Studies on Redesign and Solution Structure Determination of Nonribosomal Peptide Synthetases and Redox Regulation of Phosphatases

by

Cheng-Yu Chen

Department of Biochemistry
Duke University

Date:_______________________

Approved:

___________________________
Bruce R. Donald, Supervisor

___________________________
Pei Zhou

___________________________
Terrence Oas

___________________________
David Richardson

___________________________
Thomas LaBean

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2013
ABSTRACT

Studies on Redesign and Solution Structure Determination of Nonribosomal Peptide Synthetases and Redox Regulation of Phosphatases

by

Cheng-Yu Chen

Department of Biochemistry
Duke University

Date:_______________________

Approved:

___________________________

Bruce R. Donald, Supervisor

___________________________

Pei Zhou

___________________________

Terrence Oas

___________________________

David Richardson

___________________________

Thomas LaBean

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2013
Abstract

We present a computational structure-based redesign of the phenylalanine adenylation domain of the non-ribosomal peptide synthetase (NRPS) enzyme gramicidin S synthetase A (GrsA-PheA) for a set of non-cognate substrates for which the wild-type enzyme has little or virtually no specificity. Experimental validation of a set of top-ranked computationally-predicted enzyme mutants shows significant improvement in the specificity for the target substrates. We further present enhancements to the methodology for computational enzyme redesign that are experimentally shown to result in significant additional improvements in the target substrate specificity. The mutant with the highest activity for a non-cognate substrate exhibits 1/6 of the wild-type enzyme/wild-type substrate activity, further confirming the feasibility of our computational approach. Our results suggest that structure-based protein design can identify active mutants different from those selected by evolution.

Knowledge about the structures of individual domains and domain interactions can further our redesign of the NRPS enzymes for new bioactive nature product. So far, little structure information has been available for the auxiliary domains such as the epimerization domains and how they interact with the NRPS modules. Solution structure studies by nuclear magnetic resonance (NMR) provide advantages for understanding the dynamics of the domains and reveal active conformations that

iv
sometimes are not represented by the crystal structures. However, the large size of the NRPS proteins present challenges for structure studies in solution. In chapter 3, we study the solution structure of the 56 kDa epimerization domain of GrsA (GrsA-PheE) by NMR. We use multidimensional backbone resonance experiments as well as specific labeling strategy to assign the backbone resonances of GrsA-PheE. Secondary structures are determined by sets of residual dipolar couplings (RDCs) measured in multiple alignment media. To determine the global fold of the protein, we obtain long-range distance restraints by measuring the paramagnetic relaxation enhancements (PREs) from 15 site-directed spin labeling samples.

In chapter 4, we investigate the redox regulation of phosphatases. The activity levels of protein tyrosine phosphatases (PTPs) in cells are highly regulated in various ways including by phosphorylation, localization and protein-protein interaction. Additionally, redox-dependent modification has emerged as a critical part in attenuating PTPs activity in response to cellular stimuli. The tandem Src homology 2 domain-containing PTPs (SHPs) belong to the family of nonreceptor PTPs. The activity level of SHPs is highly regulated by interaction of SH2 domain, phosphorylation level of C-terminal tail and by reversible oxidation. In vivo evidence has shown the reversible oxidation of catalytic cysteine inhibits SHPs activity transiently as a result, affecting the phosphorylation level of its target proteins. In this chapter, we investigate in vitro the reversible oxidation of full-length and catalytic domain of SHP-1 and SHP-2 by using
kinetic measurements and mass spectrometry. We have confirmed the susceptibility of the active site cysteines of SHPs to oxidative inactivation, with rate constants for oxidation similar to other PTPs (2-10 M⁻¹s⁻¹). Both SHP-1 and SHP-2 can be reduced and reactivated with the reductants DTT and glutathione, whereas only the catalytic domain of SHP-2 is subject to reactivation by thioredoxin. Unlike PTPs whose oxidation contains a catalytic cysteine disulfide bonding to a backdoor cysteine or forms a sulfenylamide bonding to nearby backbone nitrogen, we have found that in the reversibly oxidized SHPs, the catalytic cysteines is re-reduced while two conserved backdoor cysteines form a disulfide linkage. Knocking out either of the backdoor cysteine preserves the reversibility of the oxidized SHPs with a disulfide formation between the catalytic cysteine and the remaining backdoor cysteine. However, removal of both backdoor cysteines leads to irreversible oxidative inactivation, demonstrating that these two cysteines are necessary and sufficient for ensuring reversible oxidation of the SHPs. Our results extend the mechanisms by which redox regulation of PTPs is used to modulate intracellular signaling pathways.
Contents

Abstract ................................................................................................................................................. iv

List of Tables ............................................................................................................................................. xi

List of Figures ............................................................................................................................................ xii

Acknowledgements ..................................................................................................................................... xiv

1. Introduction ........................................................................................................................................... 1
   1.1 Nonribosomal Peptide Synthetase and Computational Redesign of the Enzyme.. 1
      1.1.1 Nonribosomal Peptide Synthetase ................................................................. 1
      1.1.2 Computational Structure-Based Redesign ...................................................... 3
   1.2 Structure Determination of the Epimerase Domain of GrsA ............................... 5
      1.2.1 Structures and Dynamics of NRPSs .............................................................. 5
      1.2.2 Structure Determination of NRPS by NMR ...................................................... 7
   1.3 The Tandem Src homology 2 (SH2) Domain-containing Protein-tyrosine
       Phosphatases (SHPs) ................................................................................................. 11
      1.1.1 Regulation of SHPs by Protein-protein Interaction ........................................ 15
      1.1.2 Redox Regulation of PTPs ............................................................................... 16

2. Redesign and Kinetic Studies of GrsA-PheA ................................................................................. 19
   2.1 Introduction ................................................................................................................................. 19
   2.2 Experimental materials and methods ....................................................................................... 23
      2.2.1 Materials ....................................................................................................................... 23
      2.2.2 Mutagenesis of PheA ................................................................................................. 23
      2.2.3 Expression and purification of PheA ........................................................................... 24
2.2.4 PPI Release Assay ........................................................................................................25
2.2.5 Stopped-Flow Tryptophan Fluorescence Quenching ..............................................26
2.2.6 Free Energy Calculation .............................................................................................27
2.3 Computational Method .................................................................................................29
  2.3.1 Active Site Mutation Prediction ...............................................................................29
  2.3.2 Bolstering mutation prediction ................................................................................32
2.4 Results ..........................................................................................................................34
  2.4.1 Steady-state Kinetic Analysis ...................................................................................35
  2.4.2 Redesign for Leu ......................................................................................................41
  2.4.3 Redesign for Charged Amino Acids .......................................................................47
  2.4.4 Other Experimentally-tested Mutants ....................................................................48
  2.4.5 Pre-steady-State Binding Analysis .........................................................................51
  2.4.6 Structural Analysis ..................................................................................................53
2.5 Discussion .....................................................................................................................57
3. NMR Studies of the Epimerization Domain of GrsA .......................................................63
  3.1 Introduction ..................................................................................................................63
  3.2 Experimental procedures .............................................................................................66
    3.2.1 Materials ................................................................................................................66
    3.2.2 Cloning and Mutagenesis of PheE ....................................................................67
    3.2.3 Protein Expression, Purification and Sample Preparation for NMR ..............69
    3.2.4 NMR Spectroscopy and Backbone Resonance Assignment ............................70
    3.2.5 RDC Sample Preparation and Measurement ....................................................71
3.2.6 Site-directed Spin Labeling and PRE Measurement ...........................................73
3.2.7 Structure Computation by RDC-ANALYTIC ..................................................75
3.3 Results ..................................................................................................................76
3.3.1 Backbone Resonance Assignment ....................................................................76
3.3.2 Residual Dipolar Coupling Analysis ..................................................................80
3.3.3 Paramagnetic Relaxation Enhancements .........................................................84
3.4 Discussion .............................................................................................................87
4. Redox Regulation of SHPs by A Novel Backdoor-backdoor Cysteine Disulfide .........90
4.1 Introduction ..........................................................................................................90
4.2 Materials & Methods ..........................................................................................93
4.2.1 Materials .........................................................................................................93
4.2.2 Cloning, Expression, and Purification of Full-length SHP-1 and SHP-2 ............93
4.2.3 Cloning, Expression, and Purification of Catalytic Domains of SHP-1 and SHP-2 .........................................................................................................................94
4.2.4 Measurement of the Inactivation Rate by H$_2$O$_2$ .............................................96
4.2.5 Measurement of Reactivation Rate by DTT, GSH and TR/TRR .......................97
4.2.6 MALDI-MS Spectrometry .................................................................................98
4.3 Results .................................................................................................................99
4.3.1 Inactivation of Full-length and the Catalytic Domain of SHPs by H$_2$O$_2$ ..........99
4.3.2 Reactivation of the SHPs by DTT, GSH and TR/TRR .......................................102
4.3.3 Identification of the Backdoor Cysteine Residues in the Catalytic Domain of the SHPs ..................................................................................................................104
4.3.4 Identification of Disulfide Intermediates by Mutagenesis of the Backdoor Cysteines

4.3.5 The Two Backdoor Cysteines are Necessary and Sufficient for the Protection of the Catalytic Cysteine

4.4 Discussion

5. Summary and Further Directions

5.1 Redesign and Kinetic Analysis of GrsA-PheA

5.1.1 Enzymatic Mechanism of GrsA-PheA

5.2 Interdomain Interaction Studies of GrsA-PheE

5.3 Redox Regulation of SHPs

5.3.1 pKa of the Active Site Cysteines of SHPs and the Conformation of Reversible Oxidation State

Bibliography

Biography
List of Tables

Table 2.1: Mutagenesis primers for redesign of PheA ................................................................. 24

Table 2.2: Top ten 2-point active site mutants predicted by $K^*$ for each of the five target substrates (Leu, Arg, Glu, Lys, and Asp) ........................................................................................................ 32

Table 2.3: Top bolstering mutations from the 3-point MinDEE/A* mutation search ...... 34

Table 2.4: Mutant enzymes with experimentally-observed specificity ($k_{cat}/K_M$, $k_{cat}$, and $K_M$ for a target substrate) ........................................................................................................ 41

Table 2.5: Other experimentally-tested mutants ........................................................................... 50

Table 2.6: Rate constant and free energy change of binding with wild-type and mutant PheA .......................................................................................................................... 53

Table 2.7: MolProbity analysis for the five lowest-energy T278D and T278D/A301G structures from the respective $K^*$ ensembles with Arg as substrate .................................................................... 55

Table 2.8: Kinetic constants and free energy difference for WT, T278L, A301G and T278L/A301G PheA .................................................................................................................. 62

Table 3.1: Cloning and Mutagenesis primers for PheE ................................................................. 68

Table 4.1: Rate of Oxidation and Reactivation for SHP-1 and SHP-2 ........................................... 104
List of Figures

Figure 1.1: Schematic of a nonribosomal peptide synthetase (NRPS) assembly line........ 2

Figure 1.2: Domain structure of the human SHPs......................................................... 14

Figure 2.1: Schematic of the module structures of gramicidin S synthetase. ................. 20

Figure 2.2: SDS-PAGE showing the homogeneity of the WT and mutant PheA............... 36

Figure 2.3: Guanidine-HCl titration of wild-type and mutant PheA proteins.................... 36

Figure 2.4: Representative steady-state kinetics curves for: WT PheA with Phe (A) and
Leu (B), A301G/A322V with Phe (C) and Leu (D), and T278M/A301G with Phe (E) and
Leu (F).................................................................................................................. 37

Figure 2.5: Representative steady-state kinetics curves for: T278L/A301G with Phe (A)
and Leu (B), T278L/A301G/S447N with Phe (C) and Leu (D), and V187L/T278L/A301G
with Phe (E) and Leu (F)........................................................................................................... 38

Figure 2.6: Representative steady-state kinetics curves (black) for I277L/T278L/A301G
with Phe (A) and Leu (B), I277L/T278L/A301G/S447N with Phe (C) and Leu (D), and
T278H/A301G with Phe (E) and Glu (F). .................................................................................. 39

Figure 2.7: Representative steady-state kinetics curves (black) for T278K/A301G with Phe
(A) and Asp (B), and T278D/A301G with Phe (C), Lys (D), and Arg (E)....................... 40

Figure 2.8: $K^*$-predicted structure of the lowest-energy T278L/A301G conformation with
Leu as substrate................................................................................................................. 44

Figure 2.9: Specificity ratio ($k_{cat}/K_M$)$_{Leu}/(k_{cat}/K_M)_{Phe}$ for WT and mutant PheA in the Leu
redesigns. .......................................................................................................................... 45

Figure 2.10: Fluorescence quenching of WT PheA by phenylalanine in the presence of
ATP and Mg$^{2+}$. .................................................................................................................. 52

Figure 2.11: $K^*$-predicted structure of the second lowest-energy T278D/A301G
conformation with Arg as substrate.............................................................................. 56
Figure 2.12: The lowest-energy S447N conformation predicted by MinDEE/A*..............59

Figure 3.1: Framework of RDC-ANALYTIC.................................................................66

Figure 3.2: $^{1}$H$^{15}$N-HSQC-TROSY spectrum of GrsA-PheE........................................78

Figure 3.3: Backbone resonance assignment using 4D HNCCb spectrum ....................79

Figure 3.4: Comparison of the model and predicted secondary structure elements for PheE. .................................................................................................................................80

Figure 3.5: Back-calculated RDCs from the model vs experimental RDCs. .................82

Figure 3.6: Distribution of $^{1}$H-$^{15}$N$^{15}$ RDCs collected in phage and acid gel. ........83

Figure 3.7: Bar graphs of intensity ratio vs primary sequence for spin label position Cys323 (orange star). ........................................................................................................85

Figure 3.8: Model distances vs experimental distances...............................................86

Figure 4.1: Backdoor cysteines of the SHPs.................................................................92

Figure 4.2: Time dependence of inactivation of SHPs by H$_2$O$_2$.................................101

Figure 4.3: Time dependence of reactivation of SHPs by DTT...............................103

Figure 4.4: MALDI-MS of reduced and oxidized ΔSHP-1 and ΔSHP-2. ...............106

Figure 4.5: MALDI-MS of reduced and oxidized mutants of ΔSHP-1......111

Figure 5.1: Reaction pathway for the adenylation reaction of NRPS. .......................120

Figure 5.2: Active site of SHP-1. Green: Full-length, Cyan: Catalytic domain...........123
Acknowledgements

The dissertation studies would not have been possible without the support from numerous people. First, I would like to thank my advisor Bruce R. Donald who has guided me to the field of Computational Biology, and supported me in every possible way throughout my graduate career. The GrsA-PheA redesign project would not have been possible without Ivelin Georgiev. I thank Ivelin for his collaboration and great work on the development of the MinDEE and $K^*$ algorithms. I would like to thank Pei Zhou for his collaboration on the NMR project. Pei is not only a great collaborator but plays an important role on advising me through all of the NMR experiments. His academic advice and example have played a vital role throughout my graduate career. Thanks to Chittaranjan Tripathy (Chittu) for his hard work and great collaboration on the development of RDC-ANALYTIC. I thank people in the Donald lab who have contributed ideas, support, and feedback to my works and presentations. Especially I would like to thank John MacMaster for his endless support and help in the wet lab. The Phosphatase project would not have been possible without Johannes Rudolph. I want to thank Johannes for guiding me to the enzyme kinetics, and for supporting me throughout the project.

I am very grateful to my committee members, Terrence Oas, David Richardson, Pei Zhou, and Thosmas LaBean for their support and advices in every committee meeting. They have provided ideas and valuable feedback on the experiments, and
suggested improvements for my dissertation. I also would like to thank people in the NMR Center at Duke, Ron Venters, Tony Ribiero and Don Mika for their help and advices on the NMR experiments.

Getting a Ph.D. has been a long road with a serious commitment to the research. The graduate life would not have been so joyful and meaningful without friends and family. I want to thank all the Taiwanese friends and friends I have made at Duke during these years. I am very grateful and lucky to have a lovely family. I want to thank my wife Kaifeng Lisa Lin. Lisa is not only a mother of two children, she is also a great immunologist. Throughout these years, Lisa has not only brought happiness to my life, she is also a good friend and a life advisor. Finally I want to thank my parent Chio-Fu Chen, Chun-Mei Lin and my big brother Kuan-Hung Chen, who have supported me on everything throughout my life.
1. Introduction

1.1 Nonribosomal Peptide Synthetase and Computational Redesign of the Enzyme

1.1.1 Nonribosomal Peptide Synthetase

Over decades, natural products produced by microorganisms have received a tremendous amount of attention due to their broad spectrum of biological activity that provides valuable therapeutic potential. Among them is the large subclass of non-ribosomal peptides (NRPs), which have been applied as therapeutic drugs such as vancomycin (antibacterial), bleomycin (antitumor) and cyclosporine (immunosuppressant). NRPs are synthesized by a family of multi-domain enzymes called non-ribosomal peptide synthetases (NRPSs) (1). These multi-function enzymes are structurally organized into modules which serve as templates for incorporation of amino acid building blocks and their derivatives, and catalyze the formation of peptide bonds. A typical NRPS module comprises three core domains: adenylation (A, ~60 kDa), peptidyl carrier protein (PCP, ~9 kDa), and condensation (C, ~45 kDa) domains (Figure 1.1). The A domain recognizes and activates the amino acid substrate by formation of an aminoacyl adenylate intermediate through hydrolysis of ATP. The aminoacyl adenylate intermediate is transferred to the terminal thiol group of 4’-phosphopantetheine (Ppant) cofactor on the PCP domain via a thioester linkage. The Ppant cofactor is ~20 Å in length and functions along with the PCP domain to provide the flexibility for moving the bound substrate between domains. The C domain accommodates the upstream peptidyl
unit and the downstream aminoacyl unit and catalyzes the condensation reaction for peptide bond formation. Released from the C domain, the peptide remains bound to the downstream PCP domain and is carried to the next C-terminal module where the peptide continues to be elongated until it reaches the termination module. In addition to the core domains, there are auxiliary domains that perform extra modifications such as epimerization, cyclization, methylation, and oxidation, to the peptides. Peptide elongation stops and the peptide remains bound on the termination module where peptides usually undergo macrocyclization and are released as final products by the thioesterase (TE) domain. The catalytic steps of NRPSs occur in a fashion such as an assembly line. The number and the organization of modules as well as amino acid compositions determine the complexity of the peptide products.

Figure 1.1: Schematic of a nonribosomal peptide synthetase (NRPS) assembly line. Figure is adapted from (2).

The modular architecture of NRPSs enables microbes to have diverse metabolites and it also suggests ways to create “unnatural” products through protein engineering. It is well known that substrate specificity of NRPSs is determined mainly by the A
domain. Therefore, to create an “unnatural” NRP from NRPS, people have replaced or swapped the A domains of different substrate specificities to produce chimeric constructs. Although with limited success, the resulting proteins often suffer from a dramatic reduction in overall activity due to incorrect folding of the entire protein or incompatibility between domains. The result reveals the importance of inter-domain communication which is still poorly understood. Directed evolution has been applied to improve the efficiency of the hybrid system with ~10-fold increase in catalytic activity (3). An alternate way to change the substrate specificity of NRPS is to reengineer the substrate recognition site of the A domain. The substrate binding pocket of the A domain from gramicidin S synthetase A (GrsA) has been identified from the crystallographic studies (4). Analyzing the residues in the binding pocket and using amino acid sequence alignment of A domains, Stachelhaus et al. identified a set of residues that contribute to the substrate specificity (5). Mutations of these residues have resulted in improving substrate specificity for non-cognate substrates; however, the improvement is relatively small and limited to a few types of amino acid substrates, while the mutant enzymes still prefer the wild-type substrate.

1.1.2 Computational Structure-Based Redesign

Despite recent successes, enzyme design has posed significant challenges for both computational and purely experimental approaches. Until recently, computational enzyme design approaches have met with limited success (6-8), making experimental
techniques, such as directed evolution, the preferred method for designing new enzymes (3, 9-11). Advances in both algorithms and modeling recently resulted in the first computationally-driven de novo structure-based design of active enzymes (12, 13). A fully-automated computational approach that is applicable to general enzyme design problems, however, is yet to be developed.

A major advantage of computational structure-based protein design over the purely experimental approaches lies in its ability to efficiently (and inexpensively) search a significantly larger portion of the available space of candidate mutations. Unfortunately, computational approaches must rely on simplified models that only approximate real proteins and their interactions. Among the typical simplifying model assumptions are: a rigid protein backbone, a rotamer library of discrete side-chain conformations (14, 15), and a pairwise energy function (16, 17). To improve the accuracy of the model, some recent advances in computational protein design have incorporated continuous flexible rotamers (18) and continuous (19) or discrete (20, 21) backbone flexibility. More accurate energy functions are sometimes used as a post-processing step to re-evaluate and re-rank the top-scoring predictions from the initial model (22).

Despite the imperfections of the underlying models, the computational approaches have yielded successful designs of proteins with improved target properties (7, 22-25). Designing for enzyme activity, however, has proven to be far more elusive. The difficulty of designing enzymes via computational methods can be attributed to the
more poorly understood catalytic enzyme machinery as well as the increased inability of
the simplified models to accurately represent the catalytically-relevant interactions (and
especially the high-energy transition states) accurately.

In Chapter 2, I describe how we used a novel protein redesign algorithm, $K^*$, to
redesign the adenylation domain of GrsA (GrsA-PheA). The redesigned GrsA-PheA
mutants have switched substrate specificity to non-cognate amino acid Leu. We also
redesigned mutants of GrsA-PheA to catalyze charged amino acids Arg, Glu, Lys, and
Asp, which was not possible for the wild-type enzyme.

