WM335 (glutamate racemase deficient) cells when plated on minimal media containing low arginine. Sequencing of colonies verified that surviving colonies largely contained wild-type AEE, although some mutants also contained silent mutations. Complementation and survival was dependent on induction of variant expression, growth on minimal media, and

co-transformation with pNADGA and AEE, suggesting that both are necessary for D-Glutamate production.

Mutants with alanine substitutions at the catalytic lysines (149, 151, 247) of AEE were unable to complement when co-transformed with NADGA, suggesting that the plasmid-derived AEE is converting L-Ala-L-Glu to L-Ala-D-Glu, and confirming that the selected Ala-Glu epimerase reaction is mediated by the mechanism characteristic of other enolase superfamily members. Additionally, the AEE homologs from *B. subtilis* and BT1313 were able to complement at a much higher rate than the NAAAR from *Amycolaptosis* or an empty vector control (Table I), providing additional evidence in support of the hypothesis that the selection we observe is for this specific function

Promiscuous Activity by NADGA

Initially, we expected that the AEE component of the constructed pathway would be able to promiscuously racemize N-acetyl-L-glutamate to produce N-acetyl-D-glutamate for hydrolysis by NADGA. However, incubation of AEE with N-acetyl-L-glutamate failed to produce N-acetyl-D-glutamate (data not shown), suggesting an alternative mechanism for complementation. As we knew that the AEE component of the constructed pathway could epimerize dipeptides (14) to produce L-Ala-D-Glu, we suspected that production of D-glutamate (and subsequent complementation) was occurring via hydrolysis of the AEE reaction product by NADGA. Although NADGA has been annotated as specific for N-acetyl-D-Glutamate hydrolysis, we suspected that it promiscuously also hydrolyzed L-Ala-D-Glu as it is known to turn over other N-modified D-glutamate substrates (16). Testing this hypothesis reveals that the enzyme indeed is able to cleave L-Ala-D-Glu with an efficiency of $\sim 1.1 \times 10^5$ (Table II). This activity is metal dependent, consistent with the mechanism of other members of the amidohydrolase superfamily. This level of activity is approximately 19-fold higher than another characterized N-acetyl-D-glutamate amidohydrolase from *Bordetella bronchiseptica* that does the same reaction (Cummings J *et al.*, manuscript in preparation). The high

efficiency of NADGA for the L-Ala-D-Glu hydrolysis reaction suggests that it may represent a native reaction for the enzyme, rather than simply a promiscuous side reaction. The identification of this reactivity highlights the utility of the selection, as it may allow the identification of additional novel N-Ala-D-glutamate amidohydrolases when tested using this dual selection system.

Building a Selectable-Output Metabolic Pathway

In previous work, we identified operons that contained enolase and amidohydrolase superfamily members evolved to provide a pathway for scavenging L-amino acids (4). While these pathways may provide a selective advantage through the removal of toxic metabolic or antibiotic degradation products, they do not provide an output suitable for laboratory-based selection under standard growth conditions. However, because these pathways utilize the conserved chemical capabilities of the enolase and amidohydrolase superfamilies to perform each chemical step, we hypothesized that other superfamily member components could be recruited to change the specific reactivity at each step and thereby produce a novel output. Thus, by redesigning the pathways observed in nature, we have created a novel pathway that links the activity of both superfamily functions to the production of a selectable compound.

As shown in Figure 1B, our designed pathway converts L-Ala-L-Glu dipeptides to D-glutamate for use in cell wall synthesis, in contrast to previously identified pathways that convert modified D-amino acids to L-amino acids for use in metabolism (Figure 1A)(4). By inserting this pathway into a glutamate racemase knockout strain (10), we are able to select for the pathway output. As shown in our previous work, this also allows us to tune the stringency of the selection through product supplementation to enrich and recover weakly active enzyme variants (RAN, DST and PCB, manuscript in preparation). In addition, the complementation studies using both AEE and NADGA suggest that the selection can be used to identify both enolase superfamily (AEE activity) and/or amidohydrolase superfamily (the hydrolysis of L-Ala-D-Glu) function.

Applications for Functional Annotation and Protein Engineering

Based on results from directed evolution experiments, it has been suggested that new functions may evolve through mutations that have little effect on a native function but large effects in enhancing a promiscuous function (6, 17). The finding that some members of the enolase and amidohydrolase superfamilies are promiscuous for other reactions that are consistent with the fundamental capabilities of their respective superfamilies suggests that these enzymes may be especially useful as starting structures for engineering new reactions through the optimization of promiscuous capabilities. In addition to guiding engineering experiments, identification of promiscuous members may also inform studies of evolutionary relationships within each superfamily.