**1.2 Structure Determination of the Epimerase Domain of GrsA**

**1.2.1 Structures and Dynamics of NRPSs**

Structures of both individual NRPS domains and multi-domain constructs have
been determined by X-ray crystallography and NMR. The structures include the
adenylation domains from gramicidin S synthetase A (PheA) (4), bacillibactin (DhbE)
(26), acinetobactin (27) and the eukaryote N. lolii siderophore biosynthetic pathways(28);
the PCP domains from tyrocidine (PCP-C didomain) (29), enterobactin (PCP-TE
didomian) (30) and surfactin biosynthetic pathways; the condensation domains from
vibriobactin (VibH) (31), tyrocidine (PCP-C didomain)(29) biosynthetic pathways; the
thioesterase domains from surfactin, fengycin, and enterobactin (PCP-TE didomain)
biosynthetic pathways (32-34); and the epimerization (E) domain from the initiation
module of tyrocidine synthetase. These structures not only provide atomic details of
substrate binding and active site conformations, they also reveal dynamic nature of NRPSs. The A domain consists of a large N-terminal domain and a small C-terminal domain, with the active site located at the junction of these two sub-domains. Crystal structure of the A domain of gramicidin S synthetase A (GrsA-PheA) showed that the active site of the holo-form structure accommodates a phenylalanine and an AMP molecule. By comparing the crystal structures of the holo-GrsA-PheA domain and other adenylate-forming enzymes including the aceteyl-CoA synthetase (Acs) and the D-alanyl carrier protein ligase (DltA), it was suggested that in the second half reaction, the C-terminal domain undergoes a ~140° rotation in order for the phosphopantetheinyl arm to attack the adenylated intermediate. The structural heterogeneity was also observed in the PCP domains. Solution structures of tyrocidine PCP domain revealed two conformations: the A and H states when the domain was in apo and holo (loaded with phosphopantetheine cofactor) form respectively (35). An intermediate A/H state was also observed in both apo and holo forms of the PCP domain. NMR analysis of these states suggested a slow exchange between the A and H states of the PCP domain. Further NMR titration experiments using the phosphopantetheine transferase (PPTase) enzyme, Sfp and the upstream C domain showed that the PCP domain uses different state conformers for interaction with these two domains (34).

The structures of multi-domain constructs provided molecular details on understanding inter-domain interactions of NRPS. The X-ray structure of the
termination module from tyrocidine synthetase (C-A-PCP-TE, 144 kDa) provides a complete view of inter-domain interactions within a module of NRPS (32). In this structure, the PCP domain adopts an A/H conformation interacting closely with the C and TE domains. The long distance (~57 Å) between the active sites of the C and the A domain in the structure comparing to the length of the Ppant arm (~20 Å) suggests a large movement and reorientation of the PCP domain is required for transferring the aminoacyl substrate between the A and C domains. The dynamic inter-domain motions were also revealed in the solution structure of the PCP-TE didomain from the E. coli enterobactin synthetase, EntF (34). The NMR analysis of the PCP-TE showed that two α-helices on the TE domain (α4 - α5) hovered between the active sites of PCP and TE domains to allow access by the 4′-phosphopantetheine arm. Titration of the phosphopantetheinyl transferase enzyme, Sfp and the EntF C domain to the PCP-TE showed different binding interfaces on the PCP domain.

1.2.2 Structure Determination of NRPS by NMR

The X-ray crystal structures of NRPSs have provided molecular details on many individual domains and several multidomain constructs; however, in order for crystallization, many of these structures require inactive mutations or are trapped in an inactive conformation by inhibitors. For example, the termination module of the surfactin biosynthetic pathway was solved by using a Ser to Ala mutation on the PCP domain that results in a removal of the attachment site for the Ppant cofactor (32). The
crystallized mutant protein has the PCP domain adopting an H/A conformation. Also in the crystal structure of the epimerization domain of tyrocidine synthesis (PDBid=2XHG), a helical segment at C-terminus was removed. This segment was proposed to mediate domain-domain recognition between modules (36). Complement to X-ray crystallography, nuclear magnetic resonance provides advantages on characterizing dynamic properties of proteins and identifying interfaces between domains. Because proteins are in solutions, active conformations can often be identified and studied. Solution structure studies of NRPSs by NMR suggested intra- and inter-domain mobility that plays important part for a successful product release (34).

In NRPSs except for the small PCP domain, most domains and multi-domain constructs are large proteins in solution, which make structure determination by NMR a challenge. Over decades, advance NMR technologies such as TROSY (transverse relaxation-optimized spectroscopy) (37), RDC (residual dipolar coupling) (38) and PRE (paramagnetic relaxation enhancement) (39) as well as labeling strategies (40, 41) have helped push the size limited up to ~100 kDa. Traditional NMR structure determination approaches rely heavily on a sufficient number of accurate distance restraints from nuclear Overhauser effect (NOE). Assignment of NOE data is considered as a major bottleneck in protein NMR structure determination because the data often contains ambiguities due to chemical shift degeneracy. For small size proteins, the problem can often be resolved by using multidimensional NOESY experiments. However as the size
of protein goes larger, the problem of assignment ambiguities becomes more serious while the decrease in NMR signal due to shorter transverse relaxation time makes the data interpretation difficult. To overcome these problems, proteins with the size of more than 25 kDa usually require the use of deuteration (42). High level of deuteration, however, eliminates a majority part of NOE restraints that are required for a correct global fold determination. The use of selective labeling on methyl groups reintroduces side chain methyl protons in Val, Leu, and Ile residues, hence regains a limited number of NOEs (41). Global fold can be determined by merely using these NOEs, however, the resulting structures are usually lower resolution. The number of NOEs from the labeling strategy also highly depends on the composition and location of the hydrophobic amino acids. Also in traditional NMR approaches, several iterations of structure calculation using stochastic techniques such as simulated annealing (SA) and molecular dynamics (MD) are required to obtain a better NOE assignment and gradually improve the structure (43). However, these stochastic techniques can be trapped into local minima and result in missing of the global minimum structure solution which can subsequently lead to incorrect NOE assignments.

In contrast to NOE restraints, residual dipolar couplings (RDCs) provide global orientational restraints on internuclear vectors such as backbone NH and CaC' with respect to a global alignment frame. Incorporation of RDCs has become a routine protocol in the refinement stage at the end of structure determination by NMR.
However, this approach doesn’t directly address the issues arising from the ambiguities of NOE restraints. RDC data can be measured with high precision and assigned much faster than NOEs. To complement the drawback of using NOE restraints for structure determination, a polynomial time de novo algorithm, called RDC-EXACT, has been proposed to compute high-resolution backbone structures of secondary structure elements (SSEs) using a minimum amount of RDC data (44-46).

Paramagnetic relaxation enhancement (PRE) occurs in protein samples containing paramagnetic electrons which induce a distance-dependent line broadening of nuclear magnetic resonances. The PRE effect can be observed by site-directed spin labeling (SDSL) approach which introduces paramagnetic nitroxide centers into proteins (47). The measured PREs are used to calculate the distances from the unpaired electron of the nitroxide to the affected nuclei. The distances arising from PREs depend on the type of spin labeling. In general, PREs given by nitroxide labeling provide distance restraints of 15-24 Å, which can complement the shorter range of NOE restraints. PRE has been used for global fold determination of a medium size protein and a membrane protein with a limited number of NOE restraints (48, 49).

In chapter 3, we present the solution structure studies of the epimerization domain of gramicidin S synthetase A. We reported a complete backbone resonance assignment of the 488 residues protein. To determine the global fold of the protein, we
measured RDCs for secondary structure determination and PREs for packing the secondary structures.

1.3 The Tandem Src homology 2 (SH2) Domain-containing Protein-tyrosine Phosphatases (SHPs)

In eukaryotic cells, many essential cellular processes such as proliferation, differentiation, migration, and immunogenic responses are regulated by protein tyrosine phosphorylation. Aberrant protein tyrosine phosphorylation has been linked to various diseases such as cancer, diabetes, neurological disorders, and immune diseases (50). In cells, the level of phospho-tyrosine is mainly controlled by two classes of enzymes, protein-tyrosine kinases (PTKs) and protein-tyrosine phosphatases (PTPs). The activity of these enzymes is tightly regulated at multiple levels: spatio-temporal expression, subcellular localization and post-translational modification. Over past few decades, studies have focused on understanding the regulation of PTKs, in particular the ligand-induced activation of receptor tyrosine kinases (RTKs) or receptor-associated PTKs (51). PTPs which function as counterpart of PTKs have not received the same amount of attention partly due to lack of clear physiological substrates which makes identification and characterization of PTPs difficult. Human genomic and proteomic studies have helped the discoveries of several important PTPs including PTP1B (52), the receptor-like PTPs (53), the cell cycle regulator Cdc25 phosphatases (54) and the tumor suppressor PTEN (55). These findings suggest the importance of PTPs is involved in multiple levels of cellular signaling and in balancing the level of tyrosine phosphorylation.
In humans, the PTP family constitutes a big family of ~107 genes (56). Except for the small subfamily of Asp-dependent PTPs, the majority of PTPs are Cys-dependent enzymes. The PTP family can be broadly subdivided into three groups: receptor-like PTPs (RPTP), cytoplasmic PTPs and dual specificity phosphatase (DSPs), and are characterized by the presence of a conserved active-site signature motif (H-C-X₅-R; single amino acid code wherein X represents any amino acid) in their catalytic PTP domains (57). In the motif, the invariant Cys and Arg residues are essential for catalysis. In the active form of PTPs, the active site motif adopts a conformation that favors the formation of Cys thiolate. The thiolate side chain is often stabilized by a nearby Ser residue as found in PTP1b (52). The typical Cys-mediated catalysis involves a 2-step mechanism (Scheme 1.1). In the first step, the phosphate on the pTyr residue of the substrate undergoes nucleophilic attack by the thiolate sulphur of the catalytic Cys. The phosphate is further stabilized by forming hydrogen bonds to the Arg side chain and the surrounding backbone amides. Substrate binding induces a conformational change that brings the invariant Asp residue on the WPD loop close to the pTyr side chain of the substrate. The Asp functions as a general acid that protonates the tyrosyl leaving group of the substrate (58). In the second step, the deprotonated Asp functions as a general base that facilitates hydrolysis of the cysteinyl-phosphate intermediate and brings the catalytic Cys back to the thiolate form.
In addition to the catalytic domain, most PTPs have additional domains that serve as regulators for catalytic activity or as adaptors for localization of PTPs. In chapter 2, we focus our studies on a specific PTP family, the tandem Src homology 2 (SH2) domain-containing protein-tyrosine phosphatases (SHPs), which belong to the group of cytoplasmic PTPs. This family includes SHP-1 and SHP-2 in humans and is characterized by two tandem SH2 domains at the N-terminus (N-SH2 and C-SH2) followed by a catalytic PTP domain and an inhibitory C-terminus (Figure 1.1) (59). In addition to the PTP domain which recognizes and dephosphorylates several protein substrates, the two SH2 domains bind to phospho-tyrosyl proteins and other adapter proteins and thereby contribute to the regulation of the SHPs through localization and substrate binding. The N-SH2 domain also contributes to basal auto-inhibition of the
SHPs. Crystal structure studies of SHP-1 and SHP-2 have shown that the N-SH2 domain adopts an auto-inhibitory conformation by blocking the active site pocket (60-62). Further studies showed that the PTP activity can be recovered by binding of the SH2 domains to a phospho-tyrosyl peptide or by deleting the SH2 domains (62). Mutations of residues on the interface of the N-SH2 and PTP domains increase the level of basal activity. The C-terminal domains of SHPs contain several tyrosine and serine residues whose phosphorylations have been proposed to participate in the regulation of PTP activity (63). However the actual function of this poorly conserved domain and their physiological binding partners are still unclear.

Figure 1.2: Domain structure of the human SHPs. Percent identities, the catalytic cysteine (in red color) and catalytic aspartic acid (in blue color) are shown.
1.1.1 Regulation of SHPs by Protein-protein Interaction

Despite that SHP-1 and SHP-2 share a high sequence homology (57% overall) and similar structures (rmsd=1.1 Å in the catalytic domains), they appear to play opposite regulatory roles in signaling pathways (59). SHP-1 is mainly expressed in hematopoietic cells, whereas SHP-2 is expressed ubiquitously. Different expression patterns might explain their different roles however, some cells such as lymphocytes with high expression levels of both proteins display dramatically different outcomes when either of the SHPs loses its normal function (64). In general, SHP-1 is a negative regulator for several intracellular transmembrane signaling receptors including growth factor receptors (e.g. c-kit, CSF-1, TrkA and EGF), cytokine receptors (e.g. Epo-R, IFNα/β-R, IL-3R and IL-2R) and receptor tyrosine kinases (RTKs) (65, 66). In contrast, SHP-2 acts as a positive regulator in the Ras-extracellular signal-regulated kinase (Ras-ERK), T cell receptor (TCR) and the cytokine receptor signaling pathways (67). SHPs are recruited to these receptors by binding of the SH2 domains to the cytosolic domain of the receptors or through an adaptor protein. Peptide binding studies have found that the N-SH2 and the C-SH2 domains preferentially bind to peptides containing the immuno-receptor tyrosine-based inhibitory motifs (ITIMs) (68, 69). Upon binding to the receptors, SHPs become activated by opening the catalytic pocket which then specifically dephosphorylates a variety of substrates. Studies using chimeric SH2 domains and catalytic domains have shown that the catalytic domains from the two SHPs are not
interchangeable (70). This study suggested that the catalytic domains contribute to the substrate specificity of SHPs in addition to their SH2 domains. Interestingly, both SHP-1 and SHP-2 have been reported to dephosphorylate the cytokine receptor-associated kinase JAK2, however, dephosphorylation by the two SHPs results in dramatically different consequences for downstream signaling (71). Dephosphorylation of JAK2 by SHP-1 deactivates JAK2 and terminates the phospho-tyrosine signal while dephosphorylation by SHP-2 prevents degradation of JAK2 and sustains the phospho-tyrosine signal (72). How these two SHPs seem to both selectively dephosphorylate the JAK2 kinase with such different outcomes may be in part a result of their specific and unique interactions with protein substrates.

1.1.2 Redox Regulation of PTPs

Recently, there has been increasing evidence demonstrating a more primary mode for regulation of PTP activity involving reversible oxidation by reactive oxygen species (ROS) such as hydrogen peroxide (H$_2$O$_2$) (73-76). In PTPs, the active site conformation and the surrounding environment lower the pK$_a$ of the catalytic cysteine thiol to a reactive thiolate (Cys-S), which not only provides the catalytic driving force but also becomes highly susceptible to oxidative inactivation. Extracellular stimuli, including hormones, growth factors, and cytokines induce activation of cell surface receptors which, in addition to the primary response (e.g., receptor dimerization and downstream signaling), can lead to production of intracellular ROS to concentrations as
high as 0.1-1 mM (77, 78). The ROS is further consumed by other reducing mechanisms rapidly. However, the amount is sufficient to oxidize the catalytic cysteine thiolate.

Oxidation of this highly reactive catalytic cysteine yields a reversibly oxidized sulfenic acid (Cys-SO) that is susceptible to further irreversible oxidation to the sulfinic (Cys-SO₂⁻) or sulfonic (Cys-SO₃⁻) species. Because the inactive yet reversibly oxidized enzyme serves as a switch in controlling signaling pathways, cysteine-based phosphatases have evolved mechanisms to preserve the reversible oxidation state. For many phosphatases, including the low molecular weight phosphatase (LMW) (79), the cell cycle phosphatases Cdc25 (80-83), phosphatase and tensin homolog (PTEN) (84), and kinase associated phosphatase (KAP) (85), biochemical and structural studies have shown that the reversibly inactivated state can be stabilized by transferring the oxidation at the active site cysteine to an intracellular disulfide bond using a backdoor cysteine. On the other hand, for PTP1B (86, 87) and PTPα (88), the initial oxidation of the active site cysteine is trapped as a reversible sulfenylamide bond to the adjacent amino acid. The disulfide and sulfenylamide bonds are resistant to further oxidation under physiological levels of ROS and can be re-reduced to the active thiolate by reductants such as dithiothreitol (DTT), glutathione (GSH) and/or thioredoxin/thioredoxin reductase (TR/TRR). SHPs share the conserved PTP active site motif and were reported to participate in several redox regulatory signaling pathways. In chapter 2, we investigate kinetically the difference of the two SHPs in response to oxidation and identified a novel
protection mechanism using disulfide bond formed between two conserved backdoor cysteines.

This dissertation can be presented with respect to our peer-reviewed publications.

Chapter 2 is based on the following publications:


Chapter 4 is based on the following publication:


Finally, chapter 3 presents unpublished material.
2. Redesign and Kinetic Studies of GrsA-PheA

This chapter has been adapted mainly from the following published manuscript which is joint work with Ivelin Georgiev, Amy Anderson, and Bruce Donald:


2.1 Introduction

Gramicidin S synthetase (Grs) has been studied as a model NRPS system for understanding fundamental mechanism and as a template for enzyme redesign. Grs is composed of two multi-domain subunits, GrsA and GrsB. GrsA, in concert with GrsB, makes the cyclic decapeptide antibiotic gramicidin S. GrsA subunit (or PheATE) comprises three domains, a N-terminal ~62 kDa adenylation (PheA) domain followed by a ~9 kDa thiolation (PheT) domain (or peptidyl-carrier protein, PCP domain) and a ~56 kDa C-terminal epimerization (PheE) domain (Figure 2.1). GrsA initiates the formation of peptide chain by activating and epimerizing the L-form substrate phenylalanine into a D-Phe-S-phosphopantetheinyll intermediate which is covalently linked to the PheT domain. The condensation domain on the first module of GrsB recognizes the intermediate and catalyzes the peptide bond formation. Peptide is elongated along with four other modules that activate L-form amino acids, Pro, Val, Orn and Leu to form a pentapeptide. At the end, the terminal thioesterase domain catalyzes cyclization of two identical pentapeptides and releases the antibiotic gramicidin S.
Previous kinetic and thermodynamic studies have revealed the basis of substrate selectivity for each of the PheATE domains (90). Substrate binding studies using fluorescence titration with PheA or PheATE suggested that the PheA domain binds indiscriminately to several L-form non-cognate amino acids including Tyr, Leu, Val, Ala, His, Arg, Lys, Asp, Glu and Pro as well as the D-form Phe. However from the steady-state kinetics analysis of the PheATE with these substrates, only L-form and D-form Phe are preferred substrates for the adenylation reaction. Trp, Tyr, and Leu can be activated by the PheA domain but the catalysis is 13-, 310-, and 16-fold less efficient, respectively.
By using stopped-flow fluorometry to study pre-steady state substrate binding of PheA, Stevens et al. found that in presence of ATP and Mg$^{2+}$, wild-type PheA selectively binds to L-form Phe in the first 500 ms (8). Non-cognate substrates eventually bind when equilibrium was reached. In this study, a two-step binding process was suggested. At the first step, binding of ATP induces a conformation change in the binding pocket that favors binding of L/D-form Phe. The second step is for the structure to adopt the catalytically competent conformation that involves hydrolysis of ATP and formation of the phenylalaninyl adenylate.

The crystal structure of GrsA-PheA in complex with the wild-type (WT) substrate Phe and the AMP cofactor, has been determined, thus making this domain a suitable target for structure-based redesign. In this chapter we present a computational structure-based redesign of the GrsA-PheA domain for a set of non-cognate substrates including L-form amino acids Leu, Arg, Glu, Lys, and Asp. Computational methodology and experiments for the redesign were conducted by Ivelin Georgiev. The results from redesigning NRPS enzymes can be divided into three categories: (1) Switch the enzyme specificity from the WT substrate to the target substrate, so that the redesigned enzyme prefers the target over the WT substrate; (2) Improve (but not switch) the enzyme specificity for the target substrate (in the case where the WT enzyme already has activity for the target substrate); and (3) Create activity for the target substrate (in the case where the WT enzyme has no activity for the target substrate). In previous work, we reported
structure-based redesigns of the active site of GrsA-PheA that were experimentally confirmed to improve (but not switch) substrate specificity for Tyr (8) (category (2) results). Those redesigns were based on older versions of our $K^*$ algorithm (91), which computes partition functions over molecular ensembles defined by continuously-flexible rotamers and/or backbones. In this chapter, we present the application of improved versions of $K^*$ that incorporate several recently-described algorithmic enhancements (18, 19, 92) to redesign the active site of GrsA-PheA in order to improve its specificity for a set of non-cognate substrates for which the WT enzyme has little or virtually no specificity. Detailed kinetic experiments for a set of the top-ranked computational predictions confirm the desired improvement in specificity for five non-cognate substrates (Leu, Arg, Glu, Lys, and Asp). Several of the Leu redesigns show a switch of specificity from Phe (category (1) results). While the WT enzyme has virtually no activity for Arg, Glu, Lys, and Asp, the redesigns for these substrates successfully create the desired activity (category (3) results). Further algorithmic enhancements for predicting mutations outside of (both close to and far away from) the enzyme active site aiming at additional improvement in the substrate specificity are described and validated experimentally. The mutant with the highest activity for a non-cognate substrate exhibits 1/6 of the WT enzyme:WT substrate activity, further confirming the feasibility of our computational approach. The majority of our validated computational predictions have not been previously tested. Our results also suggest that structure-based protein design
can identify active mutants different from those selected by evolution and from the predictions of other computational approaches.

2.2 Experimental materials and methods

2.2.1 Materials

Amino acid substrates, compounds, and enzymes for the pyrophosphate release assay were purchased from Sigma-Aldrich (St. Louis, MO). Vector pQE60 and Escherichia coli strain M15 were purchased from Qiagen (Valencia, CA). Plasmid pQE60 containing WT and A301G mutant PheA genes from B. brevis (GI: 39366) were obtained as previously described (8).

2.2.2 Mutagenesis of PheA

Mutagenesis was performed using the QuickChange site-directed mutagenesis system (Stratagene) in accordance with the manufacturer’s instructions by using primers summarized in the Table 2.1. Preparation of the plasmid DNA was done in E. coli DH5α following standard procedures. All constructs were confirmed by DNA-sequencing at Duke University DNA Analysis Facility.
Table 2.1: Mutagenesis primers for redesign of PheA

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T278L-f</td>
<td>5’- GTTATTTTGTACCCTACCTATGTAG -3’</td>
</tr>
<tr>
<td>T278L-r</td>
<td>5’- GGTAACAAATAACAGTGATTTCCCTTTTG G -3’</td>
</tr>
<tr>
<td>T278M-f</td>
<td>5’- GTTATTATGTACCACCTACCTATGTAG -3’</td>
</tr>
<tr>
<td>T278M-r</td>
<td>5’- GGTAACATAATAACAGTGATTTCCCTTTTG G -3’</td>
</tr>
<tr>
<td>T278H-f</td>
<td>5’- CTGTTATCCACTACCTACCTATGTAG -3’</td>
</tr>
<tr>
<td>T278H-r</td>
<td>5’- GTGGTAAGTGAATAACAGTGATTTCCCTTTTG G -3’</td>
</tr>
<tr>
<td>T278D-f</td>
<td>5’- CTGTTATGACCTACCTACCTATGTAG -3’</td>
</tr>
<tr>
<td>T278D-r</td>
<td>5’- GTGGTAAGTGAATAACAGTGATTTCCCTTTTG G -3’</td>
</tr>
<tr>
<td>T278K-f</td>
<td>5’- CTGTTATTAAGTTACACCTACCTATGTAG -3’</td>
</tr>
<tr>
<td>T278K-r</td>
<td>5’- GTGGTAACATAATAACAGTGATTTCCCTTTTG G -3’</td>
</tr>
<tr>
<td>A322V-f</td>
<td>5’- CATAATGTCTATGGCCCTACGGAAAC -3’</td>
</tr>
<tr>
<td>A322V-r</td>
<td>5’- GCCATAGACATTTATGTAAGTTAC -3’</td>
</tr>
<tr>
<td>S447N-f</td>
<td>5’- GAAGTTGAGAATATTCTTCTAAAGCATATG -3’</td>
</tr>
<tr>
<td>S447N-r</td>
<td>5’- GAATATTCTCAACTTCTTCTAGTTCAACTCG -3’</td>
</tr>
<tr>
<td>I277L-f</td>
<td>5’- GTTCTTTTGTACCACCTACCTATGTAG -3’</td>
</tr>
<tr>
<td>I277L-r</td>
<td>5’- GGTAACAAAAAGACAGTGATTTCCCTTTTG G -3’</td>
</tr>
<tr>
<td>V187L-f</td>
<td>5’- GCTTATCTTATTTATACCTTCTGGTACAACAGGC -3’</td>
</tr>
<tr>
<td>V187L-r</td>
<td>5’- GTATAAATAAGATAAGCAAGATCGGTTGATTACTTGG -3’</td>
</tr>
</tbody>
</table>

2.2.3 Expression and purification of PheA

Vector pQE60 containing constructs of WT or mutant PheA with a C-terminal his-tag was transformed into *E. coli* M15 [pREP4] cells for expression. The proteins were expressed by induction of mid-log cells (OD~0.8) with 0.2 mM IPTG and an addition of
10 mM MgCl₂ overnight at 18 °C. The double mutant T278L/A301G as well as the triple mutants T278L/A301G/S447N, I277L/T278L/A301G, and V187L/T278L/A301G were induced with 0.05 mM IPTG and expressed at 18 °C overnight to increase protein solubility. Double mutants T278D/A301G, T278H/A301G, T278K/A301G and the quadruple mutant I277L/T278L/A301G/S447N were expressed at 16 °C with 0.05 mM IPTG. In a typical preparation of 2 L culture, 7g of cell pellets were re-suspended in 35 mL of buffer A (100 mM Tris-HCl, pH 7.5, 250 mM NaCl, and 5 mM TCEP) supplemented with a protease inhibitor cocktail. The cells were lysed by French-press and cell debris was removed by centrifugation at 20000xg for 30 min. The resulting supernatant was incubated with Ni-NTA agarose (10mL) in buffer B (50 mM Tris-HCl, pH 8.0, 400 mM NaCl, 20 mM imidazole, 0.5 mM TCEP) at 4 °C for one hour. The agarose was then washed extensively with buffer B. His-tagged proteins were eluted with 400 mM imidazole (pH 6.8) and further purified by a Superdex 200 gel filtration chromatography (GE Healthcare, Chalfont St. Giles, United Kingdom) in buffer C (50 mM Tris-HCl, pH 7.5, 1 mM TCEP). The purified proteins (>95% pure by SDS-PAGE) were concentrated to >20 mg/mL using an Amicon Ultra-15 concentrator with addition of glycerol (10% final) and rapidly frozen by liquid nitrogen for storage at -80 °C.

2.2.4 PPi Release Assay

The rate of pyrophosphate (PPi) release was measured using a coupled, continuous, spectrophotometric assay (93). In the Reaction of 100 μL total volume, PheA
or mutants (0.1 - 1 µM) were incubated at 25 °C with varying concentrations of amino acids (1 µM - 64 mM) in buffer containing 100 mM Tris-HCl (pH 7.5), 1 mM uridine-diphosphate-glucose, 375 µM glucose-1,6-bisphosphate, 1 mM β-nicotinamide adenine dinucleotide, 10 mM MgCl₂, 2 mM adenosine 5'-triphosphate, 5 mM dithiothreitol, 2 U/mL uridine-5'-diphosphoglucose pyrophosphorylase, 4 U/mL phosphoglucomutase, 4 U/mL glucose-6-phosphate dehydrogenase. Mutants T278D/A301G, T278H/A301G, and T278K/A301G were assayed in 50 mM Tris-HCl (pH 7.5) buffer. Reactions were initiated by addition of the enzymes following a 10-min incubation to allow the removal of any contaminating PPi. The absorbance at 340 nm (NADHε340= 6,317 M⁻¹ cm⁻¹) was monitored using an Agilent 8453 spectrophotometer. Substrate concentrations covering 0.2 to 5 K_M were used to determine the complete steady-state curve. The initial velocity of each substrate concentration was determined by comparison with mock-treated enzyme and fitted with the Michaelis-Menten equation to obtain k_cat and K_M.

2.2.5 Stopped-Flow Tryptophan Fluorescence Quenching

Tryptophan quenching was monitored for 0.5 s under stopped-flow conditions using an Applied Photophysics SX.18 MV stopped-flow Analyser equipped with a photomultiplier tube (PMT) to monitor fluorescence. The sample was excited at 285 nm, and emission was monitored using a 320 nm band-pass filter. A solution of 2 µM PheA or mutants in 50 mM Tris-HCl (pH 7.5), 5 mM DTT was combined 1:1 with a solution of phenylalanine or leucine (20-500 µM), resulting in 1 µM enzyme and 10-250 µM.
substrate in the sample cell. When present, ATP and MgCl₂ were used at a final concentration of 1 mM and 10 mM respectively. Between 10 and 15 transients were collected for each substrate concentration and averaged. Fluorescence data from the WT PheA with Phe was fit to the double exponential function

$$Fluorescence = Ae^{-k_{obs1}t} + Be^{-k_{obs2}t} + C$$  

(2.1)

The observed rate $k_{obs1}$ and substrate concentrations were fit to a linear function

$$k_{obs1} = k_1[S] + k_{-1}$$  

(2.2)

The observed rate $k_{obs2}$ was fit to a hyperbolic function

$$k_{obs2} = \frac{K_D[S]}{k_2 + [S]}$$  

(2.3)

From the fitting functions, $k_1$ and $k_{-1}$ are the microscopic rate constants for the association and dissociation of substrate, and $k_2$ is the forward rate of the second binding mode.

Fluorescence data from the WT and mutant PheA with Leu was fit to the single exponential function

$$Fluorescence = Ae^{-k_{obs1}t} + C$$  

(2.4)

The observed rate $k_{obs1}$ and substrate concentrations were fit to linear function to obtain $k_1$ and $k_{-1}$.

### 2.2.6 Free Energy Calculation

According to transition state theory, the equilibrium constant between enzyme (E) + substrate (S) and the transition state ES‡ is proportional to the activation energy
ΔG\textsuperscript{‡} from k_{cat}/K_{M}. The activation energy is composed of two terms, an energetically unfavorable term ΔG\textsuperscript{‡}, due to the activation energy of the chemical steps of bond making and breaking, and a compensating energetically favorable term ΔG\textsubscript{S}, due to the realization of the binding energy. That is,

\[ \Delta G_T^{\ddagger} = \Delta G^\ddagger + \Delta G_S \] (2.5)

The equation can be rewritten in terms of transition state theory to express k_{cat}/K_{M},

\[ RT \ln \frac{k_{\text{cat}}}{K_M} = RT \ln \frac{kT}{h} - \Delta G^\ddagger - \Delta G_S \] (2.6)

where \( k \) is the Boltzmann constant, \( h \) is the Planck constant, R is the gas constant and T is the absolute temperature. Therefore the activation energy \( \Delta G_T^{\ddagger} \) can be calculated by rearranging the equation to

\[ \Delta G_T^{\ddagger} = \Delta G^\ddagger + \Delta G_S = RT \ln \frac{kT}{h} - RT \ln \frac{k_{\text{cat}}}{K_M} \] (2.7)

With the assumption that the active site residues and the side-chain of the amino acid substrate are not directly involved in the chemistry of the reaction, the difference of the activation energy is considered to be the difference in binding energy of the enzyme and transition state. Let ΔΔG_{Phe-Leu} be the difference in binding energy between substrate Phe and Leu for WT and mutant PheA; let ΔΔG_{WT-Mut} be the difference in binding energy between WT and mutants for Phe and Leu; and let ΔΔG_{int} be the coupling energy (interaction energy) between T278L and A301G for Phe and Leu. Calculation of the free energy difference was done using following equations:
\[ \Delta \Delta G_{\text{Phe-Leu}} = -RT \ln \left( \frac{k_{\text{cat}}}{K_M} \right)_{\text{Phe}} \]

\[ \Delta \Delta G_{\text{WT-Mut}} = -RT \ln \left( \frac{k_{\text{cat}}}{K_M} \right)_{\text{WT}} \]

\[ \Delta \Delta G_{\text{int}} = (\Delta \Delta G_{\text{WT-A30G}} - \Delta \Delta G_{\text{WT-T278L}}) - \Delta \Delta G_{\text{WT-A30G/T278L}} \] (2.10)

Microscopic rate constant of association, \( k_1 \) was used to calculate the free energies of transition states using the function

\[ \Delta G^z_T = -RT \left[ \ln(k_1) - \ln \left( \frac{k_0}{Th} \right) \right] \] (2.11)

Where \( k_0 \) is Boltzmann’s constant, \( h \) is Planck’s constant, and \( T \) is the temperature in Kelvin. The free energy of equilibrium is calculated as

\[ \Delta G = -RT \ln K_{eq} = -RT \ln \left( \frac{k_1}{k_{-1}} \right) \] (2.12)

**2.3 Computational Method**

The computational experiments in this section was mainly designed and performed by Ivelin Georgiev.

**2.3.1 Active Site Mutation Prediction**

For a given protein-substrate complex, the \( K^* \) algorithm computes partition functions over conformational ensembles, where the contribution of each conformation...
to the partition function is weighted using Boltzmann probabilities. The ratio of the partition functions for the bound complex and unbound protein and ligand is then used to compute a provably-accurate $\varepsilon$-approximation to the binding constant for the given protein-substrate complex. $K^*$ scores were computed for each candidate protein sequence with the target substrate; sequences with higher $K^*$ scores are predicted to have better specificity for the target substrate. For computational efficiency, $K^*$ uses the minimized Dead-End Elimination (MinDEE) (18) and the Backbone DEE (BD) (19) algorithms as an initial pruning filter, and the $A^*$ branch-and-bound search (94) for the subsequent conformation enumeration (18). MinDEE and BD are DEE-based algorithms that, unlike previous DEE algorithms (95), guarantee the identification of the Global Minimum Energy Conformation (GMEC) for, respectively, a model with continuously-flexible rotamers and a flexible backbone. Combined with $A^*$, MinDEE and BD also output conformations and sequences in the precise order in which they are ranked by the model, so that no low-energy solutions are missed by the algorithm.

$K^*$ runs were performed for the following substrates: Arg, Glu, Leu, Lys, and Asp. The crystal structure of GrsA-PheA (PDB id: 1amu (4)) was used in the computational redesigns. The seven active site residues 236, 239, 278, 299, 301, 322, and 330 were modeled using continuously-flexible rotamers and were allowed to mutate. In addition, the AMP cofactor and a steric shell consisting of all residues within 8 Å from the ligand or within 3 Å from any of the seven active site residues were included as part
of the input structure. The ligand substrate was also modeled using continuously-flexible rotamers and was allowed to rotate/translate. Rotamers were obtained from the Penultimate rotamer library modal values (15). The energy function consisted of the Amber electrostatic, vdW, and dihedral terms and the EEF1 pairwise implicit solvation energy term (96). A distance-dependent dielectric of 6 and a solvation-energy scaling factor of 0.8 were used. Conformations with an initial steric overlap of more than 1.5 Å were pruned.

The top ten $K^*$ 2-point mutation predictions for each of the target substrates are shown in Table 2.2. As an orthogonal check for the computational predictions with Leu as substrate, additional $K^*$ runs that allowed flexible backbones were performed. These runs applied the BD algorithm in the DEE pruning stage and continuous backbone minimization in the A* enumeration stage, but were otherwise identical to the continuously-flexible rotamers $K^*$ runs using MinDEE. The goal of these additional runs was to determine whether the top Leu mutants predicted by the MinDEE $K^*$ runs and selected for experimental validation, would also be among the top predictions when the backbone was allowed to flex. Indeed, 278L/301G was again ranked 1st, 278M/301G – 10th, and 322V/301G – 4th. The fact that these three mutants were still within the top ten predictions even when backbone flexibility was allowed, further increased the confidence in the feasibility of our predictions.
Table 2.2: Top ten 2-point active site mutants predicted by $K^*$ for each of the five target substrates (Leu, Arg, Glu, Lys, and Asp)

<table>
<thead>
<tr>
<th></th>
<th>Leu</th>
<th>Arg</th>
<th>Glu</th>
<th>Lys</th>
<th>Asp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>278L/301G</td>
<td>278D/301G</td>
<td>239R/322R</td>
<td>278D/299D</td>
<td>239M/278R</td>
</tr>
<tr>
<td>2</td>
<td>278I/301G</td>
<td>299D/301G</td>
<td>278H/301G</td>
<td>278D/299E</td>
<td>278H/301G</td>
</tr>
<tr>
<td>3</td>
<td>299W/301G</td>
<td>278E/301G</td>
<td>239R/278H</td>
<td>278E/299D</td>
<td>278K/301G</td>
</tr>
<tr>
<td>4</td>
<td>299F/301G</td>
<td>278A/301G</td>
<td>322Q/301G</td>
<td>278D/301G</td>
<td>239K/278R</td>
</tr>
<tr>
<td>5</td>
<td>236M/301G</td>
<td>299E/301G</td>
<td>278K/301G</td>
<td>278E/299E</td>
<td>278K/330M</td>
</tr>
<tr>
<td>6</td>
<td>236L/301G</td>
<td>236G/278A</td>
<td>239K/278H</td>
<td>278D/299M</td>
<td>239R/278K</td>
</tr>
<tr>
<td>7</td>
<td>330F/301G</td>
<td>322G/301G</td>
<td>278N/322K</td>
<td>278D/299Q</td>
<td>239K/322K</td>
</tr>
<tr>
<td>8</td>
<td>278M/301G</td>
<td>278S/301G</td>
<td>278N/301G</td>
<td>278D/299W</td>
<td>239K/322R</td>
</tr>
<tr>
<td>9</td>
<td>322V/301G</td>
<td>299L/301G</td>
<td>278H/299K</td>
<td>278D/299F</td>
<td>236M/278K</td>
</tr>
<tr>
<td>10</td>
<td>278F/301G</td>
<td>236G/301G</td>
<td>278H/299R</td>
<td>278D/322G</td>
<td>278K/299M</td>
</tr>
</tbody>
</table>

2.3.2 Bolstering mutation prediction

The $K^*$ algorithm allows us to identify mutations within the active site of an enzyme. The kinetics experiments showed these $K^*$-predicted mutations yielded highly active mutants for Leu. We then investigated whether additional improvement in the Leu specificity could be achieved by introducing additional mutations outside of the active site. Previously, in other design protocols, this was done by performing multiple rounds of directed evolution on the active-site mutants (13). As an alternative, we applied a purely computational approach for predicting mutations outside of the enzyme active site. As a starting point for these computational experiments, we selected the highest-activity $K^*$ mutant for Leu (T278L/A301G). We then applied a SCMF entropy-based method (97) combined with our MinDEE/A* (18) algorithm to predict mutations both close to and far away from the enzyme active site to obtain further
improvement in the target substrate specificity. The SCMF entropy-based method heuristically selects residue positions, anywhere in the protein, that may be tolerant to mutation. Eight residue positions with the highest computed residue entropy (45, 187, 207, 210, 238, 239, 277, and 447) were selected for the MinDEE/A* mutation search. The lowest-energy rotamer conformation for the highest-activity \( K^* \) mutant for Leu (T278L/A301G) was used as input to the MinDEE/A* redesign algorithm. The MinDEE/A* input structure included the residue positions in 1amu within 8 Å of the ligand or the eight high-entropy positions identified in the entropy step. Residues 45, 187, 207, 210, 238, 277, and 447 were modeled using continuously flexible rotamers and allowed to mutate; residue 239 was modeled using continuously-flexible rotamers but was not allowed to mutate, since it is part of the enzyme active site. Mutations to these residue positions are then predicted by using the MinDEE/A* algorithm. We refer to these mutations as “bolstering.” Table 2.3 shows the top bolstering mutations from the 3-point MinDEE/A* mutation search. The addition of the bolstering mutations aims at further stabilizing the mutant enzyme and may counteract a possible destabilizing effect from the introduction of the active-site mutations. The active-site mutations plus bolstering mutations were then tested by creating mutant proteins containing both sets of mutations, and measuring the kinetic parameters.
Table 2.3: Top bolstering mutations from the 3-point MinDEE/A\* mutation search. These mutations are in addition to the 2-point active site mutation 278L/301G.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>187L/238I/447L</td>
</tr>
<tr>
<td>2</td>
<td>187L/238I/447N</td>
</tr>
<tr>
<td>3</td>
<td>187L/238I/447I</td>
</tr>
<tr>
<td>4</td>
<td>187L/277L/447L</td>
</tr>
<tr>
<td>5</td>
<td>187L/238I/447A</td>
</tr>
<tr>
<td>6</td>
<td>187L/447L</td>
</tr>
<tr>
<td>7</td>
<td>187L/207L/447L</td>
</tr>
<tr>
<td>8</td>
<td>187L/277L/447N</td>
</tr>
<tr>
<td>9</td>
<td>187L/277L/447I</td>
</tr>
<tr>
<td>10</td>
<td>187L/447N</td>
</tr>
<tr>
<td>11</td>
<td>187L/238I/447L</td>
</tr>
</tbody>
</table>

2.4 Results

The $K^*$ algorithm (18, 91) was applied to predict mutations to the active site of GrsA-PheA to switch the enzyme specificity from the WT substrate Phe toward the target L-form non-cognate substrates Leu, Arg, Glu, Lys, and Asp. For each of the redesign targets, sets of the top computational predictions were then visualized and selected for experimental validation. For the Leu redesigns, additional mutations outside of the active site were further selected using a novel computational protocol combining a Self-Consistent Mean Field (SCMF) entropy-based method (97) with our MinDEE/A\* (18) algorithm. As with the active site mutations, sets of the computationally-predicted mutations outside of the active site were visualized and selected for experimental validation.
2.4.1 Steady-state Kinetic Analysis

To confirm the desired improvement in specificity for the computationally-predicted mutants, we performed detailed steady-state kinetic experiments on a set of top-ranked computational predictions for each of the target substrates. WT and mutant PheA were over-expressed and purified to homogeneity as shown in the SDS-PAGE (Figure 2.2). The stability of PheA mutants T278L/A301G, I277L/T278L/A301G and I277L/T278L/A301G/S447N were further confirmed by comparing changes of circular dichroism in the guanidine titration experiment to the WT protein (Figure 2.3). The adenylation activity of the WT and mutant PheA was measured by monitoring the PPi release rate using a continuous spectrophotometric assay (93). The assay measures the degree of ATP consumption in an amino acid concentration-dependent manner, which reflects the rate of the enzyme to form and turn over aminoacyl adenylate. Among the proteins tested, all of them, except for the T278K/A301G mutant, showed typical hyperbolic curves with the initial velocity approaching saturation as the concentration of amino acid increases (Figure 2.4-2.7). A mock control experiment in the absence of the amino acid substrate showed a slow background ATP hydrolysis whose rate was subtracted from the rate in the presence of the substrate. The values of the kinetic constants $k_{cat}$, $K_M$ and $k_{cat}/K_M$ for different proteins with different substrates are given in Table 2.4.
Figure 2.2: SDS-PAGE showing the homogeneity of the WT and mutant PheA. 1-2 mg of protein was loaded for each sample lane. M: Marker, 1: WT, 2: A301G, 3: T278L, 4: A301G/A322V, 5: T278M/A301G, 6: T278L/A301G, 7: T278L/A301G/S447N, 8: T278L/I277L/A301G, 9: V187L/T278L/A301G, 10: I277L/T278L/A301G/S447N, 11: T278D/A301G, 12: T278H/A301G, 13: T278K/A301G.

Figure 2.3: Guanidine-HCl titration of wild-type and mutant PheA proteins.
Figure 2.4: Representative steady-state kinetics curves for: WT PheA with Phe (A) and Leu (B), A301G/A322V with Phe (C) and Leu (D), and T278M/A301G with Phe (E) and Leu (F).
Figure 2.5: Representative steady-state kinetics curves for: T278L/A301G with Phe (A) and Leu (B), T278L/A301G/S447N with Phe (C) and Leu (D), and V187L/T278L/A301G with Phe (E) and Leu (F).
Figure 2.6: Representative steady-state kinetics curves (black) for I277L/T278L/A301G with Phe (A) and Leu (B), I277L/T278L/A301G/S447N with Phe (C) and Leu (D), and T278H/A301G with Phe (E) and Glu (F). For comparison, the steady-state kinetics curve for the WT enzyme with Glu (F, blue) is also shown.
Figure 2.7: Representative steady-state kinetics curves (black) for T278K/A301G with Phe (A) and Asp (B), and T278D/A301G with Phe (C), Lys (D), and Arg (E). For comparison, the steady-state kinetics curve for the WT enzyme with Asp (B, blue), Lys (D, blue), and Arg (E, blue) are also shown.
Table 2.4: Mutant enzymes with experimentally-observed specificity ($k_{cat}/K_M$), $k_{cat}$, and $K_M$ for a target substrate (the target substrate is the same as the respective redesign target) and the WT substrate (Phe) are shown along with their respective ranks as determined by the algorithm: (i) $K^*$ active-site mutants with their respective ranks from the 2-point mutation searches for the different substrates; (ii) Bolstering mutations added to the T278L/A301G mutant with Leu as substrate. 

For clarity, the WT enzyme:WT substrate rates are shown only once. ND, not detectable. $K_M$ and $k_{cat}/K_M$ cannot be determined accurately because the solubility of Asp (~50 mM in water) limits the reaction velocity under the experimental condition, in which the velocity remains linearly dependent on the concentration of the substrate.