Despite the value of information regarding promiscuity, large scale mapping of the promiscuous activities of members has not been performed to date in part because both superfamilies are large and there are few high-throughput methods available for a comprehensive functional screening. Using the pathway selection we have developed, we expect that members of both superfamilies can now be assayed for the two specific functions required for complementation via this pathway, extending our ability to characterize promiscuous activity.

CONCLUSIONS

Although pathways have previously been engineered to provide a variety of useful outputs (18), we believe this to be the first example of using superfamily information to guide the choice of pathway components to tailor specific outputs. Following nature's example for evolving a pathway that depends on two functions that are each conserved among members of these two superfamilies, we have substituted other superfamily members to introduce variation at each step to create a novel pathway with a selectable output. This dual selection for function thereby provides an efficient screen for identification of enolase superfamily enzymes that perform the L-Ala-L-Glu epimerase

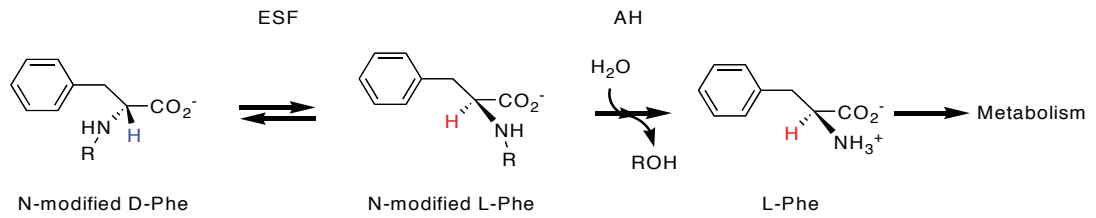
reaction, as well as amidohydrolase superfamily members that can hydrolyze L-Ala-D-Glu to produce D-glutamate. Using this selection, we have determined that the latter reaction is a promiscuous (but highly efficient) capability of the N-acetyl-D-glutamate amidohydrolase from *Achromobacter xylosoxydans*.

While our previous work has shown that the robust catalytic function of individual superfamily members is tied to conserved structural architectures (19), previous work (4) and the work reported suggest that this conservation may also hold at the higher level of complexity represented by metabolic pathways, where overall metabolic function is dependent on the conserved reactivity provided by the specific superfamily enzymes that are used as components. Based on this example, extending our analysis to characterize the association of other characterized superfamily members in metabolic pathways may provide insight for the prediction of metabolic functions in many organisms. Alternatively, the high-throughput nature of this selection for function can facilitate a survey of some aspects of promiscuity in both of these large superfamilies, contributing to a foundation for large-scale analysis of sequence and function relationships.

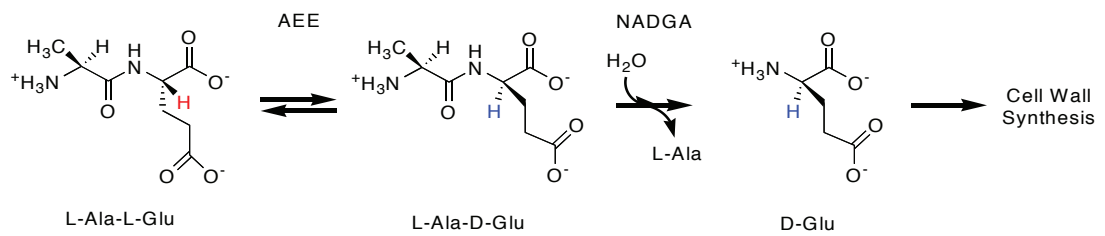
FIGURES:

Figure 1:

A.



B.



(A) Diagram showing reaction pathway of paired enolase superfamily (ESF) and amidohydrolase superfamily (AH) members to convert modified D-Phe to L-Phe (4). (B) Scheme showing turnover of L-Ala-L-Glu to yield D-Glu for cell wall synthesis

Table I: Survival rates for AEEs from *E. coli*, *B. subtilis*, and BT1313, NAAAR from *Amycolatopsis*, and pASK 7C plasmid (empty vector)

	Number of Colonies
<i>E. coli</i> AEE	~3800
<i>B. subtilis</i> AEE	~1000
BT1313 AEE	~700
<i>Amycolatopsis</i> NAAAR	17
pASK7C	3

Table II: Kinetic Constants for Hyrdolysis of L-Ala-D-Glu by *A. xylosoxydans* N-acetyl-D-Glutamate Amidohydrolase

	k_{cat} [s ⁻¹]	K_M [M]	k_{cat} / K_M [M ⁻¹ s ⁻¹]
N-acetyl-D-Glutamate amidohydrolase	120	1.1 x 10 ⁻³	1.1 x 10 ⁵