<table>
<thead>
<tr>
<th>Redesign</th>
<th>Enzyme</th>
<th>Target Substrate</th>
<th>WT Substrate (Phe)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$k_{cat}$</td>
<td>$K_M$</td>
</tr>
<tr>
<td>(i)</td>
<td>Leu</td>
<td>T278L/A301G</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T278I/A301G</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>A327V/A301G</td>
<td>9</td>
<td>4.18 ± 0.52</td>
</tr>
<tr>
<td></td>
<td>WT†</td>
<td>28.74 ± 1.58</td>
<td>6.98 ± 1.60</td>
</tr>
<tr>
<td>Arg</td>
<td>T278D/A301G</td>
<td>1</td>
<td>0.238 ± 0.007</td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Glu</td>
<td>T278I/A301G</td>
<td>2</td>
<td>0.2773 ± 0.035</td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Lys</td>
<td>T278D/A301G</td>
<td>4</td>
<td>1.09 ± 0.08</td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Asp</td>
<td>T278K/A301G</td>
<td>3</td>
<td>&gt; 0.25</td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>(ii)</td>
<td>Leu</td>
<td>T278L/A301G/S447N</td>
<td>0.85 ± 0.11</td>
</tr>
<tr>
<td></td>
<td>D77L/T278L/A301G</td>
<td>1.52 ± 0.09</td>
<td>0.013 ± 0.002</td>
</tr>
<tr>
<td></td>
<td>V187L/T278L/A301G</td>
<td>1.69 ± 0.22</td>
<td>0.011 ± 0.001</td>
</tr>
<tr>
<td></td>
<td>L275L/T278L/A301G/S447N</td>
<td>0.37 ± 0.04</td>
<td>0.0054 ± 0.0007</td>
</tr>
</tbody>
</table>

2.4.2 Redesign for Leu

The WT PheA shows a rather strong specificity to its natural substrate Phe with $k_{cat}/K_M$ value ~229-fold higher than the non-cognate amino acid Leu. A previous binding study showed that without binding of ATP, the WT PheA can accommodate most of the non-cognate amino acid substrates (90). Our results, however, show that the WT protein...
can only activate certain types of amino acids including Phe, Leu, and Val, but not charged amino acids. To switch substrate specificity of PheA from Phe to Leu, we applied the $K^*$ protein redesign algorithm (18) using as input the crystal structure of WT PheA in complex with the Phe substrate and AMP. The top-ranked $K^*$ mutation sequence was T278L/A301G (Table 2.4). The lowest-energy T278L/A301G structure from the $K^*$ ensemble with Leu as substrate is shown in Figure 2.6. The double mutant protein showed a $\sim$19-fold increase of $k_{cat}/K_M$ with Leu and a $\sim$27-fold decrease of $k_{cat}/K_M$ with Phe from the WT PheA, which results in $\sim$2.3-fold higher $k_{cat}/K_M$ for Leu than for Phe (Figure 2.8). As a result, the double mutant protein makes a $\sim$521-fold switch in specificity given the $k_{cat}/K_M$ ratio of Leu over Phe is only $\sim$0.0043 for WT PheA. The difference in the $k_{cat}/K_M$ value between the WT and the T278L/A301G PheA with Leu and Phe is mainly driven by the $K_M$ values, which have a $\sim$465-fold decrease with Leu and a $\sim$54-fold increase with Phe in the T278L/A301G mutant. As a result, the $K_M$ value with Leu becomes $\sim$6-fold lower than with Phe in T278L/A301G. The switch suggests that the double mutant protein now binds tighter to Leu than to Phe. WT PheA has a rather high $k_{cat}$ value with Leu while it is relatively low with Phe. The $k_{cat}$ value of the T278L/A301G mutant with either Leu or Phe remains at the same level as the WT PheA with Phe. The measurement of $k_{cat}$ is limited by the rate of product release because of the tight binding of the aminoacyl-AMP. Therefore, the high $k_{cat}$ value with Leu for the WT protein might be due to the loose binding of the leucyl-AMP product given its high $K_M$ value. The
double mutant T278M/A301G is ranked 8th by $K^*$ for the Leu redesign. This double mutant was previously predicted by a sequence alignment-based method and verified experimentally to activate Leu (5). We have confirmed that the T278M/A301G mutant has a $k_{cat}/K_M$ value ~5-fold higher with Leu and ~73-fold lower with Phe than the WT PheA. The T278M/A301G mutant selects ~1.5-fold more Leu than Phe (Figure 2.8). $K^*$ also predicted the double mutant A301G/A322V, which had a $k_{cat}/K_M$ value ~59-fold lower with Phe and ~2.2-fold higher with Leu than WT PheA.
Figure 2.8: $K^*$-predicted structure of the lowest-energy T278L/A301G conformation with Leu as substrate. Shown are the Leu substrate (CPK ball-and-stick and gray space-filling representations), the AMP cofactor (green), the two active site mutations 278L and 301G (orange sticks and CPK dots), and the other eight active site residues, including the remaining five mutable residues (CPK sticks and dots); C331 is hidden behind D235.
Figure 2.9: Specificity ratio \((k_{\text{cat}}/K_{M})_{\text{Leu}}/(k_{\text{cat}}/K_{M})_{\text{Phe}}\) for WT and mutant PheA in the Leu redesigns. The WT PheA shows a ratio of 0.0043 with its \(k_{\text{cat}}/K_{M}\) values of 4.15 (mM\(^{-1}\) min\(^{-1}\)) for Leu and 951.4 (mM\(^{-1}\) min\(^{-1}\)) for Phe. A301G/A322V still prefers Phe with a ratio of 0.57. T278M/A301G prefers Leu over Phe with a ratio of 1.5, while T278L/A301G shows a ratio of 2.3. The three triple mutants have ratios of 7.8 for T278L/A301G/S447N, 7.0 for I277L/T278L/A301G and 9.4 for V187L/T278L/A301G. The quadruple mutant has a ratio of 9.3.

To further improve the specificity of the double mutant T278L/A301G for Leu, we identified distal bolstering mutations outside the active site by applying the computational protocol described in the computational redesign section 2.3.2. Up to 3-point bolstering mutation search (in addition to the T278L/A301G active site mutant) was performed for the mutable positions, and the top mutations V187L, I277L and
S447N, were selected and tested. Position 238 was ignored since its side chain rotamer is too close to the active site. All the three triple mutants gave 1 to 2-fold additional improvement of the specificity with Leu over the T278L/A301G mutant. Among them, the T278L/A301G/S447N showed an additional ~2-fold higher $k_{cat}/K_M$ value for Leu with a ~2.7-fold decrease of $K_M$ and a slightly lower $k_{cat}$. The $K_M$ values with Leu are slightly lower for both I277L/T278L/A301G and V187L/T278L/A301G with Leu compared to the T278L/A301G mutant, while their $k_{cat}$ values are both slightly higher. All three triple mutants have a decreased specificity toward Phe from the T278L/A301G mutant. As a result, the difference of $k_{cat}/K_M$ between Leu and Phe became ~7.8-fold in T278L/A301G/S447N, ~7-fold in I277L/T278L/A301G and ~9.4-fold in V187L/T278L/A301G toward a better selection of Leu (Figure. 2.8). These mutants gave a switch of ~1796-fold in T278L/A301G/S447N, ~1614-fold in I277L/T278L/A301G and ~2168-fold in V187L/T278L/A301G from the WT PheA, exhibiting up to 1/6 of the WT enzyme:WT substrate activity (absolute values of $k_{cat}/K_M$). We next tested whether the quadruple mutant combining S447N and I277L could give additional improvement. However, while its $K_M$ with Leu is as low as the T278L/A301G/S447N triple mutant, its $k_{cat}$ is ~2-fold lower than any of the triple mutants and the T278L/A301G mutant. Nevertheless, it showed a significant result with its $K_M$ value of Phe close to I277L/T278L/A301G and its $K_M$ of Leu close to T278L/A301G/S447N.
2.4.3 Redesign for Charged Amino Acids

The active site of PheA contains mainly a hydrophobic pocket and shows no observable activity with charged amino acids from our activity assay (Figure 2.6, panel F and Figure 2.7, panels B, D, F). We next tested our redesign algorithm for the activation of charged amino acids, Arg, Lys, Glu, and Asp by predicting mutations around the active sites of WT PheA. Table 2.2 shows top ten 2-point active site mutants predicted by \( K^* \) for each of the four charged amino acid substrates (Arg, Glu, Lys, and Asp). Among these prediction mutants, A301G was again predicted in most of the top ten predictions for binding of the charged amino acids. The change from alanine to glycine at position 301 might provide the backbone more flexibility, therefore creates a more promiscuous active site that can accommodate other substrate side chains. We therefore focused the 2-point active site mutants that have the A301G mutation. For the second mutable positions, as expected, the algorithm predicted mainly negatively-charged side chains to bind Arg and Lys, and positively-charged side chains to bind Glu and Asp in the active site of PheA. The prediction resulted in the double mutant T278D/A301G, which ranked first to bind Arg and fourth to bind Lys. This double mutant showed small but significant activity with both Arg and Lys under the same conditions as the Leu redesign. The activity was improved when the Tris-HCl concentration was lowered to 50 mM. The T278D/A301G mutant showed substrate concentration-dependent kinetics with both the Arg and the Lys substrate (Figure 2.6, panels C-E). Both substrates
showed much higher $K_M$ values, which suggest a weak binding between the mutant protein and the substrates. Their $k_{cat}$ values with this double mutant are also several fold lower (~7-fold for Arg and ~1.5-fold for Lys) than the WT PheA with Phe. Among the top-scored sequences to bind Glu and Asp, T278H/A301G was ranked 2nd to bind Glu and T278K/A301G was ranked 3rd to bind Asp. Both mutants showed substrate concentration-dependent kinetics with their substrates (Figure 2.5, panels E-F and Figure 2.6, panels A-B). The rate of the T278K/A301G increased linearly without approaching saturation as the concentration of Asp approached its maximum solubility. As a result, only a lower bound on $k_{cat}$ was determined. The three double mutants T278D/A301G, T278K/A301G and T278H/A301G, while acquiring new substrate activity, showed decreased specificity for Phe. Unlike the increase of $k_{cat}$ for Phe observed in all of the Leu-redesigned mutants, the $k_{cat}$ values are significantly lower in T278D/A301G, T278K/A301G and T278H/A301G. The lower $k_{cat}$ value with the natural substrate Phe suggests that introduction of charged side-chains in the active site might have an influence on the enzyme catalysis.

2.4.4 Other Experimentally-tested Mutants

Table 2.5 presents the experimental results for several mutants not shown in Table 2.4. These mutants are divided into three categories: (1) Computationally-predicted mutants that do not exhibit the desired specificity for a target substrate (Table 2.5i); (2) Single-point mutations that, by themselves, are not predicted by our algorithms,
but that are components of other computationally-predicted mutants (Table 2.5ii); (3) Mutants predicted by our algorithms and selected for experimental verification that were difficult to purify due to solubility issues (Table 2.5iii). The double mutants T278E/A301G and T278D/I299E designed to bind Arg and Lys, respectively, both went into inclusion bodies in all expression conditions and could thus not be studied.
Table 2.5: Other experimentally-tested $^a$mutants. (i) The $^d k_{\text{cat}}$, $^e K_m$, and $^f k_{\text{cat}}/K_m$ for a set of mutants that showed little or no specificity for the target (Leu) vs. the WT (Phe) $^b$substrate. The $^c K^*$ ranks for the respective mutants with Leu; $^g$this mutant was a top-ranked prediction by the algorithm of [7], but was pruned by the version of $K^*$ described here. $^h$Not detectable. $^i$ $K_m$ and $k_{\text{cat}}/K_m$ cannot be accurately determined because the solubility of Phe ($\sim$180 mM in water) limits the reaction velocity under the experimental condition, in which the velocity remains linearly dependent on the concentration of the substrate. (ii) Single-point $^a$mutants that are components of $K^*$-predicted double-mutants or of predictions from a GMEC-based approach [5]. The $^d k_{\text{cat}}$, $^e K_m$, and $^f k_{\text{cat}}/K_m$ for each mutant with the respective $^b$substrate are shown. (iii) $^a$Mutants difficult to purify due to $^g$solubility issues are shown with their $K^*$ $^c$ranks and the target substrates.

<table>
<thead>
<tr>
<th>Mutant $^a$</th>
<th>Substrate $^b$</th>
<th>Rank $^c$</th>
<th>$k_{\text{cat}}$ $^d$ (min$^{-1}$)</th>
<th>$K_m$ $^e$ (mM)</th>
<th>$k_{\text{cat}}/K_m$ $^f$ (mM$^{-1}$ min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A301G/I330F</td>
<td>Leu</td>
<td>7</td>
<td>3.66</td>
<td>122.24</td>
<td>0.03</td>
</tr>
<tr>
<td>A301G/I330F</td>
<td>Phe</td>
<td></td>
<td>3.1</td>
<td>59.0</td>
<td>0.05</td>
</tr>
<tr>
<td>A301G/I330W</td>
<td>Leu</td>
<td>$^g$</td>
<td>$^g$</td>
<td>$^g$</td>
<td>$^g$</td>
</tr>
<tr>
<td>A301G/I330W</td>
<td>Phe</td>
<td>$^h$</td>
<td>&gt; 0.32 $^f$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mutant $^a$</th>
<th>Substrate $^b$</th>
<th>$k_{\text{cat}}$ $^d$ (min$^{-1}$)</th>
<th>$K_m$ $^e$ (mM)</th>
<th>$k_{\text{cat}}/K_m$ $^f$ (mM$^{-1}$ min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A236M</td>
<td>Phe</td>
<td>8.1</td>
<td>0.0834</td>
<td>97.12</td>
</tr>
<tr>
<td>A236M</td>
<td>Leu</td>
<td>21.1</td>
<td>217.2</td>
<td>0.097</td>
</tr>
<tr>
<td>A322M</td>
<td>Phe</td>
<td>20.2</td>
<td>23.2</td>
<td>0.87</td>
</tr>
<tr>
<td>A322M</td>
<td>Leu</td>
<td>9.4</td>
<td>640</td>
<td>0.015</td>
</tr>
<tr>
<td>T278E</td>
<td>Arg</td>
<td>1.4</td>
<td>80.0</td>
<td>0.018</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mutant $^a$</th>
<th>Rank $^c$</th>
<th>Soluble $^g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T278E/A301G</td>
<td>3 (Arg)</td>
<td>no</td>
</tr>
<tr>
<td>T278D/I299E</td>
<td>2 (Lys)</td>
<td>no</td>
</tr>
</tbody>
</table>
2.4.5 Pre-steady-State Binding Analysis

The steady-state analysis of WT and mutant PheA have provided evidences on substrate selection upon catalysis. However, the $K_M$ values are not completely correlated well to the reported $K_d$ values (90). To further our understanding on substrate binding, we dissected the adenylation reaction by measuring the rates of substrate association and dissociation before substrate turnover. Fluorescence quenching experiments were performed under stopped-flow conditions to obtain the microscopic rates of association and dissociation ($k_1$ and $k_{-1}$), as well as to determine dissociation constants $K_d$. Similar to previous report (δ), no fluorescence quenching in the first 500 ms was detected when using WT PheA and its natural substrate Phe in the absence of ATP and Mg$^{2+}$. Titration of Phe into WT PheA in the present of saturated ATP and Mg$^{2+}$ showed a double exponential quenching curve that indicates a two-step binding event (Figure 2.10). For substrate Leu binding to WT PheA, previous report showed no quenching was detected in the first 500 ms in the presence of ATP and Mg$^{2+}$ (δ). However, in our experiment, fluorescence quenching was observed in the first 500 ms by titrating Leu into WT PheA. The quenching curve can fit into a single exponential function and $k_1/k_{-1}$ was therefore derived from fitting the $k_{obs1}$ to the linear function (Table 2.6).
Figure 2.10: Fluorescence quenching of WT PheA by phenylalanine in the presence of ATP and Mg²⁺. The insets show the fitting of $k_{obs1}$ and concentration of Phe (left) and the fitting of $k_{obs2}$ and concentration of Phe (right). The lower panel shows the residual trace of the double exponential fitting (blue line).
We next measured the fluorescence quenching using mutant PheAs including A301G, T278M/A301G, and T278L/A301G with substrate Phe and Leu in presence of ATP and Mg$^{2+}$. No quenching was detected within 500 ms using Phe with all the mutants while they all showed single exponential quenching by binding to Leu. All three mutants showed improved $K_d$ for the substrate Leu over WT PheA (Table 2.6).

Table 2.6: Rate constant and free energy change of binding with wild-type and mutant PheA.

<table>
<thead>
<tr>
<th></th>
<th>$k_1$ (μM$^{-1}$ s$^{-1}$)</th>
<th>$k_2$ (s$^{-1}$)</th>
<th>$k_d$ (μM)</th>
<th>$\Delta G$ (kcal/mole)</th>
<th>$\Delta G^\ddagger$ from $k_1$ (kcal/mole)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT – L-Phe</td>
<td>$9.2 \times 10^{-1}$</td>
<td>20.73</td>
<td>22.53</td>
<td>-6.34</td>
<td>2.56</td>
</tr>
<tr>
<td>WT – L-Leu</td>
<td>$2.7 \times 10^{-5}$</td>
<td>0.65</td>
<td>23700</td>
<td>-2.22</td>
<td>8.75</td>
</tr>
<tr>
<td>A301G – L-Leu</td>
<td>$4.6 \times 10^{-3}$</td>
<td>2.42</td>
<td>526.1</td>
<td>-4.47</td>
<td>5.70</td>
</tr>
<tr>
<td>T278M/A301G – L-Leu</td>
<td>$4.0 \times 10^{-3}$</td>
<td>0.13</td>
<td>32.5</td>
<td>-6.12</td>
<td>5.78</td>
</tr>
<tr>
<td>T278L/A301G – L-Leu</td>
<td>$1.8 \times 10^{-2}$</td>
<td>0.08</td>
<td>4.44</td>
<td>-7.30</td>
<td>4.89</td>
</tr>
</tbody>
</table>

2.4.6 Structural Analysis

Structural comparison between the predicted mutant structures and the WT can reveal insights into the reasons for the switch of specificity in the Leu redesigns. We thus generated and visualized the structures in the $K^*$ bound-state ensemble for the T278L/A301G mutant. When overlayed with the WT structure, all conformations of the
Leu substrate found in the $K^*$ ensemble clash sterically with the side-chain of Ala at position 301. The mutation A301G appears to free up the space necessary to accommodate the Leu side-chain. In addition, the mutation T278L fills up the enzyme pocket to partially compensate for the change in the substrate size from the bulkier Phe to the smaller Leu. The lowest-energy T278L/A301G structure with Leu as substrate is shown in Figure 2.8.

To analyze the effect of the double mutation T278D/A301G on the enzyme specificity for Arg, we generated and visualized the five lowest-energy structures from the $K^*$ ensemble for this double mutant, as well as the five lowest-energy structures for the single point mutations T278D and A301G. To facilitate the structural comparison, we used MolProbity (98) server. The results are summarized in Table 2.7. The comparison between the structures for A301G (which includes the WT Thr at position 278) and T278D/A301G suggests that the addition of the T278D mutation allows the side-chain of the Arg substrate to participate in stronger hydrogen bonding and/or electrostatics interactions with the charged carboxyl group of Asp at 278. The Arg substrate conformations in all five T278D/A301G structures fill the space that is otherwise occupied by the Ala side-chain at 301. In the T278D structures, the Arg atoms are pushed away from the A301 side-chain mainly through changes in the Arg side-chain dihedrals. This mostly alleviates the Arg-A301 steric clashes (although in some structures of the single-point mutant these clashes are still significant), at the cost of
introducing new significant clashes with other residues in the active site. Similarly to the Leu redesigns, the mutation A301G therefore appears to free up the space necessary to accommodate the Arg substrate. The second lowest-energy T278D/A301G structure from the $K^*$ ensemble with Arg as substrate is shown in Figure 2.11.

Table 2.7: MolProbity analysis for the five lowest-energy T278D and T278D/A301G structures from the respective $K^*$ ensembles with Arg as substrate. The clashscore for a given structure is defined as the number of atom pairs with van der Waals overlap greater than 0.4 Å, per 1000 atoms; lower clash scores are better since they correspond to fewer significant clashes. For each structure, the list of steric overlaps > 0.4 Å between an atom from the Arg substrate and atoms from residues in the enzyme active site are shown. For each structure, the Arg side-chain dihedrals (in parenthesis) are shown along with the corresponding closest Arg rotamer.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Structure</th>
<th>Clashscore</th>
<th>Arg Atom$^e$</th>
<th>Arg Clashes$^c$</th>
<th>Other Atom$^f$</th>
<th>Overlap$^g$</th>
<th>Arg Rotamer$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T278D</td>
<td>1</td>
<td>18.02</td>
<td>NH2</td>
<td>CD1 (W239)</td>
<td>CB (A236)</td>
<td>0.544</td>
<td>mmt180 (307,286-4,189,189)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>22.52</td>
<td>NH1</td>
<td>HA (A236)</td>
<td>HB3 (A301)</td>
<td>0.555</td>
<td>mmt-85 (289.6,283,189,284)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>18.02</td>
<td>NH2</td>
<td>CD1 (W239)</td>
<td>CB (A236)</td>
<td>0.544</td>
<td>mmt180 (307,286-4,189,189)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>22.52</td>
<td>NH1</td>
<td>HA (A236)</td>
<td>HB3 (A301)</td>
<td>0.551</td>
<td>mmt-85 (289.8,283,189,284)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>24.02</td>
<td>HH1</td>
<td>CB (A301)</td>
<td>CB (A301)</td>
<td>0.862</td>
<td>mmt85 (306.9,288,3,171,76)</td>
</tr>
<tr>
<td>T278D/A301G</td>
<td>1</td>
<td>15.08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>mmt85 (284.3,189,171,80.1)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>16.59</td>
<td>HH12</td>
<td>CG (D278)</td>
<td>–</td>
<td>0.401</td>
<td>mmt-85 (301.7,171,173,284)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>15.08</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>mmt85 (284.7,189,171,79.6)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>16.59</td>
<td>HH12</td>
<td>CG (D278)</td>
<td>–</td>
<td>0.402</td>
<td>mmt-85 (301.7,171,173,1,284)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>16.59</td>
<td>HH11</td>
<td>HA3 (G301)</td>
<td>–</td>
<td>0.406</td>
<td>mmt85 (301.5,284,7,179,3,77)</td>
</tr>
</tbody>
</table>
Figure 2.11: $K^*$-predicted structure of the second lowest-energy T278D/A301G conformation with Arg as substrate. Shown are the Arg substrate and the two active site mutations 278D and 301G (CPK), the other active site residues (cyan), and the AMP cofactor (gray). Interactions between the substrate side-chain with 278D (the distance between $N_{\eta 1}$ (Arg) and $O_\delta 2$ (278D) is 2.92 Å; the distance between $N_{\eta 2}$ (Arg) and $O_\delta 1$ (278D) is 3.15 Å) and the substrate backbone with D235 and K517 are shown with dashed yellow lines. Viewing angle is chosen to show side-chain interactions between Arg and 278D.
2.5 Discussion

Several studies have shown that the substrate specificity of the adenylation domain can be modified by the mutation of the active site residues (5, 8, 99). By using a multiple sequence alignment approach to redesign the substrate specificity of GrsA-PheA, Stachelhaus et al. (5) successfully improved the activity of the enzyme for the non-cognate amino acid Leu with the introduction of a double mutation T278M/A301G, and altered the substrate specificity of an aspartate-activating domain AspA to Asn by a single mutation H322E in the active site. The sequence-based approach identifies active site residues important for the substrate specificity by comparing the corresponding moieties among different adenylation domains. However, its accuracy depends heavily on the number and diversity of available sequences. In contrast, our $K^*$ algorithm uses the structure of the PheA domain as well as an amino acid rotamer library and a molecular mechanics energy function as inputs. For a given amino acid substrate, the algorithm was able to search a space of thousands of sequences with hundreds of millions of conformations. By computing the partition functions over the conformational ensembles, the $K^*$ algorithm scores sequences based on their approximation to the binding constant. As a result, the top-scored sequences were expected to have a lower $K_d$ and consequently a lower $K_M$ for the target substrate. The feasibility of the $K^*$ algorithm was shown by the lower $K_M$ value with Leu of the top-scored mutants, T278L/A301G, T278M/A301G and A301G/A322V.
Sequence-based methods are limited to the active site signature motif. Hence, the sequence alignment approach can only identify local regions (such as the active site) where a significant sequence homology exists. It has been suggested that distal residues outside the active site might play critical roles in stabilizing protein function (100). This idea was incorporated into our computational protocol with the identification of the bolstering mutations outside the active site. The addition of the predicted bolstering mutations in the Leu redesigns had a significant impact on the substrate specificity of the enzyme. Since residue 277 is adjacent to the active site mutation T278L, the mutation I277L could directly affect the conformation of the enzyme active site, also affecting the substrate specificity. Residues 187 and 447, however, are distal from the ligand binding site and their impact is likely due to indirect and/or long-range interactions.

Interestingly, structural analysis of the lowest-energy S447N conformation predicted by the MinDEE/A^* algorithm (18) shows that the Asn side-chain reaches across a solvent channel inside the protein, making a hydrogen bond with backbone carbonyl oxygen of H344 (Figure 2.12). To further verify if the distal mutations affect the stability of the proteins, we performed guanidine-HCl titration experiment which monitored the circular dichroism change upon denature of wild-type and mutant PheA proteins. Both triple mutant I277L/T278L/A301G and quadruple mutant I277L/T278L/A301G/S447N showed almost identical denaturing curves as the double mutant T278L/A301G and WT PheA. The finding suggested that the distal mutations did not affect the overall stability.
of the protein. The improvement of the specificity for the non-cognate substrate might come from a more stable catalytic conformation contributed by the distal mutations.

Figure 2.12: The lowest-energy S447N conformation predicted by MinDEE/A*. The hydrogen bond between H344 and N447 (the distance between N\(\delta_2\) (Asn) and the backbone carbonyl oxygen of His is 2.94 Å) is shown with a dashed yellow line. The Leu ligand is shown in orange.

The ability of the algorithm to search a large space of sequences and conformations enables us to redesign the active site for a diverse set of substrates. We tested this capability of the algorithm by predicting mutations for charged amino acids whose activity was not found in PheA. In order to stabilize the charged side-chain of the substrates, the algorithm introduced polar or charged residues in the active site, which resulted in our successful mutations, T278D/A301G to bind Lys and Arg, T278H/A301G to bind Glu, and T278K/A301G to bind Asp. Interestingly, residue positions 278 and 301
were again chosen by the algorithm but with a different residue type at position 278. A previous report has shown that mutation at a single key position His322 (to Glu322) in the active site of the adenylation domain AspA from the surfactin synthetase B is sufficient to obtain the specificity switch from Asp to Asn (99). This finding, combined with our results, suggest that in GrsA-PheA, positions 278 and 301 might play key roles in the recognition of the substrate. Structural analysis of the $K^*$ models of the mutants suggests that Gly301 might alleviate steric clashes in order to bind different substrates. Residue 278 might be involved in direct interactions with the substrate side-chain.

A comparison of our computationally-predicted mutant active sites to a set of NRPS enzymes selected by evolution showed that although the amino acid identities at mutated positions were found as constituents of longer signature sequences, none of our exact mutant active sites could be found in that enzyme set. Moreover, a comparison to the predictions from two sequence-based methods showed that our structure-based method could identify active mutants different from the sequence-based predictions.

The mechanism of substrate recognition by the adenylation domain of NRPS has been puzzling. Luo et al. claimed that the discrimination of the amino acid substrate begins when the transition state is formed during the catalysis (90). Stevens et al. suggested that a conformational change toward a catalytically-relevant intermediate occurs in the adenylation process of PheA (8). In our results, the double mutant T278L/A301G dramatically lowers the value of $K_M$ for the non-cognate amino acid Leu
from the WT PheA, requiring changes to only two residues in the active site. It is therefore intriguing to see if there exists any interaction between Leu278 and Gly301 in binding of Leu. Hence, we then investigated the double mutant T278L/A301G and the single mutants T278L and A301G by analyzing their free energy change upon adenylation of Phe and Leu. Detailed free energy calculations are described in the section 2.2.6 and the kinetics constants as well as difference in free energy are listed in Table 2.8. The result shows that the WT protein has a difference in binding energy of 3.22 kcal/mol in favor of Phe while the energy difference becomes 0.49 kcal/mol in favor of Leu in the T278L/A301G mutant. The free energy barrier required for the discrimination of Phe and Leu in the WT protein was decreased to favor Leu in the mutant protein. Moreover, a coupling energy ($\Delta \Delta G_{int}$) of 1.69 kcal/mol was observed when comparing the free energy difference of the T278L/A301G mutant ($\Delta \Delta G_{WT-T278L/A301G}$) and the two corresponding single mutants T278L ($\Delta \Delta G_{WT-T278L}$) and A301G ($\Delta \Delta G_{WT-A301G}$) in binding of Leu. The coupling energy suggests that the two active site residues, Leu278 and Gly301 might interact to provide a favorable conformation for the recognition of the Leu substrate.
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td>1.73 ± 0.29</td>
<td>0.0018 ± 0.0004</td>
<td>951.4 ± 111.2</td>
<td>11.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A301G</td>
<td>0.20 ± 0.03</td>
<td>0.0005 ± 0.0001</td>
<td>353.7 ± 29.4</td>
<td>12.31 0.59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T278L</td>
<td>1.41 ± 0.13</td>
<td>0.046 ± 0.009</td>
<td>30.74 ± 3.43</td>
<td>13.76 2.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A301G/T278L</td>
<td>3.37 ± 0.08</td>
<td>0.097 ± 0.013</td>
<td>34.94 ± 4.76</td>
<td>13.68 1.96 0.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td>28.74 ± 1.58</td>
<td>6.98 ± 1.00</td>
<td>4.15 ± 0.36</td>
<td>14.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A301G</td>
<td>0.89 ± 0.05</td>
<td>0.042 ± 0.008</td>
<td>21.48 ± 3.22</td>
<td>13.97 -0.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T278L</td>
<td>23.68 ± 2.30</td>
<td>26.28 ± 1.62</td>
<td>0.90 ± 0.04</td>
<td>15.85 0.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A301G/T278L</td>
<td>1.16 ± 0.10</td>
<td>0.015 ± 0.002</td>
<td>79.49 ± 13.67</td>
<td>13.19 -1.75 1.69</td>
<td></td>
</tr>
</tbody>
</table>
3. NMR Studies of the Epimerization Domain of GrsA

3.1 Introduction

One unique feature of nonribosomal peptides (NRPs) is the presence of D-amino acids, which provide NRPs unique conformations and configurations for orientating the growing peptides as seen in penicillin and vancomycin (101). In nonribosomal peptide synthetase (NRPS), incorporation of D-amino acids can occur by two routes: 1) direct activation of a D-amino acid via an A domain, such as the sixth A domain in the fusaricidin NRPS (102) or 2) in situ epimerization of the Ca center of the PCP-bound L-amino acid during peptide elongation. The second mechanism is the more common route and is catalyzed by the ~450 a.a. epimerization (E) domain of NRPS. E domains are auxiliary domains that are embedded either in the initiation modules as in gramicidin S synthetase or in internal modules as in the third module of AcvA synthetase (103).

Sequence alignment and secondary structure predictions of the E domain family have suggested that the E domain shares similar structure and sequence motifs with the C domain family of NRPS (31). The epimerization reaction is reversible, as a result, an equilibrium between both isomers can be found as the end products. However, only the D-aminoacyl- or peptidyl-PCP-bound substrate can be recognized by the donor site of the downstream C domain.

Biochemical and mutational studies of the E domain in the initiation module of gramicidin S synthetase (GrsA-PheE) have identified several key catalytic residues
including the H753 of the His-motif, D757, and Y976 to be essential for proton transfer at the Ca center of the PCP-bound L-Phe to form D-Phe (104). Substrate specificity of E domains has been puzzling. As shown in the studies of GrsA-PheE, the E domain selectively epimerized the PCP-bound cognate substrate L-Phe but not non-cognate substrates or free amino acids (90). It was therefore proposed that the E domain, in concert with the PCP domain, contributes to some degree of substrate specificity of NRPSs. Other studies using E domain of the initial module and the peptidyl-bound domains as the substrates showed that the E domain in the initial module was able to epimerize the peptidyl intermediates (105). More recently, a short fragment from the C-terminal region of E domain has been identified to play a role in the recognition of downstream modules (36). This short fragment also called communication-mediating (COM) domain specifically recognizes an acceptor region of downstream modules. The structural basis of the inter-module communication and substrate specificity provided by the E domains has been unknown.

The large size of the epimerization domain presents a challenge for structure determination by traditional NMR approach, which requires a large number of NOE distance restraints for global fold determination. In this study, we used a novel structure determination framework, called RDC-ANALYTIC (Figure 3.1). RDC-ANALYTIC is a suite of programs for high-resolution protein backbone structure determination from residual dipolar couplings (RDCs) and residual chemical shift anisotropies (RCSAs) and
requires only two RDCs/RCSAs per residue in one or two alignment media with a sparse set of nuclear Overhauser effect (NOE) data (106, 107). In the RDC-ANALYTIC framework, we first compute the orientations and conformations of secondary structure elements (SSEs) from RDCs. SSEs were then packed by sparse NOE restraints and/or by long range distance restraints obtained from paramagnetic relaxation enhancements (PREs). The conformations of loops can also be computed from the RDCs by the POOL module (107). The computed loops along with the packed SSEs are fed to structure refinement protocol in XPLOR/XPLOR-NIH for final backbone structure calculation.

In this chapter, we present the NMR studies of GrsA-PheE. The backbone resonance assignments are reported, that provide a foundation for future studies of both substrate binding and domain-domain interactions. For global fold determination, we report measurements of $^1$H-$^{15}$N$^\alpha$ and $^{13}$C$^\alpha$–$^{13}$C$'$ RDCs in three alignment media, and PRE distance restraints from 15 site-specific spin labeling samples.
3.2 Experimental procedures

3.2.1 Materials

All reagents, unless otherwise specified, were purchased from VWR International (West Chester, PA). Stable isotopes were purchased from Cambridge Isotope Laboratory (Andover, MA). Pf1 filamentous phage was purchased from ASLA, Ltd. (Riga, Latvia). Spin-label reagents, (1-oxy-2,2,5,5-tetramethyl-D3-pyrroline-3-methyl)methanethiosulfonate (MTSSL) and (1-acetyl-2,2,5,5-tetramethyl-D3-pyrroline-3-methyl)methanethiosulfonate (dMTSSL) were purchased from Toronto Research Chemicals Inc (Toronto, Ontario, Canada). Human $\alpha$–thrombin was purchased from Haematologic Technologies Inc. (Essex Junction, VT).
3.2.2 Cloning and Mutagenesis of PheE

The gene fragment of PheE domain were PCR amplified from the full-length PheATE using the primers containing NdeI and EcoRI site (Table 3.1). PheE was inserted into the pET28b expression vector with an N-terminus His-tag followed by a thrombin cutting site and the gene. The plasmid was then transformed into *Escherichia coli* BL21 (DE3) cells for expression.

Mutagenesis was performed using the QuickChange site-directed mutagenesis system (Stratagene) in accordance with the manufacturer's instructions by using primers summarized in Table 3.1. Preparation of the plasmid DNA was done in *E. coli* DH5α following standard procedures. All constructs were confirmed by DNA-sequencing.
### Table 3.1: Cloning and Mutagenesis primers for PheE

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PheE_N_NdeI:</strong></td>
<td>ATATACATATGAGT GAGCAAGGTATTTGTG GAAGGTGAG</td>
</tr>
<tr>
<td><strong>PheE_C_EcoRI:</strong></td>
<td>ATATAGAATTCTTGCTTACTAGGTAAATCGGCAACAAATCG</td>
</tr>
<tr>
<td><strong>C106A-F:</strong></td>
<td>GGTGATTGCTGAAGAATCTGCTCG</td>
</tr>
<tr>
<td><strong>C106A-R:</strong></td>
<td>CTTCAGCATGTTCAATGATGGCC</td>
</tr>
<tr>
<td><strong>C453A-F:</strong></td>
<td>GAACATGC TGTACAGAAGGAAGATAC</td>
</tr>
<tr>
<td><strong>C453A-R:</strong></td>
<td>GTCCAATCTCACCACACAAATACCTTGCTC</td>
</tr>
<tr>
<td><strong>E11C-F:</strong></td>
<td>GAGCAAGGTATTGTGTTGGTATGGAGATGGAC</td>
</tr>
<tr>
<td><strong>E11C-R:</strong></td>
<td>GTTCTCTCCACACACAAATACCTTGCTC</td>
</tr>
<tr>
<td><strong>Q27C-F:</strong></td>
<td>GCATTGAGGCTTGGACAAAAATGCAACCAGTGG</td>
</tr>
<tr>
<td><strong>Q27C-R:</strong></td>
<td>CCAATGGTCTATTGTATTTTACAAATGCAACCAGTGG</td>
</tr>
<tr>
<td><strong>N44C-F:</strong></td>
<td>GTTGTATTAGACCATGTTCGGGTTTGATAAAGAG</td>
</tr>
<tr>
<td><strong>N44C-R:</strong></td>
<td>CTCTTTATCAAACACATGGTCTATACAAC</td>
</tr>
<tr>
<td><strong>D182C-F:</strong></td>
<td>GAGAAAACAGATTCAGACATGCTGGTCTATTGAATTAG</td>
</tr>
<tr>
<td><strong>D182C-R:</strong></td>
<td>CTAATTCAATAGACCACATTAATGCTGTTATTC</td>
</tr>
<tr>
<td><strong>K189C-F:</strong></td>
<td>GACTGGTCTATTGAATTGTTAGTGGAACAGCGAAT</td>
</tr>
<tr>
<td><strong>K189C-R:</strong></td>
<td>ATTCGCTGTTCGATAACATTTCAATAGACCAGATG</td>
</tr>
<tr>
<td><strong>E201C-F:</strong></td>
<td>GCGAATTATTCCTAGAAGAAGCTTGTTATTGGCATTTG</td>
</tr>
<tr>
<td><strong>E201C-R:</strong></td>
<td>CAAATGATGCAATCCAAACATGGTTCTCTAGGAATATTC</td>
</tr>
<tr>
<td><strong>Y208C-F:</strong></td>
<td>GGCATCATTTGAATTTGTTATACCGAAGACGTTG</td>
</tr>
<tr>
<td><strong>Y208C-R:</strong></td>
<td>GAGCTTTTTAGCTTAAAACTCAGGAAGATG</td>
</tr>
<tr>
<td><strong>E242C-F:</strong></td>
<td>GAGTTAACAATAGAAGAGACATGCTGGTCTATTGAATTAG</td>
</tr>
<tr>
<td><strong>E242C-R:</strong></td>
<td>CATTTTTCAATTTCAATGCTGTTATAGCTAATC</td>
</tr>
<tr>
<td><strong>K314C-F:</strong></td>
<td>GTTGTACTTGAATGCAATGGTGATGATGG</td>
</tr>
<tr>
<td><strong>K314C-R:</strong></td>
<td>CAATGCTCCGAACATGGCATATCAAGTACAAAC</td>
</tr>
<tr>
<td><strong>K323C-F:</strong></td>
<td>GGATGATTGCTTATCAATGCTGTTTAGAAGAAAGAATTTACG</td>
</tr>
<tr>
<td><strong>K323C-R:</strong></td>
<td>CGTAAATTTTTCTTACATTAAACAGATTTGATAAGAAATCATCC</td>
</tr>
<tr>
<td><strong>R440C-F:</strong></td>
<td>CCATTTGAAATTTGAGCTGTAGGATTGCATGGAT</td>
</tr>
<tr>
<td><strong>R440C-R:</strong></td>
<td>GATGTTGCTTATAGCAGTCAATGCTGAATGG</td>
</tr>
<tr>
<td><strong>T434C-F:</strong></td>
<td>GCAGTATACAGGAGGATGCTGACATGCAATGAGCC</td>
</tr>
<tr>
<td><strong>T434C-R:</strong></td>
<td>GCTCAATTGCTGAATGCAATCTCCTCTATATAC</td>
</tr>
</tbody>
</table>

68
3.2.3 Protein Expression, Purification and Sample Preparation for NMR

Except for the selectively $^{15}$N-enriched samples, wild-type and mutant PheE were overexpressed in *Escherichia coli* BL21 (DE3) cells on a pET28b vector. The samples labeled with $^{15}$N-Val, Ile, Leu, Lys, Phe or Tyr were overexpressed in *Escherichia coli* DL39 auxotrophic strain on a pET15b vector. The sample labeled with $^{15}$N-His was overexpressed in *Escherichia coli* histidine auxotroph BL21 (DE3) (ΔhisB) on a pET28b vector. The sample labeled with $^{15}$N-Met was overexpressed in *Escherichia coli* B834 (DE3) auxotrophic strain on a pET28b vector. Cells used for preparing NMR sample were grown in M9 minimal medium with different isotopes enrichment. Isotope enriched protein was prepared using $[^{13}$C]glucose and $^{15}$NH$_4$Cl as the sole carbon and nitrogen sources. Deuterium labeling was carried by growing the cells in M9 with 100% D$_2$O and deuterated $[^{13}$C]glucose.

The labeled PheE proteins were expressed by induction of mid-log cells (OD~0.8 for normal M9 culture, OD~0.4 for deuterated M9 culture) with 0.2 mM IPTG at 30 °C overnight. In a typical preparation of 1.0 L M9 culture, cell pellets were re-suspended in 30 mL of buffer A (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 2 mM TCEP) supplemented with a protease inhibitor cocktail. The cells were lysed by French-press and cell debris was removed by centrifugation at 20000xg for 30 min. The resulting supernatant was incubated with Ni sepharose 6 fast flow resin (GE Healthcare, Chalfont St. Giles, United Kingdom) at 4 °C for one hour. The resin was then washed extensively.
with buffer B (50 mM Tris-HCl, pH 8.0, 250 mM NaCl, 20 mM imidazole and 1 mM TCEP). The his-tagged proteins were eluted with 350 mM imidazole (pH 7.5) and further purified by a Superdex 200 gel filtration chromatography (GE Healthcare, Chalfont St. Giles, United Kingdom) in buffer C (20 mM Tris-HCl, pH 8.0, 1 mM TCEP). The purified proteins were concentrated to ~1.0 mL and incubated with thrombin (10~30 unit) to remove His tags. The thrombin and uncleaved proteins were removed by passing the reaction mixture through Ni sepharose resin and a benzamidine sepharose column. The purified proteins were further concentrated and buffer-exchanged to NMR buffer (20 mM Tris, pH 8.0, 10% D$_2$O, 0.25 mM TCEP). For deuterated sample, the purified protein were further denatured in 8.0 M urea and incubated for at least 2 hours at room temperature to allow a complete $^1$H-hydrogen exchanged back to amide proton. To refold the protein, the protein-urea solution was diluted at least 10-fold rapidly into the refolding buffer (20 mM Tris pH 8.0, 1mM TCEP). The refolded proteins were concentrated and buffer-exchanged in 20 mM deuterated Tris-HCl pH8.0, 10% D$_2$O, 0.25 mM TCEP for NMR experiments.

**3.2.4 NMR Spectroscopy and Backbone Resonance Assignment**

All NMR experiments were performed on either a 600 MHz or 800 MHz Varian Innova spectrometer at 45 °C except for the RDC and PRE measurements which were performed at 32 °C and 37 °C respectively. Each spectrometer was equipped with a triple-resonance cryogenically cooled probe (25 °K) designed to accommodate 5 mm
samples. FIDs were processed using NMRPIPE (Delaglio et al. 1995) and datasets analyzed using Sparky (T. D. Goddard, and D. G. Kneller, SPARKY 3, University of California, San Francisco).

For backbone resonance assignment, a suit of TROSY based triple-resonance experiments, including HN(Ca), HN(Co)Ca, HN(Ca)Cb, HN(CoCa)Cb, HNCo and a “just in time” HN(Ca)Co (108) were performed using uniformly $^2$H$^{13}$C$^{15}$N-labeled PheE. Selectively $^{15}$N-enriched PheE including $^{15}$N-Val, Ile, Leu, Lys, Asp, Phe, Tyr, His, and Met were used to collect HSQC spectra. A 4D TROSY-HNCaCb and 4D TROSY-HN(Ca)NH (109) experiments were performed with sparse sampling using the uniformly $^2$H$^{13}$C$^{15}$N-labeled PheE. 3D $^{15}$N-NOESY-TROSY experiment was performed using uniformly $^2$H$^{13}$C$^{15}$N-labeled, ILV-methyl-protonated PheE. The PACES algorithm was used to identify resonance connectivity (110). TALOS+ was used for prediction of secondary structures from chemical shift derived from the backbone resonance assignments (111).