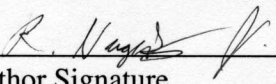
- (1) Gerlt, J. A., Babbitt, P. C., and Rayment, I. (2005) Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. *Arch Biochem Biophys* 433, 59-70.
- (2) Palmer, D. R., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. (1999) Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase. *Biochemistry* 38, 4252-4258.
- (3) Taylor Ringia, E. A., Garrett, J. B., Thoden, J. B., Holden, H. M., Rayment, I., and Gerlt, J. A. (2004) Evolution of enzymatic activity in the enolase superfamily: functional studies of the promiscuous o-succinylbenzoate synthase from *Amycolatopsis*. *Biochemistry* 43, 224-229.
- (4) Sakai, A., Xiang, D. F., Xu, C., Song, L., Yew, W. S., Raushel, F. M., and Gerlt, J. A. (2006) Evolution of enzymatic activities in the enolase superfamily: N-succinylamino acid racemase and a new pathway for the irreversible conversion of D- to L-amino acids. *Biochemistry* 45, 4455-4462.
- (5) Song, L., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Brown, S., Imker, H. J., Babbitt, P. C., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2007) Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat Chem Biol* 3, 486-491.
- (6) Schmidt, D. M., Mundorff, E. C., Dojka, M., Bermudez, E., Ness, J. E., Govindarajan, S., Babbitt, P. C., Minshull, J., and Gerlt, J. A. (2003) Evolutionary potential of (beta/alpha)₈-barrels: functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry* 42, 8387-8393.
- (7) Seibert, C. M., and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry* 44, 6383-6391.
- (8) Afriat, L., Roodveldt, C., Manco, G., and Tawfik, D. S. (2006) The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase. *Biochemistry* 45, 13677-13686.
- (9) Park, H. S., Nam, S. H., Lee, J. K., Yoon, C. N., Mannervik, B., Benkovic, S. J., and Kim, H. S. (2006) Design and evolution of new catalytic activity with an existing protein scaffold. *Science* 311, 535-538.
- (10) Dougherty, T. J., Thanassi, J. A., and Pucci, M. J. (1993) The *Escherichia coli* mutant requiring D-glutamic acid is the result of mutations in two distinct genetic loci. *J Bacteriol* 175, 111-116.
- (11) Wakayama, M., Katsuno, Y., Hayashi, S., Miyamoto, Y., Sakai, K., and Moriguchi, M. (1995) Cloning and sequencing of a gene encoding D-aminoacylase from *Alcaligenes xylosoxydans* subsp. *xylosoxydans* A-6 and expression of the gene in *Escherichia coli*. *Biosci Biotechnol Biochem* 59, 2115-2119.
- (12) Glansdorff, N. (1965) Topography of Cotransducible Arginine Mutations in *Escherichia Coli* K-12. *Genetics* 51, 167-179.
- (13) Gallo, K. A., and Knowles, J. R. (1993) Purification, cloning, and cofactor independence of glutamate racemase from *Lactobacillus*. *Biochemistry* 32, 3981-3990.
- (14) Schmidt, D. M., Hubbard, B. K., and Gerlt, J. A. (2001) Evolution of enzymatic

- activities in the enolase superfamily: functional assignment of unknown proteins in *Bacillus subtilis* and *Escherichia coli* as L-Ala-D/L-Glu epimerases. *Biochemistry* 40, 15707-15715.
- (15) Gulick, A. M., Schmidt, D. M., Gerlt, J. A., and Rayment, I. (2001) Evolution of enzymatic activities in the enolase superfamily: crystal structures of the L-Ala-D/L-Glu epimerases from *Escherichia coli* and *Bacillus subtilis*. *Biochemistry* 40, 15716-15724.
- (16) Sakai, K., Imamura, K., Sonoda, Y., Kido, H., and Moriguchi, M. (1991) Purification and characterization of N-acyl-D-glutamate deacylase from *Alcaligenes xylosoxydans* subsp. *xylosoxydans* A-6. *FEBS Lett* 289, 44-46.
- (17) Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q. G. S., Roodveldt, C., and Tawfik, D. S. (2005) The 'evolvability' of promiscuous protein functions. *Nat Genet* 37, 73-76.
- (18) Keasling, J. D. (2008) Synthetic biology for synthetic chemistry. *ACS Chem Biol* 3, 64-76.
- (19) Gerlt, J. A., and Babbitt, P. C. (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem* 70, 209-246.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

3/25/08
Date