### 3.2.5 RDC Sample Preparation and Measurement

The Phage medium for RDC measurement was prepared by gently mixing the $^2$H$^{13}$C$^{15}$N-labeled PheE with the stock solution of Pf1 filamentous phage (~50 mg/mL) to a final concentration of ~500 uM of protein and ~15 mg/mL of phage in a buffer containing 10% D$_2$O, 20mM Tris (pH 8.0) and 100 mM NaCl. The protein/phage sample
gave a D$_2$O quadrupolar splittings of ~11.0 Hz and the quadrupolar splitting stayed at a similar range after two weeks of RDCs measurement at 32 °C.

Stretch polyacrylamide gels were prepared as described previously using the devices from New Era Enterprises Inc (112). Stock solution of 40% w/v acrylamide and 2% w/v N,N'-methylenebisacrylamide were mixed in a 29:1 ratio and diluted to a final acrylamide concentration of 4% in 20 mM Tris (pH 8.0) buffer. Immediately prior to polymerization, all the solutions (H$_2$O, acrylamide/bisacrylamide mixture, and buffer) were degassed thoroughly. Polymerization was initiated by the addition of 0.1% ammoniumperoxide sulfate and 1% tetramethylethylenediamine (TEMED). A volume of 400 μL mixtures was allowed to polymerize in the 5 mm diameter gel chamber (New Era) overnight at 30 °C. The polymerized gels were first washed in deionized water for one hour and then extensively washed in Tris buffer (20 mM Tris, pH 8.0) for at least three hours and repeated the washing procedure for at least two times. The gels were then dried at room temperature for three hours. About 400 μL of protein solution (~0.8 mM protein in 40 mM Tris, pH 8.0 and 18% D$_2$O) was incubated with the dried gel at room temperature for overnight to allow sufficient time for protein diffusing into the gel. The protein-containing gel was washed briefly with NMR buffer (20 mM Tris, pH 8.0 and 9% D$_2$O) and transferred back into the 5 mm diameter gel chamber. The gel was stretched into the 4.2 mm diameter NMR tube from the chamber using the device from New Era. To introduce negative charge, stock solution of 40% w/v acrylic acid was used.
to make the polymerization mixture with a final concentration of 5% acrylic acid relative to the total copolymer (4% acrylamide). After the washing cycles, the charged gel became significant larger than a regular gel due to strong electroosmotic swelling. Therefore the gel was dried for a longer time allowing it to shrink back to the normal size.

Residual dipolar couplings for bond vectors $^1$H-$^{15}$N ($^1$D$^{	ext{NH}}$) and $^{13}$C$^\alpha$-$^{13}$C' ($^1$D$^{	ext{C}_\alpha C'}$) were determined from the difference in one bond couplings between an isotropic sample and weakly aligned samples in phage or polyacrylamide gels. A 2D $^1$H-$^{15}$N IPAP experiment and a set of 3D HNCO-based experiments for the measurement of $^1$H-$^{15}$N$^\text{H}$ and $^{13}$C$^\alpha$-$^{13}$C' dipolar couplings were performed to measure the $1^\text{DH-N}$ and $1^\text{DC}_\alpha$–C' couplings. Sparky was used for peak analysis.

### 3.2.6 Site-directed Spin Labeling and PRE Measurement

Wild-type PheE has two intrinsic cysteine residues, Cys106 and Cys 453. The two cysteines were mutated to alanine individually to create single cysteine mutants or together to create a non-cysteine construct. A single cysteine was reintroduced back into the non-cysteine construct by site-directed mutagenesis. A total of 15 single cysteine mutants were made. Uniformly $^2$H$^{15}$N-labeled single cysteine mutant proteins were overexpressed and purified as described in section 3.2.3. After gel-filtration chromatography, the proteins were buffer-exchanged at least ~10,000-fold to remove TCEP and concentrated to a volume of ~300 μL. The concentrated proteins were mixed
with 2.0 mL of 8.0 M urea and split into two equal portions for parallel labeling with MTSSL and dMTSSL. Both reagents were added from 500 mM stocks in acetonitrile at a 10-fold molar excess over protein and incubated at room temperature for 1 hour. To ensure complete labeling, another 10-fold molar excess of reagent was added and the mixture was incubated at room temperature overnight. During the incubation, the MTSSL-labeled sample was prevented from light. The final concentration of urea is around 7 M which is essential for denaturing the protein and allowing sufficient $^1$H-hydrogen back-exchanged to backbone amide. Proteins were refolded as described in section 3.2.3 and buffer-exchanged at least 10,000-fold to remove free labeling reagents. The final sample concentration ranges from 0.3 to 0.5 mM.

TROSY-based $^1$H-$^{15}$N HSQC spectra were collected at 45 °C or 37 °C on a 800 MHz Varian Unity Inova NMR spectrometer. Typically, 1024 complex points in the direct dimension ($^1$H) and 128 complex points in the indirect dimension ($^{15}$N) were collected with number of transients at least 16. Spectra were processed and analyzed with NMRPipe and Sparky. Peaks were assigned according to the wild-type PheE spectra. To avoid the inaccurate estimation of peak intensity, peaks with significant overlaps were not assigned.

PRE distances were determined using Battiste and Wagner’s method (48), in which the ratio of peak heights between the paramagnetic ($I_{para}$) and the diamagnetic
samples was used to derive the transverse relaxation rate enhancement, $R_2^{sp}$, which was contributed by the unpaired electron on the spin label using the equation,

$$\frac{l_{para}}{l_{dia}} = \frac{R_2^{exp(-R_2^{sp}t)}}{R_2^{exp}+R_2^{sp}}$$  \hspace{1cm} (3.1)$$

where $R_2$ is the transverse relaxation rate of the resonance in the diamagnetic sample, which was estimated from the half-height linewidth assuming Lorentzian line shape; $t$ is the total evolution time in the proton dimension. Before derive the rate, the height ratio was calibrated using 8 unaffected peaks. The rate was then used to calculate the distance using the modified Solomon-Bloembergen equation,

$$r = \left[ K \left( \frac{4\tau_c}{R_2^{sp}} + \frac{3\tau_c}{1+\omega^2\tau_c^2} \right) \right]^{1/6}$$  \hspace{1cm} (3.2)$$

where $r$ is the distance between the unpaired electron and the nuclear spins (the amide protons); $K$ is a constant, $1.23 \times 10^{-32}$ cm$^6$s$^{-2}$; $\tau_c$ is the correlation time for the electron-nuclear interaction, and is approximated as the global rotational correlation time of the protein; $\omega_h$ is the Larmor frequency of the proton nuclear spin.

### 3.2.7 Structure Computation by RDC-ANALYTIC

The experiments of this section were mainly performed by Chittaranjan Tripathy.

Figure 3.1 shows the framework of RDC-ANALYTIC. The input data of RDC-ANALYTIC include: (1) the primary sequence of the protein; (2) any combination of at least two RDCs (or RCSAs) per residue measured in one (or two) alignment medium; (3) a sparse set of NOEs and/or PREs; (4) secondary structure element (SSE) boundaries.
based on TALOS+ dihedral restraints; and (5) the rotamer library. RDC-ANALYTIC uses a divide-and-conquer approach to partition the protein backbone into fragments of bounded lengths consisting of Secondary Structure Elements (SSEs), and loops of bounded lengths, based on TALOS+ dihedral information. It then uses the RDCs for the SSE fragments and computes their conformations, with respect to a principal order frame (POF) of the RDCs. The SSE fragments computed by RDC-ANALYTIC are properly oriented in the POF modulo the orientational degeneracy due to the symmetry of the dipolar operators. These SSE fragments are then assembled using NOE/PRE distance restraints to determine the global fold.

3.3 Results

3.3.1 Backbone Resonance Assignment

A 2D $^{15}$N heteronuclear single-quantum correlation (HSQC) spectrum of the non-deuterated PheE sample was first collected and later compared to the $^{15}$N-HSQC-TROSY spectrum of the deuterated sample for which the sample was denatured and refolded in water-based buffer to allow $^1$H-hydrogen back-exchanged to amide proton. Both spectra showed almost identical peaks while the deuterated sample showed a much better dispersed peaks with narrower linewidth. The result indicated that the refolded protein was structurally identical to the native protein.

PheE comprises 488 amino acid residues including 13 pro residues. The large number of residues presented a challenge for backbone resonance assignment. A
significant degree of resonance overlap can be found in the central part of the $^{15}$N-HSQC-TROSY spectrum (Figure 3.2). A suit of TROSY-based triple resonance experiments as described in experimental procedures 3.2.4 were collected with a $^2$H$^{13}$C$^{15}$N-labeled protein sample. To proceed with the assignments, selectively $^{15}$N-enriched samples including Val, Ile, Leu, Lys, Phe, Tyr, Asp, Met and His were prepared for the HSQC spectra. These HSQC's with specific amino acid types serve as anchoring point to facilitate the initial assignments. Gly amide resonances were unambiguously identified based on the 3D HN(Ca)Cb experiment. In addition, many Ala and Ser/Thr residues could be identified based on their distinctive low and high C$\beta$ chemical shifts, respectively. Peaks from spectra mentioned above were picked and formulated into a spin system table with atom types as well as amino acid types. Connectivity of resonances was determined by applying the spin system table to the PACES algorithm (110) and the majority (~82%) of the resonances was assigned.Connectivity was confirmed through the backbone resonance analysis of a 4D HN(Ca)Cb experiment (Figure 3.3) and through correlate triplets of contiguous amides from 4D HN(Ca)NH experiment. Amide assignments were also confirmed through analysis of a 3D NOESY-TROSY experiment. Overall, every identifiable backbone amide peak in the $^1$H$^{15}$N-HSQC-TROSY spectrum was assigned, representing 80% of the non-proline residues in PheE. $^1$H and $^{13}$C backbone chemical shifts were used to predict secondary structure elements by TALOS+ (111). The epimerization domain from the initial module of
tyrocidine synthetase shares 61% sequence identity with GrsA-PheE. We therefore generate a homology model based on the crystal structure of the epimerization domain of tyrocidine synthetase (pdbid: 2XHG) using SWISS-MODEL server (113). Our secondary structure prediction using chemical shift information agrees very well to the model structure (Figure 3.4). The agreement suggests that PheE adopts a similar fold.

Figure 3.2: $^{1}H^{15}N$-HSQC-TROSY spectrum of GrsA-PheE collected at 800 MHz at 45°C.
Figure 3.3: Backbone resonance assignment using 4D HNCA/Cb spectrum. Residues Y91T92F93 were used as an example in the analysis. In the lower left panel, the HSQC cross peak of F93 can be identified and its NH position was used to identify Cα Cβ cross peaks of F93 and T92 in the 4D HNCA/Cb spectrum (upper left panel). The Cα Cβ position of T92 was then used to identify its NH cross peak position in the same spectrum (lower right panel).
Figure 3.4: Comparison of the model and predicted secondary structure elements for PheE. Secondary structure elements from the model structure are shown above the sequence, with cylinders and arrows denoting $\alpha$-helices and $\beta$-strands, respectively. Secondary structure elements predicted by TALOS+ based on the $^1$H and $^{13}$C backbone chemical shifts are shown below the sequence, with blue bars representing the probability of the prediction (positive: strands, negative: helices). Residues used for site-directed spin labeling were shown in orange boxes.

### 3.3.2 Residual Dipolar Coupling Analysis

Residual dipolar couplings (RDCs) were measured using TROSY-based 3D HNCO experiments in three media for $^{15}$N-$^1$H$^N$ and two media for $^{13}$C$^\alpha$-$^{13}$C$'$ as described in section 3.2.5 (114). To evaluate the quality and reliability of the data and to estimate
an error range of the RDCs, we compared the back-calculated RDCs from the model structure of PheE to the experimentally measured RDCs. The model structure is first protonated using the MolProbity server (98). The Singular Value Decomposition (SVD) module of RDC-ANALYTIC is then used with the protonated structure and the experimental RDCs to analyze the quality of RDCs. Figure 3.5 shows the fitting of the experimental RDCs to the model RDCs for $^1$H-$^{15}$N bond vectors in phage, stretched polyacrylamide gel and acid stretched gel and for $^{13}$C$^\alpha$–$^{13}$C' bond vectors in phage and stretched polyacrylamide gel. For $^1$H-$^{15}$N bond vectors, the RDCs collected in phage give the best fit with a Pearson correlation coefficient of 0.879 and a RDC RMSD of 5.596 Hz. The fitting of RDCs collected in acid gel has a correlation coefficient of 0.728 and a RDC RMSD of 4.792 Hz. The RDCs collected in stretched polyacrylamide gel gives a poor fit to the model RDCs, which indicated a poor alignment in this medium. For $^{13}$C$^\alpha$–$^{13}$C' bond vectors, we got a correlation coefficient of 0.621 and a RDC RMSD of 2.83 Hz in phage. Again the RDCs collected in stretched polyacrylamide gel give a poor fitting to the model RDCs.
Figure 3.5: Back-calculated RDCs from the model vs experimental RDCs. Upper panel shows the $^{13}C^{\alpha}-^{13}C'$ RDCs collected in phage and stretched polyacrylamide gel. Lower panel shows the $^1H-^{15}N$ RDCs collected in phage, stretched polyacrylamide gel and acid stretched gel.

Bond vectors from at least two independent alignment media are required to determine the correct orientation of secondary structures. We next investigated the $^1H-^{15}N$ RDCs in phage and in the acid stretched gel to see if the two alignment media are independent. Figure 3.6A shows a similar distribution of $^1H-^{15}N$ RDCs along the amino acid sequence. The axial component of alignment tensor and the rhombicity can be estimated from the histogram analysis, which gives $1.7 \times 10^{-3}$ and $1.2 \times 10^{-3}$ of axial component and $0.705$ and $0.58$ of rhombicity for phage and acid gel respectively (Figure 3.6B). We next analyzed the $^1H-^{15}N$ RDCs for each individual residue by visualizing the
RDCs in the principle order frame. Figure 3.6C shows the RDC curves of residue 437.

The two curves visualized using RDCs measured in phage and in acid stretched gel show a significant amount of overlaps. These results suggested that the RDCs measured in phage and in acid stretched gel might not be independent.

Figure 3.6: Distribution of $^1$H-$^{15}$N RDCs collected in phage and acid gel. A), $^1$H-$^{15}$N RDCs as function of residue number for phage (blue dots) and acid gel (red dots). B), Histograms of RDCs measured in phage and acid gel. C) Representative RDC curves derived from $^1$H-$^{15}$N RDCs of residue 437 measured in phage (red) and acid gel (green). The band represents an error range of 3 Hz.
3.3.3 Paramagnetic Relaxation Enhancements

To minimize possible interference of the protein from the cysteine mutation, polar and non-conserved residues were considered as targeted mutagenesis sites. The amino acid sequence of PhE was aligned against the Pfam family of NRPS to identify the non-conserved residues. The choices of cysteine sites were also based on the secondary structure predictions from TALOS+ using chemical shifts (Figure 3.4). Among the 15 single cysteine mutants, E11C, N44C, K314C are on the loop region, Q27C, D182C, Y208C, T434C, C453 are on the edge of secondary structure, and R53C, C106, K189C, E201C, E242C, K323C, R440C are within the secondary structure. All the mutant proteins were expressed well in per-deuterated M9 media and can be purified to a decent amount.

Labeling of both reagents caused significant changes of chemical shifts on residues close to the labeling sites and smaller changes on the rest of the residues when comparing the HSQC spectra to the unlabeled protein. The HSQC spectra of MTSSL-labeled and dMTSSL-labeled sample showed almost identical peak positions with MTSSL-labeled sample having disappeared peaks and peaks with lower intensity due to the PRE effect. Peak heights instead of peak volumes were used to determine the intensity ratios. The ratios and the peak linewidth from the diamagnetic sample were used to calculate the distances as described in method section 3.2.6. Severely overlapped peaks were completely omitted from the analysis. Shifts of peak positions due to the
spin-labeling could result in significant amount of miss-assignments. To avoid the problem and to identify locations that were affected by the spin-labeling, we plotted the ratio of intensity against the amino acid sequence. As an example shown in Figure 3.7, regions that were affected by spin-labeling were identified with significant lower intensity ratios. The patterns of the affected regions and the corresponding spin-labeling site also suggested that PRE effect is distance-dependent.

Figure 3.7: Bar graphs of intensity ratio vs primary sequence for spin label position Cys323 (orange star). Ratios are determined as described in section 3.2.6. The absence of a bar at any position indicates overlap or lack of assignment. The two red boxes indicate the regions that show PRE effects.
To evaluate the quality and reliability of the data, we compared the measured distances to the distances from the homology model of PheE. We modeled the MTSSL side chain into the homology model for each of the 15 spin labels. Figure 3.8 shows a graph of the modeled distances vs the experimental distances. The graph shows a good correlation between the modeled distances and the experimental distances; however, it also reveals a large error range (± 5 Å). There are two possible explanations for the large errors. First, the model structure might not represent the real structure in particular the orientations of secondary structure might be significantly different from the model structure. Second, we used a rigid rotamer to model the MTSSL side chain, which might not represent the real flexibility of the MTSSL.

![Graph of modeled distances vs experimental distances. The fitting line shows 90% of correlation.](image)

Figure 3.8: Model distances vs experimental distances. The fitting line shows 90% of correlation.
3.4 Discussion

The assigned backbone resonances have allowed us to accurately predict the secondary structures of GrsA-PheE using TALOS+ (111). The prediction agrees well with the crystal structure of the homologous epimerization domain of tyrocidine synthetase (Figure 3.4). Knowing the location of secondary structure within the sequence is an important step for our RDC-based structure determination approach.

The data we have collected have shown promising results in solving the solution structure of GrsA-PheE. Even though the $^{15}$N-$^1$H$_N$ RDCs measured in phage and acid stretched gel seem to be not independent, the $^{13}$C$^\alpha$-$^{13}$C$'$ RDCs in phage can provide further filtering for the incorrect orientations calculated from the $^{15}$N-$^1$H$_N$ RDCs. The PRE distance restraints have been used to lift the ambiguity of secondary structure orientations calculated from RDCs (115). Therefore, similar approach using the PRE data can be applied to resolve the orientational ambiguity from $^{15}$N-$^1$H$_N$ RDCs. More distance restraints including the amide-amide, amide-methyl and methyl-methyl NOEs from selective methyl labeling samples can provide further packing to determine the global fold of the structure.

Given the size of this protein, we are not able to obtain a complete assignment of side-chain resonances using traditional TOCSY-based methods. For large proteins, people have shown that specific methyl labeling on Leu, Ile and Val residues in conjunction with COSY-based experiment can assign the methyl side chain resonances
of these residues \(41\). Similar methyl labeling strategy can be used to obtain methyl-methyl and amide-methyl NOEs for distance restraints. Methyl side chain resonances can also be assigned by analyzing the methyl-amide NOEs with the neighboring backbone amide resonances. Furthermore, it has been proposed to use the RDC-defined global fold structure to assign both side-chain resonances and NOEs without TOCSY data \(116\). With the similar approach using NASCA (NOE assignment and Side-Chain Assignment) algorithm, we expect to obtain more side-chain resonances and NOE assignments using the global fold of PheE, which will further improve the resolution of the overall backbone structure.

It is believed the solution structures determined merely from methyl-methyl and amide-methyl NOEs are usually lower resolution because these structure determinations use ideal geometry of secondary structure as the starting model and the sparse NOE restraints are usually not enough for a precise determination of secondary structure orientation. In our RDC-ANALYTIC framework (Figure 4.1), we use RDCs for direct determination for the orientation of secondary structures. These correctly determined secondary structures can serve as the initial model for the structure determination. Furthermore, it has been proposed to use RDCs for the determination of loop structures \(107\), which are usually missing in traditional approach. Combining these approaches, our RDC-ANALYTIC framework uses a better initial model computed from RDCs and
provides a complete solution for global fold determination of large proteins in high resolution.
4. Redox Regulation of SHPs by A Novel Backdoor-backdoor Cysteine Disulfide

This chapter has been adapted mainly from the following published manuscript which is joint work with Devina Willard and Johannes Rudolph.


4.1 Introduction

The tandem SH2 (Src homology 2 domain) containing PTPs (SHPs) are among the increasing number of PTPs for which ROS-mediated redox regulation has been implicated. SHP-1 and SHP-2 perform opposing tasks in signaling pathways. Despite their opposite roles in signaling there is considerable evidence that both SHP-1 and SHP-2 are regulated by reversible oxidation. For SHP-1, in T-cells treated with hydrogen peroxide, redox-mediated inhibition of SHP-1 leads to increased phosphorylation of downstream substrates such as the ERK and p38 but not JNK kinases (117). For SHP-2, stimulation of T-cell receptors by T-cell antigen leads to production of H2O2, reversible oxidation of SHP-2, increased phosphorylation of its down-stream substrates, and integrin activation (118). Signal transduction and induction of mitogenesis through the platelet-derived growth factor receptor has also been shown to depend on transient oxidation of SHP-2 (119). Most recently, reversible oxidation of SHP-2 has been found subsequent to stimulation of cells with epidermal growth factor, playing a critical role in fibroblast proliferation and subsequent cardiac fibrosis (120).
Both SHP-1 and SHP-2 share the characteristic active-site motif (HCX5R) and conserved catalytic cysteine of all PTPs. In addition to sharing a similar protein fold with PTP1b, both SHP-1 and SHP-2 have the same five variable residues in the active site loop as PTP1b (SAGIG). Titration of the active site of SHP-1 using aromatic disulfides suggests that the catalytic cysteine in SHP-1 has a low $pK_a$ compared to other cysteines in the protein (121). In addition, S-nitrosylation of the catalytic cysteine has been reported to inhibit the phosphatase activity of SHP-1 in vivo (78). Thus, as for other PTPs, the catalytic cysteines in the SHPs appear susceptible to oxidation by ROS. In comparison to other PTPs, SHPs appear to have several possible mechanisms for protecting the initially oxidize sulfenic acid from irreversible oxidation. Based on the similarity in the active site pocket to PTP1b and PTPα, sulfenylamide formation is also possible. Alternatively, based on the crystal structures of SHP-1 (61) and SHP-2 (60), formation of intramolecular disulfides with one of two highly conserved cysteines seems possible (Figure 4.1). These two potential backdoor cysteines are located within 9 to 12 Å from the catalytic cysteine, within the range observed for PTPs and DSPs known to use a disulfide for protecting against irreversible oxidation. Towards a better understanding of the molecular mechanism for the oxidative regulation of the SHPs, we investigate herein their oxidation by hydrogen peroxide and reactivation by different reductants (DTT, GSH, and TR/TRR), with particular emphasis on the role of the two potential backdoor cysteines. Our results demonstrate transient formation of a sulfenic acid at the
active site followed by formation of a novel inactive and reversibly oxidized form wherein the catalytic cysteine is re-reduced and the two backdoor cysteines form an intramolecular disulfide.

Figure 4.1: Backdoor cysteines of the SHPs. Green: SHP-1, Purple: SHP-2. Distance (γS-γS): SHP-1: C363-C455 = 9.43 Å, C329-C455 = 12.35 Å, C329-C363 = 5.32 Å; SHP-2: C367-C459 = 9.66 Å, C333-C459 = 12.40 Å, C333-C367 = 5.07 Å. Structure alignment was done by using the CE website (http://cl.sdsc.edu) and graphed using pymol.
4.2 Materials & Methods

4.2.1 Materials

Hydrogen peroxide (H₂O₂) purchased from EM Science and dilutions to 1.0 M were prepared daily. Thioredoxin/thioredoxin reductase (TR/TRR) from *Escherichia coli* was a generous gift from JoAnne Stubbe (Massachusetts Institute of Technology, Cambridge, MA). NADPH, 3-o-methylfluorescein phosphate (mFP), dithiothreitol (DTT), iodoacetic acid (IAA), glutathione, and catalase were obtained from Sigma. Sequencing-grade trypsin was purchased from Promega and was dissolved in 50 mM acetic acid before use. Pfu DNA polymerase was obtained from Stratagene. Restriction enzymes were obtained from New England Biolabs. Cloning vectors pET28a and pET30 were obtained from Novagen.

4.2.2 Cloning, Expression, and Purification of Full-length SHP-1 and SHP-2.

The open reading frames of full-length SHP-1 and SHP-2 were amplified from a cDNA library derived from human brain by PCR (Cycle 1: 95°C, 3 min; 55°C, 1.5 min, 72°C, 2.75 min; Cycles 2 – 41: 95°C, 1 min; 53°C, 1.5 min; 72°C, 2.75 min; Final Cycle: 72°C, 8 min) using the primers TATACATATGCTCCCGTGCGGTGTTTCACCG and TATAGAATTCTCCTTTGAGGAACCCCTTGC for SHP-1 and ATATACATATGACATCGCGGAGATGGTTTCACCC and ATATAGAATTCTCTACTGAAACTTTTCTGCTGTTGC for SHP-2. After digestion of the PCR products with NdeI and EcoRI, SHP-1 and SHP-2 were cloned into the pET30a
expression vector and transformed into *Escherichia coli* BL21 (DE3) cells for expression. The proteins were expressed by induction of mid-log cells with 0.2 mM IPTG overnight at 25 °C. In a typical preparation, 30 g of frozen cell pellets were thawed in 60 mL of buffer A (20 mM MES, pH 5.5, 1 mM EDTA, and 1 mM DTT) supplemented with protease inhibitor cocktail from Boehringer Mannheim. The cells were lysed by sonication and cell debris was removed by centrifugation at 20000xg for 30 min. The resulting supernatant was bound to SP-Sepharose (7 mL) equilibrated in buffer A. The column was washed extensively with buffer A and the proteins were eluted with a linear gradient of NaCl (0 – 0.5 M) in buffer A. Fractions containing PTP activity were combined, concentrated with an Amicon YM10 concentrator, and exchanged into buffer B (50 mM Tris, pH 8.5, 1 mM EDTA, and 1 mM DTT) by multiple dilutions. Proteins were then bound to Q-Sepharose (5 mL) equilibrated in buffer B. SHPs were eluted with a linear gradient of NaCl (0 to 0.5 M) in buffer B and further purified by Superdex-200 gel filtration chromatography in TNED buffer (50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, and 1 mM DTT). The purified proteins (>90% pure by SDS-PAGE) were concentrated to 10 mg/mL using an Amicon Ultra-15 concentrator with addition of glycerol (10% final) and frozen in liquid N₂ for storage at -80 °C.

### 4.2.3 Cloning, Expression, and Purification of Catalytic Domains of SHP-1 and SHP-2

The His-tagged catalytic domains of SHP-1 (∆SHP-1, residues 247-514) and SHP-2 (∆SHP-2, residues 268-525) were PCR amplified (Cycle 1: 95 °C, 3 min; 55 °C, 1.5 min,
72 °C, 2 min; Cycles 2 – 31: 95 °C, 1 min; 55 °C, 1.5 min; 72 °C, 2 min; Final Cycle: 72 °C, 5 min) from the full-length clones using the primers

CCCATATGGGCTTCTGGGAGGAGTTTGAGAG and

ACAAGCTTCACTTCTGCGACTGCAGGACCTC for ∆SHP-1 and

GACATATGGGTCAAAGGCAAGAAAACAAAAAC and

TCAAGCTTCATAGTGTTCATATAATGCTG for ∆SHP-2. After digestion of the PCR products with NdeI and HindIII, ∆SHP-1 and ∆SHP-2 were cloned into the pET28a expression vector and transformed into Escherichia coli BL21 (DE3) cells for expression.

Following induction as for the full-length proteins, the cleared cell lysates were incubated with Ni-NTA agarose (5 mL) in buffer C (50 mM Tris, pH 8.0, 250 mM NaCl). The agarose was washed extensively with buffer C containing 20 mM imidazole. The His-tagged proteins were eluted with 500 mM imidazole and further purified by Superdex-75 gel filtration chromatography in TNED buffer. The purified proteins (>90% pure by SDS-PAGE) were concentrated to >20 mg/mL using an Amicon Ultra-15 concentrator with addition of glycerol (10% final) and frozen in liquid N2 for storage at -80 °C. Various cysteine mutants were generated by Quick-change mutagenesis (Stratagene) according to the manufacturer’s instructions using the following primers for ∆SHP-1: C363S, GGCCGAACAAATCCGTC CCATACTGG and

CCAGTATGGGACGGATTTGTCCGGCC; C329S,

ATCGCCAGCCAGGGCTCTCTGGAGGCCACGGTC and
GACCGTGGCCTCCAGAGGCGCCTGGCTGGCGAT; C455S,  
CCATCATCGTGCACTCCAGCGCCGGCATC and  
GATGCCGGCGCTGGAGTGCACGATGATGG; and for ∆SHP-2: C333S,  
ATTGCCACACAAGGCTCCCTGCAAAACACGGTG and  
CACCGTGTTTTGCAGGGAGCCTTGTGTGGCAAT; C367S,  
AGAGGAAAGAGTAAATCTGTCAAATACTGGCCT and  
AGGCCAGTATTGACAGATTACTCTTCTCTTCT. The C363S/C329S double mutation  
of ∆SHP-1 was generated from the single mutant ∆SHP-1 (C363S) with use of the C329S  
primers. The mutant proteins were purified as described for the wild-type catalytic  
domains.

4.2.4 Measurement of the Inactivation Rate by H₂O₂

The rate of inactivation of SHPs was calculated using a fixed time-point quench  
protocol. SHPs (100 – 200 µM) were incubated at 25 °C with varying concentrations of  
H₂O₂ (1 – 5 mM) in buffer H (50 mM HEPES, pH 7.5, 100 mM NaCl, 1mM EDTA) for the  
full-length enzymes and in 3C buffer (50 mM Tris, 50 mM bis-Tris, 100 mM sodium  
acetate, pH 7.5) for the catalytic domains. At varying time points (30 s to 15 min),  
 aliquots were taken and diluted (200-fold) into 3C buffer containing 10 units of catalase.  
Phosphatase activities were measured immediately using 3-o-methylfluorescein  
phosphate (mFP; 250 µM) in a continuous assay that monitors product formation at 477  
nm (ε = 27,200 M⁻¹cm⁻¹). The percent remaining activities were determined by
comparison with mock-treated samples. The observed rate constant of inactivation ($k_{\text{obs}}$) was determined by fitting with equation 4.1.

$$\text{percent \_ remaining \_ activity} = 100e^{-k_{\text{obs}} t} + \text{remaining \_ activity}$$ \hspace{1cm} (4.1)

The second-order rate constant $k_1$ (Scheme 4.1) was obtained by fitting the linear dependence of the rate of inactivation ($k_{\text{obs}}$) on the concentration of H$_2$O$_2$.

Scheme 4.1 Reaction Pathway for Oxidation/Reduction of the SHPs. The kinetic and mechanistic reaction pathway for oxidation and rereduction of the SHPs is outlined using SHP-1 as an example. Initial oxidation of the active site Cys455 by hydrogen peroxide ($k_1$[H$_2$O$_2$]) yields the sulfenic acid. Further oxidation ($k_2$[H$_2$O$_2$]) would yield higher oxidation species such as the sulfonic acid. Instead, rapid intramolecular disulfide ($k_3 > k_2$[H$_2$O$_2$]) formation with one of two backdoor cysteines (Cys329 or Cys363) followed by disulfide exchange yields the stable backdoor-backdoor disulfide. Re-reduction of either the sulfenic acid or any of the disulfides yields active enzyme ($k_4$[reductant]).

**4.2.5 Measurement of Reactivation Rate by DTT, GSH and TR/TRR**

The SHPs (100 – 200 μM) were initially inactivated by incubation with H$_2$O$_2$ (0.5 – 2 mM) for 10 to 15 min. Aliquots were then diluted 2-fold into buffer H (for full-length SHPs) or 3C buffer (for catalytic domains) containing 10 units of catalase and varying
amounts of DTT (10 – 200 mM), GSH (50 – 200 mM), or TR/TRR (0.2 – 5 molar equivalents of enzyme). Following incubation under reducing conditions for varying amounts of time (20 s – 10 min), aliquots were diluted 100-fold into 3C buffer and phosphatase activity was measured using mFP (250 µM). When using TR/TRR as the reductant, NADPH (16 – 400 uM) was also included and the ratio of TR to TRR was held fixed at 200 to 1. Control experiments verified that catalase was not required for reactivation using DTT or GSH because even the lowest concentration of reductant was sufficient for rapidly inactivating all H₂O₂. The observed rate constant of reactivation (k_{react}) was determined by fitting with Equation 4.2.

\[
\text{percent \ recovered \ activity} = 100(1 - e^{-k_{obs}t}) + \text{remaining \ activity}
\] (4.2).

The second-order rate constant k₄ (Scheme 4.1) was obtained by fitting the linear dependence of the rate of reactivation (k_{react}) on the concentration of reductants.

### 4.2.6 MALDI-MS Spectrometry

Samples for matrix-assisted laser desorption mass spectrometry (MALDI-MS) were prepared by incubating SHPs (~50 µg) with or without H₂O₂ followed by 10-fold dilution into 50 mM ammonium bicarbonate buffer at pH 7.8. Excess H₂O₂ was then rapidly removed by a G-25 spin column. To prevent disulfide exchange during the subsequent work-up, remaining free cysteines were blocked by reaction for 30 min with high concentrations of iodoacetic acid (IAA; 50 mM). Remaining IAA was removed by G-25 spin column. The SHPs were then digested with 0.5 µg of sequencing grade trypsin
by incubation overnight at 37 °C. The digested peptide mixture was desalted with a C18-
ZipTip (Millipore) and analyzed by MALDI-TOF using a Voyager DE-Pro.

4.3 Results

4.3.1 Inactivation of Full-length and the Catalytic Domain of SHPs by H2O2

In vivo studies have demonstrated the inactivation of SHP phosphatase activity
in response to external H2O2 or growth factor stimulated production of H2O2.
Additionally, preliminary studies in vitro have indicated that oxidation of SHP-1 is
readily reversible (122). However, no evidence has been found that this inactivation is a
result of direct oxidation of the active site cysteine in the SHPs. We began this study by
determining the susceptibility of SHP-1 and SHP-2 to oxidative inactivation in vitro. In
our studies we used both full-length SHPs and their catalytic domain constructs (∆SHP-
1 and ∆SHP-2). The catalytic domains reflect activated SHPs wherein the SH2 domains
are disengaged from the active sites, instead binding to specific pTyr motifs in receptor
tyrosine kinases. By measuring remaining phosphatase activity as a function of time and
concentration of H2O2, we determined the second-order rate constant for oxidation of the
SHPs (Figure 4.2, Table 4.1). SHP-1 and SHP-2 showed similar rate constants for
oxidation by H2O2 (2.0 and 2.4 M⁻¹ s⁻¹, respectively), as expected from their similar active
sites. In comparison, the rate constants for inactivation of ∆SHP-1 and ∆SHP-2 are ~4-
fold higher than for the full-length proteins (9.4 and 8.8 M⁻¹ s⁻¹, respectively). Slower
oxidation for the full-length SHPs is consistent with their reported lower phosphatase
activity, wherein the N-terminal SH2 (N-SH2) domain interacts with and alters the conformation of the catalytic pocket (60, 123). The more rapid oxidative inactivation for the catalytic domains of the SHPs suggests that oxidation targets the catalytic cysteine. The inactivation rates for the catalytic domains of the SHPs are comparable to other phosphatases implicated in regulation by direct oxidation including PTP1b (9.1 M⁻¹ s⁻¹, (124)), VHR (17.9 M⁻¹ s⁻¹, (124)), mitogen-activated protein kinase phosphatase (MKP3; 9.6 M⁻¹ s⁻¹, (125)), and Cdc25B (164 M⁻¹ s⁻¹, (83)).
Figure 4.2: Time dependence of inactivation of SHPs by H$_2$O$_2$. (A) Full-length SHP-1, the following concentrations of H$_2$O$_2$ were used: 2 mM (■), 3 mM (●), 4 mM (▲), and 5 mM (▲). In the inset, the concentration dependence of the rate of inactivation was used to derive a second-order rate constant, $k_1$ of 2.0 M$^{-1}$s$^{-1}$. (B) Full-length SHP-2, the following concentration of H$_2$O$_2$ were used: 2 mM (●), 3 mM (▲), 4 mM (■), and 5 mM (●). The inset gave a second-order rate constant, $k_1$ of 2.4 M$^{-1}$s$^{-1}$. (C) ∆SHP-1, the following concentrations of H$_2$O$_2$ were used: 0.5 mM (■), 1.0 mM (▲), 1.5 mM (●), and 2.0 mM (●). The inset gave a second-order rate constant, $k_1$ of 9.4 M$^{-1}$s$^{-1}$. (D) ∆SHP-2, the following concentration of H$_2$O$_2$ were used: 1.0 mM (■), 1.5 mM (▲), 2.0 mM (●), and 2.5 mM (●). The inset gave a second-order rate constant, $k_1$ of 8.8 M$^{-1}$s$^{-1}$. 
4.3.2 Reactivation of the SHPs by DTT, GSH and TR/TRR

We next characterized the re-reduction and concomitant reactivation of the SHPs by different reductants including DTT, glutathione (GSH), and thioredoxin/thioredoxin reductase (TR/TRR). Each SHP was first inactivated under appropriate specific conditions to yield less than 5% remaining activity and then quenched into buffer containing varying amounts of catalase and reducing agent. After further incubation to allow re-reduction of the catalytic cysteine, the recovery of activity was measured as a function of time and concentration of reductant. For DTT and GSH the rate of reactivation varied linearly as a function of reductant (Figure 4.3, Table 4.1). For SHP1, SHP-2, and ΔSHP-2, more than 80% activity was recovered with both DTT and GSH, while ~50% was recovered for ΔSHP-1. Only ΔSHP-2 could be reactivated with TR/TRR (36.5 ± 2.1 M^{-1} s^{-1}), suggesting that the SH2 domains not only prevent substrate binding but also inhibit the interaction of the catalytic domain with thioredoxin as needed for reactivation. The inability to reduce ΔSHP-1 with TR/TRR hints at a potential physiological difference in the regulation of SHP-1 and SHP-2.
Figure 4.3: Time dependence of reactivation of SHPs by DTT. (A) SHP-1, the following concentrations of DTT were used: 5 mM (●), 10 mM (▲), 15 mM (■), and 20 mM (○). In the inset, the concentration dependence of the rate of reactivation was used to derive a second-order rate constant, \( k_4 \) of 0.64 M\(^{-1}\)s\(^{-1}\). (B) SHP-2, the following concentration of H\(_2\)O\(_2\) were used: 5 mM (●), 10 mM (■), 15 mM (▲), and 20 mM (○). The inset gave a second-order rate constant, \( k_4 \) of 0.46 M\(^{-1}\)s\(^{-1}\). (C) ΔSHP-1, the following concentrations of DTT were used: 10 mM (●), 15 mM (▲), 20 mM (△), and 25 mM (●). The inset gave a second-order rate constant, \( k_4 \) of 0.33 M\(^{-1}\)s\(^{-1}\). (D) ΔSHP-2, the following concentration of H\(_2\)O\(_2\) were used: 2 mM (■), 3 mM (○), 4 mM (▲), and 5mM (★). The inset gave a second-order rate constant, \( k_4 \) of 0.84 M\(^{-1}\)s\(^{-1}\).
### Table 4.1: Rate of Oxidation and Reactivation for SHP-1 and SHP-2

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_1$ (H$_2$O$_2$) (M$^{-1}$s$^{-1}$)</th>
<th>$k_4$ (DTT) (M$^{-1}$s$^{-1}$)</th>
<th>$k_4$ (GSH) (M$^{-1}$s$^{-1}$)</th>
<th>$k_4$ (TR/TRR) (M$^{-1}$s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHP-1</td>
<td>2.0 ± 0.1</td>
<td>0.64 ± 0.08</td>
<td>0.026 ± 0.003</td>
<td>nd$^a$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(87% ± 8%$^b$)</td>
<td>(97% ± 2%)</td>
<td></td>
</tr>
<tr>
<td>SHP-2</td>
<td>2.4 ± 0.3</td>
<td>0.46 ± 0.04</td>
<td>0.16 ± 0.04</td>
<td>nd$^a$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(88% ± 3%)</td>
<td>(96% ± 3%)</td>
<td></td>
</tr>
<tr>
<td>ΔSHP-1</td>
<td>9.4 ± 0.8</td>
<td>0.33 ± 0.01</td>
<td>0.077 ± 0.014</td>
<td>nd$^a$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(46% ± 3%)</td>
<td>(65% ± 1%)</td>
<td></td>
</tr>
<tr>
<td>ΔSHP-2</td>
<td>8.8 ± 0.3</td>
<td>0.84 ± 0.03</td>
<td>0.07 ± 0.01</td>
<td>36.5 ± 2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(88% ± 3%)</td>
<td>(100% ± 10%)</td>
<td></td>
</tr>
<tr>
<td>ΔSHP-1 (C363S)</td>
<td>6.5 ± 0.2</td>
<td>0.77 ± 0.05</td>
<td>0.092 ± 0.013</td>
<td>nd$^a$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(48% ± 4%)</td>
<td>(53% ± 5%)</td>
<td></td>
</tr>
<tr>
<td>ΔSHP-1 (C329S)</td>
<td>9.6 ± 1.4</td>
<td>0.16 ± 0.03</td>
<td>0.053 ± 0.008</td>
<td>nd$^a$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(86% ± 8%)</td>
<td>(86% ± 3%)</td>
<td></td>
</tr>
<tr>
<td>ΔSHP-1 (C363S/C329S)</td>
<td>1.1 ± 0.1</td>
<td>nd$^a$</td>
<td>nd$^a$</td>
<td>nd$^a$</td>
</tr>
</tbody>
</table>

$^a$ Not detectable.

$^b$ Percent of recovered activity was obtained while the maximum reaction rate was reached after incubation with reductants.

#### 4.3.3 Identification of the Backdoor Cysteine Residues in the Catalytic Domain of the SHPs

Recovery of high levels of activity following oxidation and subsequent reduction for the SHPs (Table 4.1) and the known reactivity of sulfenic acids (126) implied the existence of a protective mechanism to prevent irreversible oxidation of the catalytic cysteines (Scheme 4.1). Given the presence of two conserved potential backdoor cysteines, we focused on the possible formation of intramolecular disulfide bonds using MALDI-MS. To reflect the active form of the enzyme and simplify the spectra, we chose to study the catalytic domain alone. To prevent disulfide exchange during the analysis,
all samples were treated with 50 mM iodoacetic acid prior to workup for MALDI-MS, thus trapping free cysteines in the carboxymethylated form but not modifying disulfides or oxidized cysteines. As expected for the non-oxidized forms of ΔSHP-1 and ΔSHP-2, we identified the active site peptides containing the carboxymethylated catalytic cysteines: QESLPHAGPIIVHC*SAGIGR in ΔSHP-1 at 2101 Da and QESIMDAGPVVVHC*SAGIGR in ΔSHP-2 at 2086 Da (Figure 4.4). Additionally, we observed the peptides containing the carboxymethylated backdoor cysteine residues (Figure 4.4). Therefore, no intramolecular disulfides exist in the reduced catalytic domains of either phosphatase. Following oxidation to a re-reducible form, we expected to find the catalytic cysteines as sulfinic acids or in a disulfide bonds with one of the potential backdoor cysteines. We were thus surprised to identify the same tryptic peptides containing the carboxymethylated catalytic cysteines in the oxidized samples of both ΔSHP-1 and ΔSHP-2 (Figure 4.4). The MALDI spectra from the oxidized ΔSHPs did contain a novel fragment not seen in the reduced samples, which corresponded to a disulfide linkage between the two potential backdoor cysteines (4515 Da for ΔSHP-1, 2465 for ΔSHP-2; Figure 4.4). Appearance of these backdoor-backdoor disulfides was accompanied by the significant reduction in the intensity of the individual peaks representing the carboxymethylated backdoor cysteines for ΔSHP-1 (Figure 4.4A).
Figure 4.4: MALDI-MS of reduced and oxidized ΔSHP-1 and ΔSHP-2. Wild-type ΔSHP-1, the catalytic mutant C455S of ΔSHP-1, and ΔSHP-2 were untreated (reduced) or treated (oxidized) with H$_2$O$_2$ followed by treatment with IAA. Peptide masses containing the catalytic cysteine and the backdoor cysteines are labeled in bold italic.
type. The schematic insets show the deduced oxidation states of the three cysteines near the active site. (A) The reduced sample from ΔSHP-1 shows the presence of the peptides containing carboxymethylated catalytic cysteine (Cys455, 2101 Da) and the two backdoor cysteines (Cys363, 1523 Da; Cys329, 3108 Da) with a mass increase of 58 Da compared to the unmodified peptides. In the oxidized sample, the two peptides containing the carboxymethylated backdoor cysteines have disappeared, and a larger peptide consistent with formation of a disulfide bond between the two backdoor cysteines is seen instead (4515 Da). The blown-up region shows the active site peptide with the catalytic cysteine in the sulfenic acid (2059 Da) and the sulfinic acid (2076 Da) forms. (B) The reduced sample from ΔSHP-2 shows the carboxymethylated peptide containing the catalytic cysteine (Cys459, 2086 Da) and one of the backdoor cysteines (Cys367, 2176 Da). The mass of the peptide containing the second backdoor cysteine (Cys367) is too small to detect under these MALDI-MS conditions (407 Da). As for ΔSHP-2, the oxidized sample shows the peptide representing the disulfide between the two backdoor cysteines (2465 Da). (C) The reduced sample from the catalytic mutant C455S of ΔSHP-1 shows the presence of the peptides containing the mutated catalytic residue (Ser455, 2027 Da) and the two backdoor cysteines (Cys363, 1523 Da; Cys329, 3108 Da). In contrast to the wild-type protein, the oxidized sample of the mutant still shows the two peptides containing the carboxymethylated backdoor cysteines (compare with panel A) and does not contain the peptide (4515 Da) representing a disulfide bond between the two backdoor cysteines (note arrow at location).

To ensure that disulfide formation between the two backdoor cysteines did not occur directly but instead depended on prior oxidation of the active site cysteine, we next generated the active site mutant C455S of ΔSHP-1. As expected, the mutant protein had no detectable phosphatase activity with the substrate mFP. Using MALDI-MS, C455S yielded the active site peptide with a serine mutation (2028 Da) and the same two peptides containing the carboxymethylated backdoor cysteines (1523 and 3108 Da, Figure 4.4C), even following oxidation by H₂O₂. The lack of an intramolecular disulfide between the two backdoor cysteines in the absence of the catalytic cysteine indicates that
oxidation of the catalytic cysteine is a prerequisite to formation of the backdoor-backdoor disulfide. Consistent with this interpretation, the MALDI-MS spectrum of oxidized ΔSHP-1 shows molecular masses that correspond to the active site peptide with the sulfenic acid (∼2059) and sulfinic acid (∼2076 Da) modifications (Figure 4.4A, blown-up region). The detection of the sulfenic acid-containing peptide is consistent with the expected intermediate on the path to disulfide formation (Scheme 4.1). The detection of the sulfinic acid-containing peptide is consistent with the incomplete reversibility of oxidative inactivation for ΔSHP-1 (Table 4.1). The absence of a sulfenic acid intermediate or sulfinic acid byproduct following oxidation of ΔSHP-2 presumably reflects the greater efficiency of transfer of the oxidation state from the active site cysteine to the two backdoor cysteines (Scheme 4.1) and is consistent with the high recovery of activity upon subsequent reduction (Table 4.1).

4.3.4 Identification of Disulfide Intermediates by Mutagenesis of the Backdoor Cysteines

To explain the mechanism by which we obtain a reduced catalytic cysteine and a backdoor-backdoor cysteine disulfide as the stable end product of oxidation, we hypothesized that the catalytic cysteine, following initial oxidation to the sulfenic acid, first forms a transient disulfide with one or the other of the two backdoor cysteines (Scheme 4.1). The proximity (<6 Å) of the two backdoor cysteines to each other would then favor formation of the backdoor-backdoor disulfide (Figure 4.1). We used site-
directed mutagenesis and MALDI-MS to probe this mechanism, wherein mutation of one of the backdoor cysteines could yield a disulfide between the catalytic cysteine and the remaining backdoor cysteine, representing a disulfide intermediate during the process of oxidation of the wild-type protein. Alternatively, removing both backdoor cysteines could lead to irreversible inactivation wherein disulfide exchange is no longer possible and the catalytic cysteine becomes oxidized to the sulfinic or sulfonic acid. We first mutated the backdoor cysteines that are further from the catalytic cysteine, namely, Cys329 in ΔSHP-1 and Cys333 in ΔSHP-2. The C329S mutant showed similar activity as wild-type ΔSHP-1, while the C333S mutant had less than 10% activity as compared to wild-type ΔSHP-2. Because the reduced activity of Cys333 in ΔSHP-2 may reflect a significant perturbation in the overall structure of the protein or a change in the microenvironment of the active site, we focused on characterizing the redox reactions of the C329S mutant of ΔSHP-1.

Purified C329S was treated with the same oxidation and reduction conditions as the wild-type protein. The C329S mutant had similar rate constants for oxidation and reactivation as the wild-type protein, indicating that the C329S mutation does not affect the chemical environment of the catalytic cysteine (Table 4.1). To probe the mechanism by which C329S achieves a reversibly oxidized form, the mutant ΔSHP-1 was treated with H₂O₂ and IAA and analyzed by MALDI-MS as for the wild-type protein. As expected, the reduced protein showed peptides containing carboxymethylated catalytic
(2101 Da) and the remaining backdoor cysteines (1523 Da) (Figure 4.5A). The sample from the oxidized protein contained a novel peak (3507 Da) representing the peptide containing a disulfide bond between the catalytic (Cys455) and the remaining backdoor (Cys363) cysteines. These data demonstrate the ready formation of a reversible disulfide between the catalytic cysteine and the nearby backdoor cysteine, representing a possible intermediate in the oxidation process of the wild-type protein (Scheme 4.1).
Figure 4.5: MALDI-MS of reduced and oxidized mutants of ΔSHP-1. C329S, C363S, and C329S/C363S mutants of ΔSHP-1 were untreated (reduced) or treated with H$_2$O$_2$ (oxidized), followed by IAA treatment. Peptide masses containing the catalytic cysteine and the backdoor cysteines are labeled in bold italic type. The schematic inset shows the deduced oxidation states of the three cysteines near the active site. (A) The C329S mutant. The reduced sample shows the peptides containing the
carboxymethylated catalytic cysteine (Cys455, 2101 Da), a carboxymethylated backdoor cysteine (Cys363, 1523 Da), and the backdoor serine mutation (Ser329, 3034 Da). The oxidized sample shows the presence of the peptide containing the disulfide between Cys455 and Cys363 (3507 Da). (B) The C363S mutant. The reduced sample shows the peptides containing carboxymethylated catalytic cysteine (Cys455, 2101 Da), a carboxymethylated backdoor cysteine (Cys329, 3109 Da), and the backdoor serine mutation (Ser363, 1449 Da). The oxidized sample shows the presence of the peptide containing the disulfide between Cys455 and Cys329 (5092 Da) and peptide containing a sulfinic acid-modified catalytic cysteine (2076 Da). (C) The C329S/C363S double mutant. Only the region containing the peptide with catalytic cysteine is shown. The same amount of C329S/C363S was treated with buffer (0 min), 1 mM H₂O₂ for 5 min, or H₂O₂ for 10 min followed by IAA treatment. The fully reduced sample (0 min) shows the peptide containing the carboxymethylated catalytic cysteine (2102 Da). The sulfenic acid- (2058 Da) and the sulfinic acid-modified (2076 Da) cysteines are present in the sample treated with H₂O₂ for 5 min while the 2102 peak representing the unoxidized catalytic cysteine is still dominant. In the sample treated with H₂O₂ for 10 min the peptides containing sulfenic acid- and sulfinic acid-modified cysteine are the major forms present.

We next generated the other backdoor cysteine mutant of ΔSHP-1, namely, C363S. (The corresponding C367S mutant of ΔSHP-2 did not express well and could thus not be studied.) Purified C363S showed only a 2-fold reduction in activity as compared to wild-type ΔSHP-1. C363S treated with H₂O₂ as for the wild-type protein showed a similar rate constant for oxidative inactivation (Table 4.1). Although the nearest backdoor cysteine is absent, the C363S mutant could be re-reduced by DTT and GSH with comparable rate constants compared to the wild-type protein, albeit to a lower extent of recovered activity. Peptides identified by MALDI-MS from the oxidized C363S mutant demonstrate that the reversibility of oxidation in the C363S mutant is due to the formation of a disulfide bond between the catalytic (Cys455) and the more distant
backdoor (Cys329) cysteines (Figure 4.5B). The oxidized sample contained a peptide corresponding to this alternative disulfide (5092 Da), whereas the reduced sample showed peptides containing the carboxymethylated catalytic (2101 Da) and more distant backdoor (3109 Da) cysteines. Thus, either of the two backdoor cysteines in ΔSHP-1 can capture the sulfenic acid that arises upon oxidation of the catalytic cysteine (Scheme 4.1).

4.3.5 The Two Backdoor Cysteines are Necessary and Sufficient for the Protection of the Catalytic Cysteine

The catalytic domains of both SHP-1 and SHP-2 contain two additional cysteines besides the catalytic and backdoor cysteines. It is theoretically possible that either of these more remote cysteines could also react with the oxidized catalytic cysteine and subsequently exchange to form the more stable backdoor-backdoor cysteine disulfide that we have identified above. Precedence for the involvement of multiple cysteines in the protection of the catalytic cysteine comes from the mitogen-activated protein kinase phosphatase (MKP3) (125). To test this possibility, we investigated the properties of the double mutation, eliminating both backdoor cysteines in ΔSHP-1, namely, C329S/C363S. The double mutant protein had a 2-fold decreased activity as compared to the wild-type protein, consistent with the activity seen for the single mutant C363S. The double mutant had a rate constant for oxidization by H₂O₂ that was 9-fold decreased as compared to wild-type ΔSHP-1 (Table 4.1). Despite a slower oxidation process, treatment of oxidized C329S/C363S with the reductants DTT or GSH did not lead to reactivation, unlike for the wild-type or either single cysteine mutation of ΔSHP-1. In
fact, even the mildest possible treatment consistently led to irreversibly oxidized protein.

To determine by MALDI-MS the oxidation state of the active site cysteine following inactivation, C329S/C363S was incubated with 1 mM H$_2$O$_2$ for 5 min (partial inactivation) or for 10 min (complete inactivation) and analyzed by MALDI-MS. With increasing inactivation, there is an increase in the peptides representing the sulfenic acid- (2059 Da) and sulfinic acid- (2076 Da) modified catalytic cysteine with concomitant loss of the carboxymethylated catalytic cysteine (2102 Da) (Figure 4.5C). Additionally, no peptides representing disulfide formation with one of the more remote cysteines could be detected.

### 4.4 Discussion

SHP-1 and SHP-2 are involved in multiple intracellular pathways responding to diverse signals such as integrins, growth factors, cytokines, and hormones. Ligand binding to the various membrane receptors in these pathways not only triggers traditional signaling events inside cells (i.e., phosphorylation, G-protein activation) but also leads to increases in local concentration of ROS up to 1 mM, thought to be generated in part by the NADPH oxidase system. These concentrations of H$_2$O$_2$ are sufficient to oxidize common regulators of these receptors, namely, the PTPs. As we have shown here in vitro, the SHPs are highly susceptible to oxidative inactivation at their active site cysteines, consistent with the identification of the sulfenic acid form of both SHPs upon activation of CD8$^+$ T-cells (127) and with previous in vivo data.
suggesting that oxidation of SHPs plays a regulatory role in signaling (117-120).

Additionally, our data provide interesting insights into the kinetic and mechanistic aspects of oxidative SHP regulation.

Kinetically, we noted no significant differences in the rates of oxidation and re-reduction (by DTT or GSH) between SHP-1 and SHP-2, whether as full-length proteins or as catalytic domains (Table 4.1). These data are reminiscent of the intriguing similarity in sequence, structure, and activity of these two phosphatases, despite their opposing roles in signal transduction. Clearly, there are aspects of intracellular localization, substrate recognition, and regulation that are insufficiently understood to explain these opposing roles. As a hint toward possible differences between these two phosphatases in their activated form (lacking the inhibitory SH2 domains), we did note that ΔSHP-2 but not ΔSHP-1 could be reactivated by TR/TRR, one of the more likely physiological reductants. Also, although the kinetic rates for reactivation were similar for ΔSHP-1 and ΔSHP-2, the extent of reactivation for ΔSHP-2 was significantly greater than for ΔSHP-1 (Table 4.1). Whether these data suggest a more permanent inactivation of SHP-1 in response to ROS in comparison to SHP-2 or the involvement of other intracellular reductants remains to be investigated in more detail. An additional hint that the true physiological reductant remains to be identified for both SHPs comes from the measured rates of reduction. In comparison to other PTPs where rates of reactivation
have been quantitated, the re-reduction of the SHPs is extremely slow. For example, Cdc25 (83) and MPK3 (125) are reactivated 10-30-fold faster by TR/TRR than ΔSHP-2.

Mechanistically, we have identified a novel end product that protects the oxidation state of the SHPs in a readily reversible form. With the exception of the peroxiredoxins, it is well-known that the higher oxidation states of cysteine (sulfonic and sulfinic acids) are essentially irreversibly oxidized (128). Previously known mechanisms for preserving the singly oxidized form of cysteine include the chemically unusual sulfenylamide found in PTP1b and PTPα and the backdoor cysteines found in the low molecular weight and Cdc25 phosphatases, among others. When present in a phosphatase, these backdoor cysteines are highly conserved, suggesting their evolutionary importance to the function and/or regulation of these PTPs. The presence of two highly conserved potential backdoor cysteines in the SHPs raised the possibility that more than one cysteine could serve as a partner for the active site cysteine in the formation of a stable disulfide. Much to our surprise, while such disulfides between the active site cysteine and either of the backdoor cysteines exist as intermediates, the stably oxidized form consists of a reduced catalytic cysteine and a backdoor-backdoor disulfide (Scheme 4.1). This mechanism is different than the one found in MKP3, where all of the multiple cysteines are involved in disulfides directly with the active site cysteine and do not appear to form disulfides with each other (125). This novel oxidized form of the SHPs also emphasizes that identifying oxidized forms of PTPs in vivo must rely on
techniques other than trapping sulfenic acids or modified active site cysteines (glutathionylation, disulfides, or sulfenylamides). That is, dissecting redox regulation of PTPs inside cells may be more difficult and complex than previously thought.
5. Summary and Further Directions

5.1 Redesign and Kinetic Analysis of GrsA-PheA

Using our suite of structure-based protein design algorithms, we successfully redesigned GrsA-PheA for a set of non-cognate substrates. A switch of substrate specificity from Phe toward Leu was observed for several of the computationally-predicted mutants. Further redesigns for Arg, Glu, Lys, and Asp were also successful experimentally, and accomplished the task of creating novel substrate activity (virtually non-existent in WT GrsA-PheA), although the preferred substrate for all these mutants was still Phe. The incorporation of an explicit negative design procedure will be important for predicting active mutants that show the desired switch of substrate specificities. However, for in vitro or biotechnology applications, it would be possible to use the designed mutants for charged amino acid adenylation by controlling the input substrates to exclude Phe. More extensive investigation of the effect of bolstering mutations on the substrate specificity of the redesigned enzymes could be an important step toward a general purely-computational algorithm for predicting enzymes with high activity by identifying mutations anywhere in the protein, both proximal and distal to the ligand binding site.

5.1.1 Enzymatic Mechanism of GrsA-PheA

We have performed steady-state and pre-steady-state kinetics on the redesigned PheA mutants for further understanding the enzymatic mechanism of the adenylation
reaction. By analyzing free energy changes between WT and mutant PheA with substrates Phe and Leu, we found out the active site residues 278 and 301 interact to each other, providing extra coupling energy to facilitate substrate binding. The pre-steady-state analysis of the WT PheA binding to Phe revealed a two-step binding process. The finding suggests an existence of two phenylalanine-binding conformations of PheA (Figure 5.1). The first conformation is a substrate-selection conformation in presence of ATP and Mg²⁺. This conformation most likely represents the active enzyme-substrate complex. The second conformation is a product-releasing conformation as shown in the crystal structure of PheA complex with AMP and Phe (4). Interestingly, by using the crystal structure of PheA as input, $K^*$ was able to predict active site mutants that actually improve $K_d$ with changes on both association and dissociation constants for the non-cognate substrate Leu (Table 2.6). The improvement in $K_d$ for the double mutant (T278L/A301G) to bind Leu suggested that the PheA mutant adopts a different substrate-selection conformation in the substrate binding step. It is likely that the AMP- and ATP-bound conformations of PheA are structurally very similar and the selection of substrates lies in few key interactions of residues in the substrate binding pocket as
shown in our free energy analysis (Table 2.8).

Figure 5.1: Reaction pathway for the adenylation reaction of NRPS. E* and E** denote the two substrate-binding conformations as described in the text.

The adenylation step is the first half step for loading of amino acid substrate to the adenylation domain of NRPS. The second half step which also occurs in the adenylation domain is the attachment of aminoacyl group to the terminal thiol of the 4’-phosphopantetheine (Ppant) cofactor, which is covalently bound to the PCP domain. It is suggested that a conformational change of the A domain occurs after the first half adenylation step so that the A domain adopts a conformation to accommodate the Ppant arm as well as provides the right chemical environment for the formation of the thioester bond. To make our redesigned PheA applicable for making the final product, it is important to show that the mutant enzyme can catalyze the non-cognate substrate in the second step. Since the substrate-bound Ppant cofactor is covalently linked to the PCP domain, we propose a single-turnover experiment by using rapid quench apparatus to measure the fast product formation. 14C-labeled amino acid substrate will be used as a read-out to monitor the product formation as a function of time.
5.2 Interdomain Interaction Studies of GrsA-PheE

It has been shown that the activity of the epimerization domains of NRPSs requires specific types of aminoacyl- or peptidyl-bound carrier proteins (129). This finding suggests specific interactions exist between the E domains and the PCP domains. Also the short COM domains found in the C-terminus of E domains are known to mediate the communication between NRPS modules (36). Our NMR studies of the epimerization domain of GrsA have established the foundation for understanding the molecule recognitions in NRPS. Chemical shift perturbation (CSP) has been used extensively in revealing protein-protein or protein-ligand interactions given assigned backbone resonances. CSPs were used to reveal the binding interfaces on the PCP-TE didomain in E. coli enterobactin synthetase for the interactions to the broad-specificity PPTase, Sfp and its upstream condensation domain (34). We therefore propose to measure CSPs by titration of the GrsA-PCP/GrsA-PheA domains or the first condensation domain of GrsB to GrsA-PheE. We expect to identify regions on the PheE domain that show significant CSPs in response to different protein partners. Substrate specificity as well as the binding sites for the Ppant arm on GrsA-PheE can be revealed using the similar approach. The PheE domain has been shown to contribute to some degree of substrate specificity, while it can interconvert the L- and D-form of the phenylalanyl-PCP substrate (90). Finding the residues that play the roles in
distinguishing the substrate specificity and the stereoisomer could benefit the redesign of the enzyme.

5.3 Redox Regulation of SHPs

In chapter 4, we have provided the first in vitro kinetic and mechanistic analysis of redox regulation for the highly homologous phosphatase SHPs. Despite their opposing roles in the regulation of signaling pathways, we have found the kinetic properties of inactivation and reactivation of SHP-1 and SHP-2 to be highly similar. Most importantly, our discovery of a backdoor–backdoor disulfide in the presence of a reduced active site cysteine extends the number of different mechanisms by which an oxidized and inactivated PTP can be stabilized in a re-reducible form.

5.3.1 $pK_a$ of the Active Site Cysteines of SHPs and the Conformation of Reversible Oxidation State

A low $pK_a$ of the active site cysteine is a unique characteristic of the cysteine-dependent PTPs. Under physiological pH, the $pK_a$ value determines the strength of the catalytic cysteine thiolate which affects both the catalytic activity and the susceptibility to oxidation. Our finding that the active site cysteines of SHPs are susceptible to oxidation has further confirmed a lower $pK_a$ for the active site cysteines in SHPs. Nevertheless, the difference in oxidation by $H_2O_2$ for the full-length and the catalytic SHPs suggests a different chemical environment, hence a different $pK_a$ value for the active site cysteine. In fact, crystal structures of the full-length and the catalytic domain SHP-1 show different active site conformation (Figure 5.2). In full-length SHP-1, the
aspartic acid (D61) from the N-SH2 precludes the conserved WPD loop from entering into the active site while in the catalytic domain SHP-1, the WPD loop is close to the active site, positioning D421 near the catalytic cysteine. In addition, there are two conserved polar side chains (S456, T350 in SHP-1 and S460, T356 in SHP-2) that might participate in stabilizing the cysteine thiolate formation. It is therefore intriguing to know if the surrounding residues in active site of PTPs affect the pKₐ of the catalytic cysteines. A pH-dependent kinetics in conjunction with site-directed mutagenesis might address this question.

Figure 5.2: Active site of SHP-1. Green: Full-length, Cyan: Catalytic domain.
Finally, given that the susceptibility to oxidation and the phosphatase activity for PTPs arises from the greatly perturbed pKₐs (4-6.5) of the active site cysteines (130), the reduced active site cysteines in the stably oxidized form of the SHPs are most likely returned to a more normal value (pKₐ ~ 9) in the stably oxidized form of the SHPs. Thus, there exists the expected correlation between oxidation and inactivation, wherein the catalytically essential thiolate is most likely in the thiol form in the oxidized enzyme (Scheme 4.1). Using X-ray crystallography it may be possible to obtain insights into the local rearrangements within the active site region that allow for a protonated thiol in the oxidized enzyme, as demonstrated by “in crystal” oxidation for the Cdc25B phosphatase (81). As shown by the mutagenesis data, a single backdoor cysteine in SHP-1 is sufficient to trap the oxidation of the catalytic cysteine in a reversible state, just as for MKP3. Why the SHPs evolved a mechanism using this novel two-backdoor cysteine trap remains unclear. There may be some advantage for effecting the transfer of oxidizing equivalents to a separate reductant, as seen for example in the complex thiol-transfer reactions of arsenate reductase (131) and ribonucleotide reductase (132). Establishing the physiologically relevant reductant for the SHPs would allow further probing of this pathway.
Bibliography


through the screening of a combinatorial phosphotyrosyl peptide library, 
*Biochemistry* 39, 13251-13260.


requirement of reversible cysteine sulfenic acid formation for T cell activation and function, *Journal of immunology* 179, 6456-6467.


Biography

Cheng-Yu Chen

Date of Birth: Nov 26th, 1978,
Place of Birth: Taichung, Taiwan

Education:

2004~2013 Doctor of Philosophy at Duke University, Durham, NC
  Graduate program in Biochemistry
  Advisor: Bruce R. Donald, PhD

2000~2002 The Degree of Master of Science, Department of Life Science, National Tsing-Hua University, HsinChu, Taiwan
  Advisor: Wen-Ching Wang, PhD

1996~2000 Bachelor’s Degree in Science, Department of Life Science, National Tsing-Hua University, HsinChu, Taiwan

PUBLICATIONS:


