BROADENING THE APPLICATION OF ENZYME-CATALYZED SYNTHESIS OF SEMI-SYNTHETIC ANTIBIOTICS

A Thesis
Presented to
The Academic Faculty

by

Janna K. Blum

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemical & Biomolecular Engineering

Georgia Institute of Technology
May 2011
BROADENING THE APPLICATION OF ENZYME-CATALYZED SYNTHESIS OF SEMI-SYNTHETIC ANTIBIOTICS

Approved by:

Professor Andreas S. Bommarius, Advisor
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Professor Christopher W. Jones
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Professor Wendy L. Kelly
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Hang Lu
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Professor Mark R. Prausnitz
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Date Approved: March 17 2011
The best way to have a good idea is to have lots of ideas.

—Linus Pauling (1901-1994)
To Theresa P. Creel (Grandma)
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my friends and family for the endless support. I would like to thank my parents: my mom who is a great sounding board and would listen to me on both my good days and bad days and my dad who was always in my corner. I would like to acknowledge my brother, Nathan, for keeping me grounded in my industrial background and reminding me of what I had to offer the field of Chemical Engineering. To Richard, without your love and support I would have never stuck through this process, also thank you for the help with \LaTeX and MATLAB.

To my family, Jill Billy, Cameron Billy, Carson Billy, Toni Creel, Mark Blum, Nancy Blum, and Natalie Blum and all of my very dear friends, in particular Michelle Alger, the support system that you provided was integral to my success. To my friends within the Georgia Tech Student Government Association and Dr. Raymond Vito, thank you for getting me out of the lab and letting me contribute to Georgia Tech. I would like to thank all of the incoming Ph.D. Chemical Engineering class of 2005, I am honored to be associated with such wonderful people. Tori, Michelle, Krystal, Ven, Ed, Angel, Richard and Alison (honorary 2005 class member), thanks for your friendships. Lastly, I would like to recognize the fallen member of our class, Christopher Russell (1982-2008); you will be sorely missed.

Secondly, I would like to thank my colleagues whom I also consider my friends. Countless hours of discussion about research helped to turn a chemical engineer who did not know the difference between an intron, a plasmid (aka “circle DNA”), and a chromosome into a biochemical engineer. I appreciate Dr. Karen Polizzi, Dr. Javier Chaparro-Riggers, Dr. Bernard Loo and Dr. Tracey Thaler for introducing me to the Bommarius Lab and teaching me the basics of cloning and molecular biology. Furthermore, I am grateful for the many technical discussions on enzyme catalysis and protein engineering with Dr. Eduardo Vazquez, Dr. Barhaheh Azizi, Dr. Thomas Rogers, Dr. Mélanie Hall, Michael Abrahamson, Michael Rood, and Soohee Cho. I would like to thank Jennifer, Hally, Amanda, Hilda, Anna and Michael R. for teaching me the ins and outs of nuclear receptors and chemical complementation. Believe it or not, I learned quite a bit from those AEG meetings. I would like to thank the rest of our lab and the IBB 3D wing for making this the best environment to work in all of Georgia Tech; I am a better scientist because of all of you. To all of the IBB and
ES&T support staff, Allen Echols, Steve Woodard, James Goodard, Floyd Wood, Todd Clarkson and Josie Giles for making this the best place to conduct research and always welcoming me with a smile in the hallways and elevators.

Lastly, I would like to thank the researchers who contributed directly to my thesis work. To my undergraduate researchers, Pat Garrett, Evelina Ponizhaylo, Dan Ricketts, and Thanh Ha, your help was tremendous and it was a joy to work and get to know all of you. To Pat and Evelina thanks for starting out on this project with me. I know that sometimes it seemed hopeless when we had no activity and no bands on the DNA gel, but together we got through the rough patches and really made this enzyme robust. To Dan and Thanh, you are both great scientists and your help on the thermostability and substrate specificity projects was invaluable. I would like to recognize Andria Deaguro and Carolina Perez for the work on the one-pot synthesis of ampicillin. Andria, I truly appreciate that you were on board with the idea and took your time and your undergrad’s time to work on this project with me. I would like to acknowledge Prabuddha Bansal for his help with the MATLAB code that was used to evaluate the synthesis kinetics. I would like to thank Dr. Nick Hud and his lab for their help with the CD instrument, and Jonathan Rubin for running the DLS samples. I would like to acknowledge Dr. Donald Doyle for use of his lab equipment, helpful discussions on biochemistry and the NIH Grant. I would like to acknowledge my advisor, Andy Bommarius, for allowing me to explore my ideas and always being my biggest cheerleader. To my committee, Dr. Mark Prausnitz, Dr. Chris Jones, Dr. Hang Lu, and Dr. Wendy Kelly, thank you for your feedback on my research, as these conversations really broadened my horizons and made me approach my work differently. Lastly, Shimaduz HPLC, without you I would not have any data.

I gratefully acknowledge support from the National Institute of Health(Grant #5R01AI064817-02) and the NSF Graduate Research Fellowship. Additionally, I would like to acknowledge funding from the Georgia Tech Presidential Undergraduate Research Assistantship (PURA) for funding Dan Ricketts, Thanh Ha, and Carolina Perez.
# TABLE OF CONTENTS

DEDICATION ................................................................. iv

ACKNOWLEDGEMENTS .................................................. v

LIST OF TABLES ............................................................ xi

LIST OF FIGURES ......................................................... xii

LIST OF SYMBOLS OR ABBREVIATIONS ............................... xiv

SUMMARY ................................................................. xvii

I  INTRODUCTION .......................................................... 1

1.1 \( \beta \)-lactam Antibiotics ............................................. 1

1.2 Manufacture of Enzymatic Semi-Synthetic \( \beta \)-lactam Antibiotics .. 6

1.2.1 Chemical synthesis ............................................... 6

1.2.2 Biocatalytic synthesis .......................................... 6

1.2.3 Comparison of chemical and biocatalytic syntheses ......... 9

1.3 Biocatalysts used to synthesize \( \beta \)-Lactam antibiotics ......... 10

1.3.1 Penicillin G Acylase(PGA) ................................... 10

1.3.2 \( \alpha \)-Amino Ester Hydrolase(AEH) .......................... 14

1.3.3 Comparison between AEH and PGA ......................... 17

1.4 Protein Engineering .................................................. 18

1.4.1 Protein Engineering toward increased thermostability .... 19

1.4.2 Protein Engineering toward altered substrate specificity ... 23

1.4.3 Link between thermostability and altered substrate specificity 26

1.5 Aims ................................................................. 27

1.5.1 Aim 1 ............................................................. 27

1.5.2 Aim 2 ............................................................. 28

1.5.3 Aim 3 ............................................................. 28

1.5.4 Recommendations and Conclusions .......................... 29

1.5.5 Publication Information ......................................... 30
II ISOLATION, EXPRESSION, AND CHARACTERIZATION OF AEH FROM
XANTHOMONAS CAMPESTRIS PV. CAMPESTRIS, ATCC 33913 . 31

2.1 Introduction .......................................................... 31
2.2 Results and Discussion .............................................. 33
  2.2.1 Isolation and Expression ...................................... 33
  2.2.2 Activity Characterization .................................. 35
  2.2.3 Thermostability ................................................. 35
  2.2.4 Role of the Carboxylate Cluster ............................ 40
2.3 Conclusions ............................................................ 44
2.4 Materials and Methods ............................................. 45
  2.4.1 Materials .......................................................... 45
  2.4.2 Bacterial Strains and Plasmids ............................... 46
  2.4.3 DNA Sequencing ................................................ 46
  2.4.4 Cloning of gaa into expression host ......................... 46
  2.4.5 Site-directed mutagenesis ................................... 47
  2.4.6 Recombinant overexpression of the AEH .................... 47
  2.4.7 Isolation of recombinant gaa gene from E.coli ............. 47
  2.4.8 Enzyme assays and determination of kinetic constants .... 48
  2.4.9 High Performance Liquid Chromatography .................. 48
  2.4.10 Circular Dichromism (CD) ................................... 48
  2.4.11 Publication Information ....................................... 49

III SYNTHESIS OF AMPICILLIN USING AEH ....................... 50

3.1 Introduction .......................................................... 50
  3.1.1 Kinetics of the Amidation reaction using AEH ............ 50
  3.1.2 One-pot synthesis using AEH and PGA ...................... 51
3.2 Results and Discussion ............................................. 53
  3.2.1 Synthesis using AEH alone .................................... 53
  3.2.2 Ampicillin Synthesis using a Two-Enzyme Cascade System
        with both α-Amino Ester Hydrolase and Penicillin G Acylase 59
3.3 Conclusions ......................................................... 64
3.3.1 AEH Synthesis ................................................. 64
3.3.2 Cascade synthesis ............................................ 64
3.4 Materials and Methods ........................................... 65
3.4.1 Materials ...................................................... 65
3.4.2 AEH Synthesis Reactions from (R)-PGME and 6-APA 65
3.4.3 One-Pot, One Step Synthesis (1P1S) ....................... 65
3.4.4 One-Pot, Two-Step Synthesis (1P2S) ....................... 66
3.4.5 One-Pot, Two-Step, Two-Stage Synthesis (1P2S-2S) 66
3.4.6 High Pressure Liquid Chromatography (HPLC) Assay 66
3.4.7 Publication Information ...................................... 67

IV IMPROVED THERMOSTABILITY OF AEH USING A MODIFIED STRUCTURE-GUIDED CONSENSUS METHOD ........................................... 68
4.1 Introduction ....................................................... 68
4.2 Results and Discussion .......................................... 72
4.2.1 Round 1: Single Variants .................................... 72
4.2.2 Round 2: Multiple Variants .................................. 78
4.2.3 Round 3: NNK libraries ....................................... 78
4.2.4 Round 4: Introduction of a disulfide-bond:G357C 83
4.2.5 Total Turnover Number ....................................... 85
4.2.6 Thermodynamic evaluation of variants with circular dichroism 87
4.3 Conclusions ....................................................... 88
4.4 Materials and Methods .......................................... 89
4.4.1 Protein ....................................................... 89
4.4.2 Mutations ..................................................... 89
4.4.3 Expression .................................................... 89
4.4.4 T30 \textsuperscript{50} measurements ................................ 89
4.4.5 High-Performance Chromatography (HPLC) ............... 91
4.4.6 CD ........................................................... 91
# LIST OF TABLES

1.1 Antibiotic Synthesis Process Impurities ........................................ 10
1.2 Selected Kinetic Parameters for PGA from *E. coli* .......................... 12
1.3 Selected Kinetic Parameters for AEH ........................................... 16
2.1 AEH from *X. campestris pv. campestris* Kinetic constants ............... 35
2.2 Alignment of active site residues of *X. campestris pv. campestris* AEH, RhCoCE, and J1 Acylase ............................. 42
3.1 6APA:(R)-PGME ratios .................................................................. 54
3.2 1P1S Ampicillin Conversions ...................................................... 59
3.3 1P2S Ampicillin Conversions ...................................................... 63
4.1 Percent amino acid identity matrix of homologous proteins to AEH from *X. campestris pv. campestris* ............................. 73
4.2 MSGC Round 1mutations ........................................................... 74
4.3 Covariance Matrix ....................................................................... 78
4.4 Results from Rounds 2–4 .............................................................. 79
4.5 Total Turnover Number Calculation .............................................. 86
4.6 Mutational Additivity ................................................................. 87
5.1 AEH CAST library ...................................................................... 101
5.2 AEH site-directed mutations with Cocaine Esterase and J1 Acylase like active-sites .................................................. 104
E.1 Key Residues from Amino Ester Hydrolases ................................. 131
LIST OF FIGURES

1.1 $\beta$-lactam nuclei ........................................ 2
1.2 Amoxicillin .................................................. 2
1.3 Penicillin G Inhibitor Complex ............................ 3
1.4 $\beta$-lactamasae hydrolysis reaction ........................ 4
1.5 $\beta$-lactamase-resistant antibiotics ........................ 5
1.6 Thermodynamic vs. Kinetic Controlled Enzyme Catalyzed Synthesis 7
1.7 PGA crystal structure, 1gm9 ................................ 11
1.8 AEH Enzyme from A.turbidans ............................ 15
1.9 Consensus Approach ........................................ 19
1.10 Results from Structure-guided consensus of Glucose Dehydrogenase (GDH) ........................................ 21
1.11 B-FIT libraries LipaseA ..................................... 23
1.12 ProSAR ......................................................... 25
1.13 Thermostatrability vs. Evolvability ........................ 27
2.1 AEH Synthesis Reaction ................................. 32
2.2 AEH Protein Gel ............................................. 34
2.3 Michaelis-Menten Equation Fits .......................... 36
2.4 AEH pH Profiles ............................................ 37
2.5 AEH Temperature Profiles ................................. 38
2.6 Wild-type CD scan rate dependance ....................... 39
2.7 AEH Half-life Studys ....................................... 41
2.8 Protease Assay .............................................. 41
2.9 AEH Active Site ............................................. 44
3.1 iPGA and AEH Cascade Scheme 2 ......................... 52
3.2 AEH Synthesis Profiles ................................... 55
3.3 6-APA:(R)PGME ratio ..................................... 56
3.4 Determination of $\beta_0$ and $\gamma$ .......................... 57
3.5 R-PG versus AMP .................................................. 58
3.6 1P1S and 1P2S Reaction Profiles ............................... 60
3.7 1P1S and 1P2S Ampicillin Conversions ........................ 62
4.1 Modified structure guided consensus (MSGC) decision tree .... 71
4.2 T_{30}^{50} measurements ............................................. 74
4.3 MSCG variant results ............................................... 75
4.4 AEH Domain .......................................................... 77
4.5 PDAB Chromogenic Assay .......................................... 80
4.6 Small-scale T_{30}^{50} experiments .................................... 80
4.7 Screening results for Libraries K34NNK-TM and E143NNK-TM ... 82
4.8 K34L and K34C SDS-PAGE Gel ................................... 83
4.9 G357C Design ........................................................ 84
4.10 G357C-TM SDS-PAGE ............................................. 85
4.11 T_{30}^{50} profiles ..................................................... 86
4.12 Circular Dichroism evaluation of Triple Variant ................. 88
5.1 AEH Library design concept ....................................... 98
5.2 homologous Residues of AEH, Cocaine Esterase, and J1 Acylase ... 99
5.3 AEH CAST libraries ............................................... 100
5.4 CD Scans of homologous Active-Site Mutations ............... 105
5.5 Leucine Auxotroph Assay .......................................... 106
5.6 Chemical Complementation ....................................... 107
5.7 Enzyme Activated Growth ........................................ 108
G.1 E143 NNK Library results ....................................... 137
LIST OF SYMBOLS OR ABBREVIATIONS

1P1S    One-Pot-One-Step.
1P2S    One-Pot-Two-Step.
6-APA   6-aminopenicillanic acid.
7-ACA   7-aminocephalosporanic acid.
7-ADCA  7-desacetoxy cephalosporanic acid.
A       pre-exponential of Arrhenius equation.
AEH     α-Amino Ester Hydrolase.
α, β, γ Synthesis Kinetic Parameters.
α_CD    Fraction of Unfolded Protein.
Amp     Ampicillin.
ATCC    American Type Culture Collection.
bp      Base Pairs.
CAST    Combinatorial Active Site-saturation Testing.
CD      Circular Dichroism.
CEX     Cephalexin.
Cm      Chloramphenicol.
Cm_r    Chloramphenicol Resistant.
∆G      Gibbs free energy (KJ·mol⁻¹).
∆H      enthalpy change (KJ·mol⁻¹).
DLS     Dynamic Light Scattering.
DNA     Deoxyribonucleic acid.
<E>     SCHEMA disruption.
E_a     Activation Energy (KJ·mol⁻¹).
E_app   Enantiomeric Ratio.
E_d     Deactivation Energy (KJ·mol⁻¹).
e.e.\textsubscript{p} Enantiomeric Excess.

ePPCR Error Prone PCR.

His\textsubscript{6} 6X Histidine Tag.

HPLC High Performance Liquid Chromatography.

IA Incubated Activity.

iPGA Immobilized Penicillin G Acylase.

ISM Iterative Saturation Mutagenesis.

Kan Kanamycin.

k\textsubscript{cat} catalytic turnover number (s\textsuperscript{-1}).

k\textsubscript{cat,obs} observed catalytic constant (s\textsuperscript{-1}).

k\textsubscript{D,obs} Deactivation rate constant (s\textsuperscript{-1}).

K\textsubscript{f} Inhibition Constant (mM).

K\textsubscript{M} Michaelis Constant (mM).

<m> SCHEMA Average Number of Mutations.

MSGC Modified structure guided consensus.

NIA Non-Incubated activity.

NNK Degenerate codon, where N=(A,T,C,G), K=(G,T).

\nu Catalytic Rate.

PCR Polymerase Chain Reaction.

PDA prominence diode array detector.

PDAB Paradimethyl-amino benzaldehyde.

pdb Protein Data Bank.

penG Penicillin.

PG Phenylglycine.

PGA Penicillin G Acylase.

PGME Phenylglycine Methyl Ester.

P\textsubscript{h} Hydrolysis Product, phenylglycine.
$[P]_{max}$  Maximum Product Concentration (mM).
$P_s$  Synthesis Product.
RA  Residual or Relative Activity.
RhCocE  \textit{Rhodacoccus} sp. Cocaine Esterase.
RMS  Root Mean Square.
SA  Specific Activity (U·mg$^{-1}$).
SDS  Sodium Dodecylsulfate.
t  time.
$T$  Temperature ($^\circ$C).
$T_{30}^{50}$  Temperature that has 50\% of residual activity after 30 min of incubation ($^\circ$C).
$\tau_{1/2}$  half-life (min).
TM  AEH Triple Variant (N186D/A275P/V622I).
$T_m$  Melting Temperature ($^\circ$C).
$T_{opt}$  Temperature of Maximum Activity ($^\circ$C).
TTN  Total Turnover Number.
U  Activity Units ($\mu$mol min$^{-1}$).
WT  Wild-Type protein.
SUMMARY

The development of enzymatic pathways to β-lactamase-resistant antibiotics allows for cheaper, faster manufacturing of second-line antibiotics used to treat antibiotic-resistant bacterial pathogens. Enzymes have been successfully applied toward the synthesis of first-line, semi-synthetic antibiotics, such as ampicillin and amoxicillin; however, they have not been used in the production of more complex β-lactamase-resistant antibiotics, such as oxacillin. Because of increasing resistance by bacteria towards the currently used antibiotics, it is crucial to develop enzymes to improve manufacturing of β-lactamase-resistant antibiotics.

Penicillin G Acylase (PGA) and α-amino ester hydrolase (AEH) are promising template enzymes for protein engineering of novel biocatalysis toward the synthesis of β-lactamase-resistant antibiotics. PGA is used commercially in both the production of 6-amino-penicillanic acid (6-APA) from penicillin G and the semi-synthesis of antibiotics such as ampicillin and amoxicillin, which are β-lactamase-sensitive. AEH catalyzes the synthesis and hydrolysis of α-amino acylases, including ampicillin. AEHs unique substrate specificity and alternative enzyme fold of homotetrameric α/β-hydrolase versus the heterodimeric penicillin G acylase make it an interesting candidate for protein engineering and improved/altered activity over penicillin G acylase; however, some challenges, such as the enzyme poor thermostability, have been identified throughout our work and in literature that require the enzyme be further improved before the screening of large libraries. The goal of this work is to develop an improved AEH which can be used in future works to alter the substrate range.

The first goal of this work focused on the characterization of a putative AEH from Xanthomonas campestris pv. campestris, ATCC 33913. We successfully isolated, cloned and expressed the enzyme in E. coli and characterized the enzyme for substrate specificity, kinetics, thermostability, and active-site mutability. The purified protein was able to perform both hydrolysis and synthesis of a variety of α-amino β-lactam antibiotics including (R)-ampicillin and cephalixin, with optimal ampicillin hydrolytic activity at 25 °C and pH 6.8. The kinetic parameters for ampicillin hydrolysis at these conditions were $k_{cat}$ of 72.5 s$^{-1}$ and $K_M$ of 1.1 mM. An extensive study of site-directed mutations around the binding pocket strongly suggests that the AEHs feature many residues critical to activity around the active site, including Y82, Y175, D207, D208, W209, Y222, and E309, in addition to those residues forming the catalytic triad.

As the second goal of this work, we characterized the synthetic properties of the
enzymes and furthermore used the enzyme in a two-enzyme cascade system using both PGA and AEH. The synthesis parameters $\alpha$, $\beta_o$, and $\gamma$ for ampicillin, determined here first for this class of proteins, are $\alpha = 0.25$, $\beta_o = 42.8 \text{M}^{-1}$, and $\gamma = 0.23$, and these parameters demonstrate the excellent synthetic potential of these enzymes. Secondly, we took advantage of the unique substrate specificity of AEH to perform a purely aqueous one-pot production of ampicillin from penicillin G and (R)-phenylglycine methyl ester, catalyzed by $\alpha$-amino ester hydrolase and penicillin G acylase. The synthesis was performed in both a one-pot, one-step synthesis (1P1S), resulting in a maximum conversion of 39%, and a one-pot, two-step (1P1S) process, resulting in a maximum conversion of 47%. The two-enzyme cascade is a promising alternative to the current enzymatic two-step, two-pot manufacturing process for semi-synthetic $\beta$-lactam antibiotics which requires intermittent isolation of 6-aminopenicillanic acid.

As third goal, we aimed to improve the thermostability of the AEHs. To improve the industrial relevance of these enzymes the thermostability needs improvement, our characterization work showed that the optimal reaction temperature was 25°C and at 30°C had a half-life of only 5 min. Additionally, improved thermostability has been linked to improved evolvability. Improved evolvability improves the applicability of the enzyme for directed evolution or protein engineering protocol designed to alter its reaction properties. Applying a variant of the structure-guided consensus approach, developed in our lab, we analyzed the proteins for thermostability by evaluating them for their $T_{30}^{50}$ value, which is the temperature at which the observed half-life is 30 minutes. The $T_{30}^{50}$ value of the wild-type protein is 27±2°C. In the first-round of single variants, we obtained nine variants with improved or equivalent thermostability to the wild-type. The best single variant V622I had a $T_{50}^{30}$ of 34°C, but it had reduced specific activity compared to the wild-type. We conducted three more rounds of protein engineering, ultimately resulting in our best overall variant N186D/A275P/V622I/E143H that had both an improved $T_{50}^{30}$ of 35°C and 1.3-fold improvement in specific activity.

Lastly, we begin to explore how to change the substrate specificity of the AEHs. We focus on the role of the carboxylate cluster (D208, E309, D310) and the homologous proteins *Rhodacoccus sp.* cocaine esterase (RhCocE) and J1 Acylase. Variants replacing the homologous residues from both of these proteins were created in both the wild-type protein and the thermostabilized triple-variant. The variants were expressed and purified. All of the variants were tested for hydrolytic activity against phenylglycine methyl ester, ampicillin, and penicillin G; unfortunately none of the variants exhibited activity above background. Folding experiments are in progress.
Overall, we have successfully improved the understanding of the AEH class of enzymes and applied a novel cascade application, demonstrating its unique applicability in the synthesis of β-lactam antibiotics. The improved thermostability will further improve the industrial relevance of AEHs.
A growing threat to public health is the resistance of disease causing microbes to existing drug therapy. To address the growing need for new antibiotics, the tools of biocatalysis and protein engineering can be used to develop novel processes for the production of therapies whose inexpensive manufacturing is not feasible with current technology. Biosynthesis has an advantage over traditional organic synthesis because these processes are highly enantiomerically selective and environmentally benign. Protein engineering can be used to alter the properties of existing enzymes toward the desired conditions and, in our case, substrates. The development of enzymatic pathways to β-lactamase-resistant antibiotics would allow for cheaper, faster manufacturing of second-line antibiotics used to treat antibiotic-resistant bacterial pathogens [1].

1.1 β-lactam Antibiotics

β-lactam antibiotics have been in clinical use for more than 60 years and are currently the most widely used group of antibiotics utilized to treat bacterial infections. The common moiety of all β-lactam antibiotics is a 2-azetidinone ring, more commonly referred to as the β-lactam ring, which is responsible for their bactericidal capabilities [2]. There are four sub-families of the β-lactam antibiotic family that are currently used in clinical practice, the penicillins, cephalosporins, carbapenems, and monobactams, which vary in their ring structure around the β-lactam ring to form unique β-lactam nuclei as shown in Fig. 1.1. The β-lactam nuclei are bonded to different acyl side chains to form different semi-synthetic β-lactam antibiotics. Acyl side chain variations affect microbial activity and chemical stability and thus can lead to expanded spectra
Figure 1.1: The $\beta$-lactam nuclei corresponding to A.) the penicillin sub-family, B.) the cephalosporin sub-family, C.) the carbapenem sub-family, and D.) the monobactam family.

![Diagram of $\beta$-lactam nuclei]

Figure 1.2: Example of the structure of a $\beta$-lactam antibiotic, amoxicillin of bactericidal activity and increased bioavailability [3]. Examples of some commonly prescribed antibiotics are amoxicillin (Fig. 1.2), ampicillin, cephalexin.

$\beta$-lactam antibiotics inhibit the final step of bacterial wall synthesis. Specifically, they act as inhibitors for D-alanyl-D-alanine transpeptidase, more commonly referred to as the penicillin binding proteins, which are responsible for elongating and crosslinking the peptidoglycan in bacterial cell walls [4–6]. $\beta$-lactam antibiotics mimic the structure of the D-alanyl-D-alanine crosslink in peptidoglycan thus acting as an irreversible inhibitor to transpeptidase enzyme as shown in Fig. 1.3 [7,8]. Peptidoglycan is a cross-linked polymer that forms a net-like structure to support the cells structure and protect the cell from the surrounding environment. Failure to form the peptidoglycan structure results in growth inhibition, cell lysis, and ultimately cell
death [5]. β-lactam antibiotics can work on both gram-positive and gram-negative bacteria, but in gram-negative bacteria they must penetrate the cell membrane to get into the cell periplasm to access the penicillin-binding proteins [5]. Even though they are active against a wide range of bacterial pathogens, they have shown very little toxicity to mammalian cells, a fact which has contributed greatly to their success in clinical practice [2]. Resistance of pathogens to first-line (i.e., penicillin G) and broad-spectrum (i.e., ampicillin, and amoxicillin) β-lactam antibiotics has continued to increase because they were introduced in the 1940’s [9]. In addition, diseases which compromise the immune system such as HIV/AIDS have further complicated the issue since bacteria can better thrive in a weakened immune environment. According to the Center for Disease Control (CDC), in the US there are 2 million cases of infectious disease annually, resulting in 90,000 deaths; 70% of cases which resulted in death are resistant to at least one antibiotic.

Over time, pathogens have developed resistance to β-lactam antibiotics in four ways [2,7]:

1. Pathogenic strains produce β-lactamases, enzymes that catalyze the hydrolysis of the β-lactam ring and render the antibiotic ineffective, Fig. 1.4.

2. Pathogenic strains produce mutated penicillin-binding proteins to decrease their
Figure 1.4: $\beta$-lactamase hydrolysis of the $\beta$-lactam ring

binding affinity of the $\beta$-lactam ring.

3. Gram-negative pathogenic strains downregulate their porin channels to limit the entry of the antibiotic into the periplasmic space and thus block their access to the penicillin-binding proteins [5].

4. Pathogenic strains are able to force the eflux of the antibiotic from the cell cytosol.

Perhaps the most drastic case of antibiotic resistance is the case of methicillin-resistant *Staphylococcus aureus*, which has developed resistance by way of downregulation of its porin channels. However, the most common reason for antibiotic resistance is the bacterial strains’ production of $\beta$-lactamase [4,5]. To date, there are more than 450 known $\beta$-lactamases [4]. $\beta$-lactamases can be expressed in both gram-negative and gram-positive bacteria and be either constitutively expressed or induced by exposure to $\beta$-lactam antibiotics [4]. In some cases, the effect of $\beta$-lactamases can be suppressed by administering a $\beta$-lactamase inhibitor along with the antibiotic. FDA approved inhibitors include sulbactam, tazobactam and clavuanate [10].

To combat $\beta$-lactamases, the side chain of $\beta$-lactam antibiotics were enlarged to introduce steric hindrance around the $\beta$-lactam ring. Introduced in 1959, methicillin (Figure 1.5) was the first $\beta$-lactamase-resistant antibiotic; however, shortly after it was introduced methicillin-resistant strains of bacteria emerged. Methicillin is no longer used clinically for treatment; however, it is the standard used to determine antibiotic
Figure 1.5: \(\beta\)-lactamase-resistant antibiotics are currently not accessible to enzymatic pathways for synthesis. Target antibiotics for synthesis include oxacillin, cloxacillin, and nafcillin.

resistance in *Staphylococcus aureus* infections, also known as MRSA—methicillin resistant *Staphylococcus aureus*. Antibiotics such as oxacillin, nafcillin, and cloxacillin (Fig. 1.5) are \(\beta\)-lactamase-resistant antibiotics that are used to treat gram-positive *Streptococcus* and *Staphylococcus* infections. These antibiotics are interesting targets for the development of an enzymatic synthesis.

Despite the growing resistance to \(\beta\)-lactam antibiotics, their use in clinical practices to treat bacterial infections remains prevalent. In 2003 they comprised about 65\% of the total world market for antibiotics, with cephalosporin dosage sales estimated to be $9.9$ billion and penicillin dosage sales estimated to be $5$ billion [11]. Furthermore, about half of the antibacterials listed on World Health Organization’s (WHO) Model List of Essential Medicines are \(\beta\)-lactam antibiotics, with seven on the core list of medicines that are needed for a basic health care system and three on the complementary list of medicines that are needed for priority diseases. The core list medicines are considered to be the most safe, efficacious and cost-effective medicines for current and future health care [12].
1.2 Manufacture of Enzymatic Semi-Synthetic β-lactam Antibiotic

1.2.1 Chemical synthesis

Chemical coupling of the β-lactam moieties with the side chain acyl donor has dominated the industrial production of semi-synthetic β-lactam antibiotics since their discovery in the early 1960s. The majority of semi-synthetic β-lactam antibiotics are still synthesized by way of the Dane anhydride process, which can achieve yields as high as 90% [13]. This process is carried out at temperatures as low as -30°C, uses highly reactive pivaloyl chloride and silylating protection groups, and requires large volumes of dichloromethane, triethylamine and acetone for solvents. Despite recycling solvents and auxiliary reagents where possible, the Dane anhydride process still generates a large amount of non-biodegradable waste. Specifically, 30-40 kg of solid waste are generated for every kg of product in the synthesis of cephalexin [1].

1.2.2 Biocatalytic synthesis

The coupling of the β-lactam nuclei with the acyl side chain can be accomplished enzymatically utilizing penicillin G acylase (PGA). A less studied enzyme, α-amino ester hydrolases (AEH), can also be used for this reaction when the acyl side chain has an amino group in the α-position. Furthermore, the coupling can be carried out under thermodynamic control, which utilizes a non-activated acyl side chain, or kinetic control, which requires an activated side chain (typically an ester or amide). See Fig. 1.6 for a comparison of the thermodynamic and enzymatic synthesis of amoxicillin.

The thermodynamically-controlled synthesis, also referred to as the direct or equilibrium-controlled synthesis, is only able to achieve minimal yields (1% or less) in aqueous medium, as the equilibrium lies very far to the left [14–17]. At typical reaction conditions the substrates are in their ionized forms, which PGA does not accept [17, 19]. Therefore, the thermodynamically-controlled enzymatic synthesis
**Figure 1.6:** The PGA catalyzed coupling of amoxicillin via A.) a thermodynamically-controlled route starting from D-HPG and 6-APA and B.) a kinetically-controlled route starting from D-HPG methyl ester and 6-APA.
usually has very slow rates in addition to very low yields. Adding organic co-solvents to decrease the water activity and shift the equilibrium towards the antibiotic has been studied with limited success [14–17]. For example, in the case of cephalexin synthesis the yield was only improved from 0.75% in aqueous medium to 1.25% in 36% (v/v) triglyme [17]. Employing a solid-to-solid reaction where the product that is formed exceeds its solubility limit has also been studied to shift the equilibrium of the reaction. The solid-to-solid synthesis of amoxicillin in aqueous solutions and with dimethylformamide as a co-solvent was reported as being thermodynamically unfeasible [15], but has shown more promising results with the addition of counter-ions to aide in product precipitation [14].

Typically, the kinetically-controlled synthesis is employed for the coupling of the β-lactam nuclei and the acyl side chain. Unfortunately, the kinetically-controlled reaction results in low yields (30-40%) because PGA has the ability to act as both a hydrolase and an acylase. As a result, PGA catalyzes the undesired primary hydrolysis of the acyl donor and the secondary hydrolysis of the antibiotic in addition to the desired synthesis reaction (Fig. 1.6).

The first published example of the biocatalytic coupling of β-lactam antibiotics on an industrial scale was not until 1997 when DSM opened a production plant for cephalexin, marketed as Purilex, in Barcelona, Spain. Initially it was reported that the penicillin acylase-catalyzed coupling reaction is performed in water between 0°C and 20°C with varying stoichiometric ratios of the nucleus and activated acyl side chain. The enzymatic coupling only achieved 40-60% conversion per cycle and had to be followed by several isolation and recycle steps [1]. Further reports indicate that the yields of the enzymatic coupling were able to be improved by using substrate concentrations as high as 0.3 to 0.5 M and using an excess of the activated acyl side chain [18]. DSM has since expanded their production of enzymatically synthesized β-lactam antibiotics under the umbrella name of DSMPureActives™ to also include
amoxicillin, marketed as Purimox®, and cefadroxil, marketed as Puridrox®. [19]

1.2.3 Comparison of chemical and biocatalytic syntheses

The obvious disadvantage of enzymatic synthesis process in comparison to chemical synthesis process is the lower yield of the product. However, the enzymatic coupling has undeniable advantages over the chemical coupling in terms of cost of raw materials, environmental impact, product quality, and ease of processing due to the fact that it is carried out at ambient temperature, pressure and pH; and does not require toxic or hazardous reagents or solvents [13]. From an environmental standpoint, by switching from the chemical coupling to the enzymatic coupling process, DSM has reduced process greenhouse emissions by over 50%, reduced process emissions to water by 50%, and reduced process waste by 90%. From a product quality standpoint, they have increased the average purity by 1.6%, 1.4%, and 1.2% as assayed “on dry” (HPLC) for Purimox®, Purilex® and Puridrox®, respectively. For an example of the impurities eliminated by switching to the biocatalytic coupling in the production of cephadroxil, see Table 1.1. Furthermore, they report an increased shelf life of the enzymatically synthesized products and better tasting end products for patients. From a processing standpoint, DSM has reduced the number of processing steps by 50% and increased the size of batch production by at least 3-fold and as much as 10-fold. The decrease in number of impurities and increase in the size of batch production translates into a significant reduction in the amount of quality control testing required for all pharmaceutical products. [19]

However, enzymes do not currently exist that are able to catalyze the acylation of larger side chains, such as those associated with the β-lactamase-resistant antibiotics. Both the chemical and enzymatic processes are used industrially, with most of the newer manufacturing facilities using an entirely enzymatic process due to the economic and environmental advantages of the enzymatic process.
Table 1.1: Comparison of Impurities Present in the Enzymatically Coupled Antibiotic Puridrox and the Industrial Average for Chemically Coupled Cefadroxil

<table>
<thead>
<tr>
<th>Impurity</th>
<th>Impurity Source</th>
<th>Puridrox®</th>
<th>Industry Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetone (%)</td>
<td>Solvent</td>
<td>&lt;0.01</td>
<td>0.01–0.08</td>
</tr>
<tr>
<td>Methylene chloride (ppm)</td>
<td>Solvent</td>
<td>NP</td>
<td>1.5–100</td>
</tr>
<tr>
<td>Pivalic acid (ppm)</td>
<td>Solvent</td>
<td>NP</td>
<td>&lt;10</td>
</tr>
<tr>
<td>Dimethylformamide (%)</td>
<td>Solvent</td>
<td>NP</td>
<td>0–0.37</td>
</tr>
<tr>
<td>n-HPG</td>
<td>Undesired product</td>
<td>0.03–0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>7-ADCA</td>
<td>Reactant</td>
<td>0.08–0.16</td>
<td>0.08</td>
</tr>
</tbody>
</table>

1.3 Biocatalysts used to synthesize β-Lactam antibiotics

1.3.1 Penicillin G Acylase (PGA)

As discussed previously, penicillin G acylase (EC 3.5.1.11) is a very important enzyme in the industrial production of semi-synthetic β-lactam antibiotics as it is used to obtain the β-lactam nuclei from naturally occurring antibiotic and can be used for the coupling of the β-lactam nuclei to acyl side side chains. Discovered in 1960, PGAs are thought to be used to generate carbon sources from aromatic compounds in vivo [20]. Bacteria, actinomycetes, yeasts, and fungi all produce PGA [21]. This discussion will be focused on the PGA strain ATCC 11105 from E. coli because it is the most studied and industrially relevant of the PGAs [18].

1.3.1.1 Structure and catalytic mechanism of penicillin G acylases

Mature PGA is a heterodimeric enzyme with a 23.9 kDa α-chain and a 62.4 kDa β-chain [22–24]. The enzyme’s two chains are closely intertwined and form a 70 x 50 x 55 Å protein that is kidney-shaped in its cross section and has a cone-shaped depression in the center of the protein surface, at the base of which lies the catalytic residue [23, 24]. PGA belongs to the N-terminal nucleophile (Ntn) hydrolase superfamily, which is characterized by a single N-terminal catalytic residue and a characteristic four layer α + β structure around the active site [22, 24]. Refer to Fig. 1.7 for the crystal structure of mature PGA and the characteristic fold of Ntn hydrolases.
Like all Ntn hydrolases, PGA is expressed as an inactive precursor that must undergo post-translational processing to render the mature form [22]. Specifically, PGA is expressed as a single cytoplasmic precursor that contains an N-terminal 26 amino acid signal sequence, a 209 amino acid α-chain, a 54 amino acid spacer sequence, and a 557 amino acid β-chain. The signal sequence is responsible for the protein’s transport from the cytoplasm and is cleaved upon the protein’s arrival into the periplasm. The linker sequence aids in the protein’s folding and is subsequently cleaved to expose the active residue. Both the sequence and linker cleavage processes are believed to be autocatalytic [22, 24, 25]. Interestingly, a precursor mutated to undergo slow post-translational processing has shown that the periplasmic precursor protein has an almost identical fold to mature PGA, the two forms have only a 0.71 Å root mean square (r.m.s.) difference between the α carbons in their α and β chains. The periplasmic precursor protein crystal structure shows that the linker peptide is compactly packed against the surface depression and the active site observed in mature PGA [25].

The catalytic mechanism of PGA has been elucidated through crystal structures and mutagenesis studies. PGA has a hydrophobic binding pocket lined with aromatic and hydrophobic residues that is specific for phenylacetyl groups and their
The range of substrate specificity and enantioselectivity of PGA for phenylacetyl derivatives is determined by the size and polarity of their substituents [26]. Residues in the active site undergo a conformation shift upon substrate binding to create a binding site for the penicillin nucleus, therefore, PGA substrate binding is an induced fit mechanism [22,24,26,27]. Bound substrates are in the proper position for the catalytic serine, located at the β1 position, to perform a nucleophilic attack on the carbonyl carbon of the substrate to form an oxyanion tetrahedral transition state and subsequently a acyl-enzyme intermediate [23]. A nucleophile, which can be a β-lactam nucleus (synthesis reaction) or water (hydrolysis reaction), then deacylates the acyl-enzyme intermediate to form products.

### 1.3.1.2 Kinetic parameters of penicillin G acylases

The kinetic parameters of penicillin G acylase are shown in Table 1.2. Ideally, the activity of PGA for the β-lactam antibiotic would be much lower in comparison to the activated acyl side chains so as to enhance the rate of the coupling reaction and decrease the rate of hydrolysis reactions that occur in the kinetically controlled synthesis of β-lactam antibiotics. However, the kinetic parameters reported show that instead $k_{cat}/K_M$ values of the antibiotic are about ten times higher than that of the activated acyl side chains [28,29].

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_M$ (mM)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$k_{cat}/K_M$ (per s per mM)</th>
<th>Temperature (°C)</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>d-PG amide</td>
<td>27.4</td>
<td>35.6</td>
<td>1.3</td>
<td>30</td>
<td>7.0</td>
</tr>
<tr>
<td>d-PG methyl ester</td>
<td>12.5</td>
<td>50</td>
<td>4</td>
<td>30</td>
<td>7.0</td>
</tr>
<tr>
<td>Ampicillin</td>
<td>3.6</td>
<td>36.7</td>
<td>10.19</td>
<td>30</td>
<td>7.0</td>
</tr>
<tr>
<td>Cephalaxin</td>
<td>1.5</td>
<td>29</td>
<td>19.3</td>
<td>30</td>
<td>7.0</td>
</tr>
<tr>
<td>Amoxicillin</td>
<td>1.07</td>
<td>17</td>
<td>15.9</td>
<td>30</td>
<td>7.0</td>
</tr>
<tr>
<td>Cefadroxil</td>
<td>0.642</td>
<td>32</td>
<td>49.8</td>
<td>30</td>
<td>7.0</td>
</tr>
<tr>
<td>Penicillin G</td>
<td>0.013</td>
<td>39</td>
<td>3000</td>
<td>30</td>
<td>7.0</td>
</tr>
</tbody>
</table>
1.3.1.3 Protein engineering of penicillin G acylases

The majority of protein engineering on penicillin G acylase has targeted increasing the synthesis of the antibiotic and decreasing the primary hydrolysis of the activated acyl side chain and the secondary hydrolysis of the antibiotic in the biocatalytic coupling of β-lactam nucleus and the activated acyl side chain. Alkema et al. and Jager et al. have attempted this feat with both rational and combinatorial design but have had limited success [28,30–32]. Alkema et al. examined the role of residue αArg145 and βArg263 using site-directed mutagenesis. αArg145 variants showed an increase in affinity for the β-lactam nucleus as a nucleophile compared to water at the expense of significantly decreased activity. βArg263 was determined to be important for precursor processing but not involved in substrate binding [30]. Jager et al. examined the importance of αArg145 and αPhe146 on PGA nucleophile selectivity via site-saturation mutagenesis. These residues were chosen based on the fact that crystal structures show that they undergo a 3.5 Å conformational shift towards the solvent upon ligand binding. Both sets of variants had improved variants in terms of synthesis to hydrolysis ratios but again at the expense of activity. For example the αArg145Trp had a \((V_{Ps}/V_{Ph})_{init}\) of 14.3, nearly a 10-fold increase over the wild type enzyme, but showed only 6% of the wild type enzyme activity [31].

Protein engineering on PGAs has also been employed to improve its thermostability. Polizzi et al. used the structure-guided consensus concept to target potentially thermostabilizing mutations on 20 residues. Of these variants, 10 (or 50%) had an increased half-life at 50°C, 2 (or 10%) had a half life equivalent to that of the wild type enzyme, four had a decreased half-life and the remaining variant’s activity was too low to determine an accurate half-life. In comparison to wild-type PGA, variants αTrp25Tyr and αAla80Arg improved the half-life of PGA by about 2.5-fold and the activity by 1.5-fold [33].
1.3.1.4 Medium engineering with penicillin G acylases

There have been numerous studies on the effects of different organic co-solvents on the synthesis of semi-synthetic β-lactam antibiotics [34–44]. There are many advantages for enzymatic syntheses in organic solvents including increased solubilities of reactants and products, shifting reaction equilibria, and, most important for the enzymatic coupling step in β-lactam antibiotic synthesis, suppressing hydrolysis reactions. However, most enzymes show significantly decreased activity in organic co-solvents [45]. Illanes and Fajardo studied a variety of co-solvents for enzyme stability and solubility of substrates to determine which they would attempt to apply in the enzymatic synthesis in systems containing 50% (v/v) co-solvent [37]. They found that ethylene glycol gave them the best yield and productivity for the reaction. Kim and Lee studied the synthesis of four different β-lactam antibiotics in water-methanol systems. They found that the rate of the reaction was only slightly lower than that of buffer in systems containing 40% (v/v) methanol or less [41]. Pan et al. studied the effect of the synthesis of ampicillin in fully organic media and attempted to determine a correlate their results for yield with log P values. They were able to obtain a yield of 37.6% in ethyl acetate with a 24 hour reaction time and were not able to correlate their results with log P values [42].

1.3.2 α-Amino Ester Hydrolase (AEH)

α-Amino ester hydrolases (E.C. 3.1.1.43), also referred to as ampicillin acylases, are a family of enzymes that are interesting because they are capable of the synthesis and hydrolysis of semi-synthetic antibiotics [46–53]. In other words, they are alternatives to PGAs. As their name implies, AEHs are highly specific for antibiotics which contain an α-amino acid, such as ampicillin and cephalexin. In fact, they have been shown to hydrolyze β-lactams containing an α-amino group, with a two-fold higher activity compared to penicillin G [54]. Their unique substrate specificity makes them
Figure 1.8: 2B4K crystal structure of the AEH enzyme from A.turbidans bound with (R)-phenylglycine. The enzyme is colored by subunits.

interesting candidates for the enzymatic one-pot synthesis of ampicillin, as the penicillin G hydrolysis product, phenylacetic acid, does not inhibit them [46,49,50,55,56]. Discovered by Takahashi in 1972, recombinant AEHS have been isolated and characterized from A. turbidans, X. citri, and Xanthomonas campestris pv. campestris (this work) [46,49,51,57]. Several other proteins have been isolated from their wild type organisms including a cephalaxin-synthesizing enzyme (CEX) from Gluconobacter oxydans [58], and novel penicillin G acylases Achromobacter sp. CCM 4824 [59] that has the substrate range of the AEHS.

1.3.2.1  Structure and catalytic mechanism α-amino ester hydrolases

The crystal structures for AEH from both A. turbidans and X. citri have been solved; the A. turbidans crystal structure, pdb:2B4K, is shown in Fig. 1.8. The protein is a homotetramer with ~72 kDa subunits, totaling ~288 kDa [46,49]. The ligand binding site is on the interior of the homotetramer and must pass through one of two 15 Å wide entrances to reach the binding region. [49] The enzymes belong to the subgroup of serine hydrolases from the α/β-hydrolase fold family. Hence, the enzyme has a classical Ser-His-Asp catalytic triad and oxyanion hole.

AEHS are unique in that an acidic cluster (Asp-Glu-Asp), termed a carboxylate cluster, has been identified in the active site of the protein. This cluster is associated with the enzyme’s unique α-amino substrate specificity [50,60]. The AEH enzyme
Table 1.3: Selected Kinetic Parameters for AEHs

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}$ ($s^{-1}$)</th>
<th>$k_{cat}/K_m$ (per spore mM)</th>
<th>Temperature (°C)</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. turbidans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d-PG amide</td>
<td>&gt;13</td>
<td>&gt;43</td>
<td>3.3</td>
<td>30</td>
<td>6.0</td>
</tr>
<tr>
<td>d-PG methyl ester</td>
<td>7</td>
<td>1035</td>
<td>148</td>
<td>30</td>
<td>6.0</td>
</tr>
<tr>
<td>Ampicillin</td>
<td>1.0</td>
<td>152</td>
<td>162</td>
<td>30</td>
<td>6.0</td>
</tr>
<tr>
<td>Cephalaxin</td>
<td>0.34</td>
<td>347</td>
<td>1021</td>
<td>30</td>
<td>6.0</td>
</tr>
<tr>
<td>Amoxicillin</td>
<td>2.6</td>
<td>10</td>
<td>3.9</td>
<td>30</td>
<td>6.0</td>
</tr>
<tr>
<td>Cefadroxil</td>
<td>1.7</td>
<td>9.6</td>
<td>6</td>
<td>30</td>
<td>6.0</td>
</tr>
<tr>
<td><strong>X. citri</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d-PG amide</td>
<td>ND</td>
<td>1.6</td>
<td>NR</td>
<td>25</td>
<td>6.4</td>
</tr>
<tr>
<td>d-PG methyl ester</td>
<td>90</td>
<td>1880</td>
<td>21</td>
<td>25</td>
<td>6.4</td>
</tr>
<tr>
<td>Ampicillin</td>
<td>1.2</td>
<td>58</td>
<td>48</td>
<td>25</td>
<td>6.4</td>
</tr>
<tr>
<td>Cephalaxin</td>
<td>1.8</td>
<td>180</td>
<td>90</td>
<td>25</td>
<td>6.4</td>
</tr>
<tr>
<td>Amoxicillin</td>
<td>NR</td>
<td>NR</td>
<td>2</td>
<td>25</td>
<td>6.4</td>
</tr>
<tr>
<td>Cefadroxil</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>25</td>
<td>6.4</td>
</tr>
<tr>
<td>Penicillin G</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>25</td>
<td>6.4</td>
</tr>
<tr>
<td><strong>X. campestris pv. campestris</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ampicillin</td>
<td>1.1</td>
<td>72.5</td>
<td>64.5</td>
<td>25</td>
<td>6.8</td>
</tr>
<tr>
<td>Cephalaxin</td>
<td>2.2</td>
<td>250</td>
<td>85</td>
<td>25</td>
<td>6.8</td>
</tr>
<tr>
<td>Penicillin G</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>25</td>
<td>6.8</td>
</tr>
</tbody>
</table>

ND = not detected, NR = not reported.

is very substrate-specific for α-amino acid derivatives, such as ampicillin, but can accept 7-ACA, 7-ADCA, or 6-APA β-lactam nuclei, apparently with some preference for the 7-ADCA nucleus [46–53]. As such, the enzyme has no detected activity toward Penicillin G [57].

1.3.2.2 Kinetic parameters of α-amino ester hydrolases

The kinetic parameters of AEHs are shown in Table 1.3 [46, 49, 57, 60]. Unlike PGA, the kinetic rates for the activate acyl donor (D-PG methyl ester) is much greater than the antibiotic, resulting in an enzyme that is more favorable for antibiotic synthesis.

1.3.2.3 Protein engineering of α-amino ester hydrolases

AEHs were evaluated for use in the synthesis of the cephalosporins cephalaxin and cephaloglycine from (R)-phenylglycine methyl ester and 7-ACA or 7-ADCA, respectively. Several organisms were used in Takahashi’s studies with the highest product concentrations obtained from the organisms X. citri (IFO 3835) and Acetobacter tur- bidans (ATCC 9325) [51–53].
The synthesis parameters have been reported in the synthesis of cefalexin from 7-ADCA and (R)-phenylglycine methyl ester (DPGME). Conversions have high as 75% have been reported with a non-recombinant preparation of the AEH from X. citri under optimized conditions [61]. Barends et al. reported that mutating the adjacent tyrosine to the catalytic serine to alanine (Y206A) improved the initial synthesis to hydrolysis ratio for the enzyme along with its maximum product yield was improved from 36 to 58% conversion (based on the limiting substrate of 15mM DPGME with 30mM 6APA) [60]. However, there is still significant room for improvement on these enzymes synthesis parameters, as very little work has been done on optimization of the reaction conditions and protein engineering has barely been explored with this class of enzymes.

1.3.3 **Comparison between AEH and PGA**

Penicillin G acylase (PGA) and AEH are promising template enzymes for protein engineering of novel biocatalysis toward the synthesis of complex β-lactam antibiotics. PGA is used commercially in both the production of 6-amino-penicillanic acid (6-APA) from penicillin G and the semi-synthesis of antibiotics such as ampicillin and amoxicillin, which are β-lactamase-sensitive. However, the binding pocket of PGA is delimited by the protein backbone; thus, the binding pocket has little room to expand using amino acid mutations. Previous efforts in our lab by Loo et al. [62] mutated three residues in the binding pocket of PGA(αArg145NNK, αPhe146NNK, βPhe24NNK) and screened 30,000 variants resulted in no variants with altered substrate specificity. AEH’s alternative enzyme fold of homotetrameric α/β-hydrolase versus PGA’s heterodimeric N-terminal serine hydrolase fold render the former an interesting candidate for protein engineering toward improved and/or altered activities over PGA. Furthermore, the AEHs have higher selectivity for synthesis over that of PGA. The stability profile does favor PGA, which has a melting temperature 15°C.
higher than that of AEHs. Overall, AEHs are of interest for the engineering of novel substrate specificities.

1.4 Protein Engineering

Protein engineering is a technique used to alter the properties of a given protein through the introduction of mutations in the D.N.A. that encodes a protein of interest. Protein engineering additionally provides insight into the sequence - structure - function relationships of proteins. The process of protein engineering is constantly under review; investigators are looking for more rational techniques. Reducing screening effort and increased understanding of the protein function will result in a positive feedback loop for other protein engineers [63]. Numerous protein engineering tools have emerged in recent years for the improvement of existing biocatalysts or for their adaptation to novel substrates. A number of different enzymes have been evolved over the past few years, obtaining novel catalyst with improved thermostability, enantioselectivity, and substrate specificity. The selection of protein engineering tools depends upon the following criteria:

1. The expected involvement of amino acid residues close or distant to the active site: changes in activity or selectivity often involve residues close to the active site, where improvements in stability in most cases require targeting residues away from the active site. [64]

2. The library screening capacity: whereas combinatorial design strove for the largest possible library sizes, topped by selection (growth-based selection: up to $10^{12}$ variants) and routinely hitting $10^5 - 10^7$ variants for screening with fluorescent (such as fluorescence-activated cell sorting, FACS [2]) or spectrophotometric assays, data-driven protein engineering strives for the smallest possible library, usually lower than 100, with the techniques described below.
3. The depth and accuracy of available sequence and structure information: crystal structures accurate to <1 Å allow positioning of water, often critical to the catalytic mechanism, while many data-driven concepts depend on the availability of as many as possible sequences with the maximum possible level of identity for accurate predictions.

### 1.4.1 Protein Engineering toward increased thermostability

Thermostability is an important parameter in the development of industrially applicable biocatalyst, improved thermostability increases the life of the enzyme and the number of total turnovers that the enzyme can perform over the life of the catalyst, known a total turnover number (TTN). Several different protein engineering techniques have been successfully used to improve protein thermostability, including error-prone PCR (epPCR) [65], DNA Shuffling via SCHEMA [66–69], temperature value (B-FIT) approach [70] and the structure-guided consensus concept [33, 71] all of which except epPCR require additional structural information.

SCHEMA is a computational algorithm which estimates the amount of disruption caused upon DNA recombination. Libraries of chimeras are scored (E=average disruption) according the number of residue-residue contacts which broken when compared to the parent proteins. Screening can then be focused to those chimeras with
less structural disruption (low E). The SCHEMA algorithm ultimately results in an enhanced probability of functional chimeras with relatively low identity to the parent sequences, effectively generating a family of diverse, folded, and functional proteins. Additionally, many nonfunctional chimeras with low E have regained activity by low error-rate random mutagenesis.

Steipe et al. [72] were the first to demonstrate the consensus concept to create thermostable immunoglobulin domains. This concept determines the most prevalent amino acid at a given position of an alignment of homologous proteins. The resulting sequence of most prevalent amino acids is termed the consensus sequence (Fig. 1.9A). Consensus stabilization of proteins is based on the notion that the frequency of sequence elements correlate with stability contribution [72, 73]. Structure-guided consensus (SGC) was developed to account for scenarios to improve thermostability on proteins have few sequences and low homology (Fig. 1.9B) by introduction of heuristic rules from the structural information. In SGC, residues are targeted based on selection of the consensus residue (usually >50% consensus) from alignment with homologous sequences in addition to structural guidelines as defined in Vazquez-Figueroa, et al. [71]:

1. Must be > 6 Å away from the antibiotic binding site

2. Secondary structure propensity was considered (for example, if the mutation was found in a helix, a helix-stabilizer was not changed to a helix-destabilizer),

3. Amino acids involved in existing direct hydrogen bonds or salt bridges were left unchanged.

SGC has been used successfully in our lab to improve the thermostability of glucose dehydrogenase (GDH). The stability of the His-tagged B. subtilis GDH was improved from ~20 min at 25°C to ~3.5 days at 65°C, a 10^6-fold improvement [71].
Figure 1.10: Half-lives and specific activity for HisGDHs. A) 25C, B) 65C. Error bars for half-lives represent 95% confidence intervals. Error bars for specific activity represent standard deviations. The most stable variant in the top is the least stable in the bottom one. [71]
B-value or temperature factor is obtained from the solved crystal structure and characterizes the mobility of an atom in the crystal structure. In the B-FIT approach, amino acid residues are targeted that have average B-values (the average of the B-value of all the atoms in an amino acid) in the top 20 of all B-values as determined by the B-FITTER program [70, 74]. The higher the B-value indicates the higher mobility of an amino acid residue in the crystal structure, thus indicating lower thermostability. While the B-value approach does target residues, it does not predict what to mutate that residue to; instead NNK or similar libraries are introduced into the site and evaluated iteratively as described by the iterative saturation mutagenesis (ISM) approach [70, 74]. This protocol results in 4-5 individual combinatorial active site saturation testing (CAST) libraries each with sizes of 20-8,000 amino acid mutations, depending on how many sites in each library [75]. The B-value approach has been very successful improving the thermostability of Bacillus subtilis Lipase A with improvements in T_{60}^{50} value, the temperature at which the enzyme half-life is 60 minutes, of >40 °C resulting in 1000–fold improvements in half life in which around 8,000 mutants were analyzed, Fig. 1.11 [74].

Both the SGC and B-value approach are two different protein engineering techniques used to improve thermostability. When comparing these two methods, the B-value and the SGC approach, there is very little overlap in the residues selected for mutation. In analyzes of the Lipase A improved variant the stabilizing residues R33G/K112D/M134D/Y139C/I157M [70, 74] would not have been predicted by the structure-guided consensus approach. In some cases (K112 and Y139) the residue position would have been selected by the SGC approach; however, the consensus residues, K112R and Y189S, do not match the stabilizing residues and do not appear in any of the homologous sequences analyzed. Conversely, the B-value approach would not have selected the stabilizing residues from glucose dehydrogenase E170R/Q252L [71](Fig. 1.10). The residues E170 and Q252 were ranked 180/261
and 217/261 in average B-value as calculated using B-FITTER, respectively. The residues E170 and Q252 are on the interface between subunits which may be why they have relatively low B-values. It is of interest to determine if these methods can be synergistically combined to develop an even more rational stabilization technique.

1.4.2 Protein Engineering toward altered substrate specificity

The desire to alter substrate specificities of biocatalyst continues to be one the key objectives of protein engineering. Protein engineering techniques have been used to achieve this goal including random mutagenesis to computationally designed catalysis. Semi-rational, where target positions are randomized, and rational, where targeted positions are replaced with the desired amino acids, protein engineering designs offer the highest degree of success toward altered substrate activity and insight into the structure function relationships of the biocatalyst. However, improvements on this “new” enzyme to create an industrially viable enzyme benefits from less rational
design strategies, such as the data driven protocol protein sequence activity relationships, ProSAR [76].

One of the most difficult problems in protein engineering is obtain new function with an enzyme that has no detectable initial activity of the desired substrate. Peisajovovich and Tawfik suggested that in addition to the commonly referred to first rule of directed evolution ‘you get what you select/screen for’ it should be followed by a second rule of ‘you should select for what is already there [77].’ This initial activity (even if low) is the key to further success. While random mutagenesis has been the most common approach to approaching this problem, it is of interest to move toward more rational approaches to decrease screening effort and increase mechanistic understanding [63]. In Merck’s and Codexis’s redesign of the ATA-117 D-amino acid amino transferase for the Januvia® (Sitagliptin phosphate) process, they used a truncated substrate to overcome the lack of activity on their target substrate [78]. The wild-type enzyme had detectable activity (albeit low, \(~4\%\) conversion after 24 h) toward the truncated substrate. Savile et al. were able to use semi-rational design via transition state docking to improve the activity towards the truncated substrate which resulted in an enzyme capable aminating the target substrate at low conversion <1% conversion after 24 h. After several more rounds of targeted and random mutagenesis via ProSAR, an enzyme designed for industrial yields and operating conditions was developed, the final enzyme has 27 mutations.

Approaches that also target the active-site such as CASTing along with structural information from docking studies have generated some very well developed biocatalyst. Iterative CASTing and ISM have also been successfully used toward improved enantioselectivity. [79] This technique was most recently used to generate enoate reductases variants that were both (R)- and (S)- selective. [80]. Hu et al. discovered a cytochrome P450 BM-3 variant that altered the regioselectivity of the P450 BM-3 toward indirubin through the evaluation of a single NNK library that was designed
Figures 1.12: ProSAR technique developed by Codexis [76]

based on docking studies. [81]

Codexis’ proprietary algorithm, ProSAR, can be used to evolve the enzymes for
industrial use, keeping the number of screened variants at a minimum. After di-
versity generation via recombination, epPCR, site-directed, and saturation libraries
(Fig. 1.12), statistical models can be developed using the sequence-activity data for
~100 variants. [76]. Although the sequencing is reduced and the results are more
dramatic compared to other directed evolution protocols, the sequencing requirement
is still high at 10^4-10^5 variants and is reliant on effective screening or selection proto-
cols [76].

ProSAR [76], (in conjunction with directed evolution) was successfully used to
enhance the activity and enantioselectivity of several keto-reductases, a halohydro-
gen dehalogenase, and a transaminase toward industrially-relevant substrates. [82–85]
During this process, the resulting catalysts are also optimized toward the desired in-
dustrial reaction conditions.

Other successful approaches include the computational design of active sites. The
Baker lab has been successful using the Rosetta algorithm to design de novo bio-
catalysts including a Kemp elimination enzyme and a Diels-Alderase. [86,87] These
catalysts challenge our understanding on structure-function relationships and significantly broaden our ability to design catalysts not found in nature. Though, this technique requires significant improvement, as the catalytic rates are slow relative to naturally occurring enzymes and thus far the catalytic constants $k_{cat}$ are mostly $< 1 \ s^{-1}$, far below what is required of a viable industrial catalyst. A significantly improved situation concerns Rosetta’s application to design orthogonal nucleoside kinases: the algorithm resulted in successfully identified residues for specificity change and suggested mutations to create the concerted mechanism required for selectivity. [88]

1.4.3 Link between thermostability and altered substrate specificity

In addition to improved enzyme life, thermostability has been linked to the ability of an enzyme to tolerate destabilizing mutations(Fig. 1.13). [77,89–91] The selection of the starting template can also influence the result of gaining new catalytic function. ‘Neutral’ drifted and thermostabilized templates have higher tolerance for mutations, as such they can tolerate considerable mutation toward new specificities. ‘Neutral’ drift is achieved by introducing random mutations, via error-prone PCR(epPCR) or DNA shuffling, and selecting those variants for wild-type activity. [77,89,90] Thus, ‘neutral’ mutations are accumulated in the library in and around the active-site that may increase promiscuity. The ‘neutral’ drift method is very dependent on an effective high-throughput screen or selection assay to obtain the library of neutral mutations. However, it is very straightforward to create a consensus sequence and one does not require any high-throughput selection. The TEM-1 $\beta$-lactamase neutral drift evolution by Bershtein et al. demonstrated that all stabilizing mutations discovered using directed evolution drifted back to the consensus sequence. [90] Therefore, moving toward a thermostabilized template, via the consensus sequence as described in section 1.4.1, before introducing libraries toward altered substrate specificity will increase the number of active variants and have the highest degree of success.
**Figure 1.13:** The effect of a mutation can depend on the stability of the protein into which it is introduced. Figure adapted from Bloom et al. 2009 [91]

### 1.5 Aims

**1.5.1 Aim 1: Cloning, Overexpression, and Characterization of AEH from *X. campestris pv. campestris***

In Chapter II, we demonstrate the cloning, overexpression and full characterization of a novel AEH from *X. campestris pv. campestris*. Since only two AEHs have been fully characterized so far, it was of interest to further expand the number of characterized species in this class to enable combinatorial or data-driven protein engineering techniques. Such protein engineering techniques require multiple functional and fully characterized sequences to be fully efficient. We successfully isolated, cloned and expressed the enzyme in *E. coli* and characterized the enzyme for substrate specificity, kinetics, thermostability, and active-site mutability. The purified protein was able to perform both hydrolysis and synthesis of a variety of α-amino β-lactam antibiotics including (R)-ampicillin and cephalaxin, with optimal ampicillin hydrolytic activity at 25°C and pH 6.8. The kinetic parameters for ampicillin hydrolysis at these conditions were $k_{cat}$ of 72.5 s$^{-1}$ and $K_M$ of 1.1 mM. An extensive study of site-directed mutations around the binding pocket strongly suggests that the AEHs feature many
residues critical to activity around the active site, including Y82, Y175, D207, D208, W209, Y222, and E309, (X. campestris pv. campestris numbering) in addition to those residues forming the catalytic triad. [57]

1.5.2 Aim 2: Characterize the Synthesis Properties and Kinetics of AEH and use AEH in novel two-enzyme cascade with penicillin G acylase (iPGA)

In Chapter III, we characterize the synthesis properties and kinetics of AEH and use AEH in novel two-enzyme cascade with immobilized penicillin G acylase (iPGA). The synthesis parameters $\alpha$, $\beta_o$, and $\gamma$ for ampicillin, determined here first for this class of proteins, are $\alpha = 0.25$, $\beta_o = 42.8 \text{ M}^{-1}$, and $\gamma = 0.23$, and demonstrate the excellent synthetic potential of these enzymes [57]. Secondly, we took advantage of the unique substrate specificity of AEH to perform a purely aqueous one-pot production of ampicillin from penicillin G and (R)-phenylglycine methyl ester, catalyzed by $\alpha$-amino ester hydrolase and penicillin G acylase. The two-enzyme one-pot reaction scheme eliminated the need to isolate the intermediate 6-APA from the phenylacetic acid (PAA) by-product, which a known inhibitor of PGA with a $K_I = 70 \mu\text{M}$ but does not inhibit AEH. The synthesis was performed in both a one-pot, one-step synthesis (1P1S) resulting in a maximum conversion of 39%, and a one-pot, two-step (1P2S) process resulting in a maximum conversion of 47% [56]. Maximum conversions were achieved in one to two hours, significantly reducing the reaction times previously observed in the systems that use iPGA and ethylene glycol [92, 93]. In all cases, the two-enzyme system with iPGA and AEH out-performed the systems that used only iPGA, thus demonstrating the clear advantage of using AEH.

1.5.3 Aim 3: Improve the thermostability of AEH using a modified structure-guided consensus method

In Chapter IV, we improved the thermostability of the AEHs. To enhance the synthetic potential of these enzymes the thermostability requires improvement: we
showed that the optimal reaction temperature is 25°C and that our protein at 30°C has a half-life of only 5 min. Additionally, increased thermostability has been linked to improved evolubility [89–91, 94, 95]. Improved evolubility enhances the enzyme’s tolerance of protein engineering designed to alter its reaction properties. To increase the enzyme’s thermostability, we apply a variation of the structure-guided consensus approach, which investigates both the effects of the B-value [70, 74] and consensus approaches [33, 71]. We analyzed the protein and its variants for kinetic thermostability by evaluating them for their $T_{50}^{30}$ value, the temperature at which the observed half-life is 30 minutes. The $T_{50}^{30}$ value of the wild-type protein is 27±2 °C. In the first-round of single variants, we obtained ten variants out of eighteen with improved or equivalent thermostability to the wild-type. The best single variant V622I had a $T_{50}^{30}$ of 34 °C but reduced specific activity compared to the wild-type. Ultimately, we created a quadruple variant, E143H/N186D/A275P/V622I, that has both an improved $T_{50}^{30}$ of 34 °C and 1.3-fold improvement in specific activity resulting in a significant improvement in total turnover number at 25°C, a dimensionless number characterizing the total number of substrate turnovers per active site over the life of the biocatalyst.

1.5.4 Recommendations and Conclusions

Chapter V, we begin to explore how to change the substrate specificity of the AEHs. We focus on the role of the carboxylate cluster (D208, E309, D310) and the homologous proteins *Rhodococcus sp.* cocaine esterase (RhCocE) [96] and a glutaryl 7-ACA acylase (J1 Acylase). [97] Four mutations replacing the residues at positions D207, D208, E309 and D310 with the homologous RhCocE (D207A/D208P/E309L/D310F) or J1 acylase (D207E/D208V/E309T/D310L) residues were inserted into both the wild-type protein and the thermostabilized triple variant, creating a quadruple and a heptuple variant. The variants were expressed and purified. All of the variants were test for hydrolytic activity against phenylglycine methyl ester, ampicillin, and
penicillin G; however, none of the variants exhibited activity above background. Folding experiments have been conducted using both dynamic light scattering (DLS) and circular dichroism (CD). The results showed the variants with mutations stemming from the RhCocE active site featured the highest degree of folding, but still had some difference from the wild-type. Additionally discussed are the applications and feasibility of an improved selection methods for β-lactam antibiotics. Finally, we will conclude with the future perspectives on how to further improve the enzyme's stability, alternative analytical assays and overall conclusions of this work.

1.5.5 Publication Information

The review presented in Chapter I of this dissertation on the background and history of the industrial enzyme catalyzed synthesis of β-lactam antibiotics is published in Wiley’s Encyclopedia of Industrial Biotechnology (Ed. M.C. Flickenger, 2010, pp 535-567) with Andria Deaguero and my thesis advisor, Andreas Bommarius as co-authors. The review presented on the status of protein engineering is partially published in Current Opinion in Chemical Biology, in press (doi:10.1016/j.cbpa.2010.11.011) with Andreas Bommarius and Michael Abhramson as co-authors.
CHAPTER II

ISOLATION, EXPRESSION, AND CHARACTERIZATION OF AEH FROM XANTHOMONAS CAMPESTRIS PV. CAMPESTRIS, ATCC 33913

2.1 Introduction

Semi-synthetic $\beta$-lactam antibiotics (mostly penicillins and cephalosporins) make up approximately 65% of the total world market of $15$ billion for antibiotics [11]. One of the most efficient ways to produce semi-synthetic $\beta$-lactam antibiotics is via biocatalytic acylation of a $\beta$-lactam moiety, for example 6-aminopenicillanic acid (6-APA), with an activated carboxylic acid, for example (R)-phenylglycine methyl ester ((R)-PGME), as shown in Fig. 2.1. DSM Anti-infectives BV (Delft, Netherlands) currently manufactures amoxicillin, cephalexin, and cefadroxil with an enzymatic process that utilizes penicillin G acylase (PGA, EC 3.5.1.11) as discussed in Chapter I [98]; however, there is still room for improvement in this process. Improved enzymatic production processes lead to lower cost and more environmentally benign production of semi-synthetic antibiotics when compared to traditional chemical synthesis [1, 11, 29, 46, 49, 50, 60, 99, 100] and the existing enzymatic synthesis.

$\alpha$-Amino ester hydrolase (AEH) catalyzes the synthesis and hydrolysis of esters and amides of $\alpha$-amino acids, and can be used as an alternative to the industrially employed penicillin G acylase for the synthesis and hydrolysis of $\alpha$-amino containing antibiotics. While already discovered in 1973 by Takahashi et al. [52], AEHs from both Acetobacter turbidans (ATCC 9325) and Xanthomonas citri (IFO 3835) have only recently been cloned, overexpressed in E. coli, crystallized [49,50,60], and characterized
as to substrate specificity. These enzymes are serine hydrolases with a classical Ser-His-Asp catalytic triad and belong to the structural type of $\alpha/\beta$-hydrolases [46,49,50]. The AEHs are unique in their specificity toward $\alpha$-amino groups; this specificity has been associated with an acidic carboxylate cluster (D208, E309, D310) in the enzyme’s active site. The current understanding of the carboxylate cluster focuses on its involvement in the recognition of the $\alpha$-amino group on the substrate, thus positioning the substrate for catalysis. [50,60] Since only two AEHs have been fully characterized so far, it is of interest to further expand the number of characterized species in this class to enable combinatorial or data-driven protein engineering techniques such as gene recombination [101,102], combinatorial active-site saturation testing (CASTing) [70,74,75,79], or the consensus approach [33,71] to improve properties of these enzymes, such as thermostability, enantioselectivity, and substrate specificity. Such protein engineering techniques require multiple functional and fully characterized sequences to be fully efficient. Incidentally, _X. citri_ is inaccessible in the United States as it is regulated as a plant pathogen causing citrus canker in citrus plants [103], while its cousin _X. campestris pv. campestris_ [104] is available in the ATCC as ATCC 33913. In this investigation, the putative glutaryl 7-aminoccephalosporanic acid (GL7-ACA) acylase from _X. campestris pv. campestris_ (gaa) (GI:21113373) was selected
for characterization. The *gaa* gene was previously identified in a genome study of *X. campestris pv. campestris*; however, the gene has not been previously isolated nor has the protein been tested for activity. The *gaa* gene has a high level of nucleotide alignment (89%) and the protein features a high level of amino acid identity (93%) with AEH from *X. citri* [46]. All key catalytic residues are conserved, including the catalytic triad, carboxylate cluster, and oxyanion hole (see Appendix C for CLUSTAL W alignment).

In this work, we describe the isolation and cloning of the gene from genomic DNA, as well as expression and extensive characterization of the putative 7-ACA acylase from *X. campestris pv. campestris*. We successfully produce active AEH in *E. coli* and report substrate range, kinetic and thermodynamic properties, such as half-life $\tau_{1/2}$, melting temperature $T_m$, and the temperature of half-residual activity at 30 minutes $T_{30}^{50}$ of the purified recombinant protein. We study the role of the acidic carboxylate cluster through alanine replacement of the active site residues. Additionally, we evaluate the enzymes mutability around the active site through site-directed mutations of residues within 5 Å from the ligand.

### 2.2 Results and Discussion

#### 2.2.1 Isolation and Expression

The complete 1914 bp *gaa* gene of *X. campestris pv. campestris* was isolated from the genomic DNA via PCR with primers just outside of the desired gene. The gene was cloned into pET28 with and without the C-terminal 6X histidine tag, resulting in an active construct that is designated pETXcc and pEXccH, respectively, as described in Materials and Methods. The gene encodes a 637 amino acid polypeptide; the gene sequence was confirmed by DNA sequencing. The first twenty-two amino acids encode a signal peptide as predicted by the SignalP 3.0 Server [105], thus resulting in an active peptide subunit of 615 amino acids and a calculated molecular weight of 68 kDa.
and PI of 6.1. These constructs were transformed into *E. coli* BL21(DE3)pLysS cells. Overall expression is 3-5% of the total soluble protein content, similar to expression levels previously reported for AEHs. Activities of soluble lysate from pETXcc and pETXccH were compared indicating that the histidine tag did not impact expression of the plasmid, confirming previous work on the AEH from *A. turidans* [46,50]. The cells were lysed using sonication, this is an extremely sensitive step, as we require to dilute the protein pellet nearly 20-fold to prevent deactivation during sonication. All steps of the protein purification were completed at 4°C to maintain the integrity of the protein. Pure protein was obtained through immobilized metal ion affinity chromatography IMAC purification on Ni-NTA of pETXccH as described in the Materials and Methods section. The SDS-PAGE protein band at 68 kDa is the desired AEH protein (Fig. 2.2). The purified protein is then dialyzed in 50 mM phosphate solution, pH 7, to remove salts and imidizole.
Table 2.1: Kinetic constants for the AEH from *X. campestris pv. campestris*

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_M$ (mM)</th>
<th>$k_{cat}/K_M$ (s$^{-1}$·mM$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R)-ampicillin</td>
<td>72.5</td>
<td>1.1</td>
<td>64.5</td>
</tr>
<tr>
<td>Cephalexin</td>
<td>200</td>
<td>2.2</td>
<td>95</td>
</tr>
<tr>
<td>(R)-PGME</td>
<td>982</td>
<td>3.7</td>
<td>265</td>
</tr>
</tbody>
</table>

[a] All reactions performed at 25°C and pH 7, as described in Materials and Methods
[b] n.d.: not detected

2.2.2 Activity Characterization

The AEH family is specific for esters and amides bearing an $\alpha$-amino group, however, it is able to accept both cephalosporins and penicillins as the $\beta$-lactam moiety. For this reason a range of substrates that include these combinations was tested in hydrolysis conditions. The results are summarized in Table 2.1. The enzyme met the criteria for an AEH as it is active only on $\alpha$-amino containing $\beta$-lactam antibiotics and as thus did not accept penicillin G. The $k_{cat}$ and $K_M$ values of ampicillin hydrolysis at 25°C and pH 6.8 were 72.5 s$^{-1}$ and 1.1 mM (Fig 2.3), respectively.

The hydrolysis activity of AEH on ampicillin as a function of the pH value and buffer system was studied between pH 4 and 10 with an optimum pH of 6.8, Fig. 2.4. The calculated pI is 6.09, thus at optimal operating conditions the protein carries a net negative charge around $-5$.

2.2.3 Thermostability

Fig. 2.5A illustrates the temperature dependence of the AEH activity for the hydrolysis of ampicillin. The optimum temperature was found to be 25°C. Between 6°C
Figure 2.3: Least squares fit of the Michaelis-Menten Equation
and 25°C, the activity increased following the Arrhenius equation, with an activation energy of 56.7 kJ.mol$^{-1}$, see supplemental material for Arrhenius fits. CD scans (Fig. 2.5B) were performed on the pure protein to determine the irreversible melting temperature ($\alpha_{CD}=0.5$) of the protein. The protein was also evaluated by incubating at a range of temperatures for 30 min, resulting in a $T_{m}^{30}$ value of 27°C ±2°C, further illustrating that the enzyme is not very thermostable (Fig. 2.5C). As the rate of deactivation is not clearly 1$^{st}$ order at 30°C, we just report the observed half-life of the protein: 5 minutes (Fig. 2.5D). Furthermore, the deactivation and unfolding both occur in a broad temperature range between 25 and 40°C, indicating that the protein unfolds in a fairly non-cooperative manner. This non-cooperativity is also indicated in the scan rate dependancy of the circular dichroism observed melting temperatures as shown in Fig. 2.6. Using a two-state model, the melting temperature $T_m$ was determined to be between 32.8°C - 40.0°C depending on the scan rate. The dynode voltage as measured simultaneously on the CD is an indication of turbidity, i.e. aggregation of the protein [106]. The onset of aggregation is concurrent with the melting temperature of the protein, suggesting that the unfolding of AEH is irreversible and due to aggregation.
Figure 2.5: The temperature profile for the hydrolysis of ampicillin using the AEH from *X. campestris pv. campestris*. An Arrhenius fit of the temperature activation energy ($E_a = 57$ kJ·mol$^{-1}$, $A = 7.3 \times 10^3$ s$^{-1}$, $R^2 = 0.95$) and deactivation constants ($E_d = -82$ kJ·mol$^{-1}$, $A = 3.4 \times 10^4$ s$^{-1}$, $R^2 = 1.0$). Only the first four deactivation points were considered in the fit, initial activity at $>35^\circ$C is difficult to quantify since the deactivation occurs in <1 min. B. CD thermal scan result fit to a two-state deactivation model, resulting in a calculated $T_M = 32.8$ °C. C. Residual activity of AEH after incubation at reported temperature for 30 minutes, protein was immediately quenched on ice prior to determining residual activity at 25°C as described in the Materials and Methods section. D. Residual activity at 25°C of AEH after incubation at 30°C at reported time.
Figure 2.6: Thermal unfolding and aggregation of wild-type AEH monitored by CD at 222nm and turbidity over range of scan rates A.) CD thermal scan result fit to a two-state deactivation model for scan rates between \( \sim 8 \text{ °C} \cdot \text{hr}^{-1} \) and \( 90 \text{ °C} \cdot \text{hr}^{-1} \) with observed melting temperatures of \( 8 \text{ °C} \cdot \text{hr}^{-1} = 32.8 \text{ °C} \), \( 30 \text{ °C} \cdot \text{hr}^{-1} = 37.3 \text{ °C} \), \( 60 \text{ °C} \cdot \text{hr}^{-1} = 40.4 \text{ °C} \), and \( 90 \text{ °C} \cdot \text{hr}^{-1} = 41.0 \text{ °C} \). B.) Turbidity as monitored by dynode voltage on the CD. The onset of aggregation is \( 30 \text{ °C} \cdot \text{hr}^{-1} = 36.0 \text{ °C} \), \( 60 \text{ °C} \cdot \text{hr}^{-1} = 39.0 \text{ °C} \), and \( 90 \text{ °C} \cdot \text{hr}^{-1} = 41.0 \text{ °C} \).

There is an apparent disconnect between the kinetic thermostability, the temperature at which activity is loss, and the thermodynamic thermostability, the temperature where the protein is denatured. As shown in Figs.2.5A and 2.5C, the initial activity of the protein negligible at 37.0 °C and is irreversibly inactivated after 30 minutes at 36.0 °C, eventhough the protein should still have the majority of it’s secondary structure per the circular dichroism data. This low thermostability may be associated with subunit dissociation that may lead to kinetic inactivation of the protein without degradation of the secondary structure [107, 108]. Another explanation for the premature loss in activity, is that the uncoorporative unfolding may lead to small changes around the active site that lead to kinetic inactivation before the majority of the protein unfolds. Further discussion on this topic may be found in Chapters IV and V.
2.2.3.1 *Half-life studies and deactivation order*

The order of deactivation for the AEHs is mixed. This is most likely due to multiple deactivation mechanisms taking place such as subunit dissociation, which is a \(1^{st}\) order process [109] and aggregation [110] and autoproteolysis [111], which are both \(2^{nd}\) or higher order processes. At \(30^\circ\text{C}\) the deactivation order is strongly \(2^{nd}\) order as can be seen in Fig. 2.7. Using a levenspiel plot, at \(25^\circ\text{C}\) the order was calculated to be 1.5 with an \(R^2=0.95\) and at \(20^\circ\text{C}\) the order was calculated to be 3.6 with an \(R^2=0.86\). Second-order deactivation has been shown for \(\alpha\)-chymotrypsin due to autoproteolysis [111]. Because AEH’s are classified as serine hydrolases they may have protease activity. We followed the SIGMA enzymatic assay for protease activity using the milk protein casein as the substrate and measuring releases tyrosine. The AEH did not show any detectable protease activity, Fig. 2.8.

More complex deactivation models such as the three-state and four-state models of Rogers, *et al.* [112,113] are required to better elucidate an unfolding/deactivation mechanism and is further discussed in Chapter V.

2.2.4 *Role of the Carboxylate Cluster*

Single point mutations were constructed around the active site of the enzyme: Y82A, S174A, Y175(A,R,D,C,H,L,S,V), D207(A,L,N), D208(L,N), W209A, Y222A, D307N, E309(L,Q), and D310(L,N). These residues are all within 5 Å of the active site. S174, D307, H340 (not mutated) are the catalytic triad, Y82A,Y175A are part of the oxyanion hole, and lastly the carboxylate cluster D208, E309, D310. Barends, *et al.* found that a mutation at position Y206A (AEH from *X. campestris pv. campestris* residue Y175A) decreased the enzyme activity but led to improved synthesis properties[6]. The functions of W209 and Y222 are unknown, but are expected to provide bulk to the enzyme pocket and be involved with pi-pi stacking with the substrate. It was
**Figure 2.7:** Thermostability of AEH from *X. campestris pv. campestris* at \( \circ = 4^\circ C, \quad \square = 20^\circ C, \quad \triangle = 25^\circ C, \text{ and } \bigcirc = 30^\circ C. \) Residual activity was measured at 25°C in 20 mM ampicillin up to 8 days. At incubation of 4°C, no activity loss was observed for 30 days. The curve fits are based on second order deactivation.

**Figure 2.8:** Protease Assay, Calibration profile on the left and negative control (empty vector), AEH enzyme, and positive control (trypsin) on the right. No protease activity observed in the test as measured at 660nm on the spectrophotometer.
Table 2.2: Alignment of active site residues of X. campestris pv. campestris AEH, RhCoCE, and J1 Acylase

<table>
<thead>
<tr>
<th>X. campestris pv. campestris</th>
<th>RhCoCE[^a]</th>
<th>J1 Acylase[^b]</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y82</td>
<td>Y44</td>
<td>Y84</td>
<td>Oxyanion hole</td>
</tr>
<tr>
<td>S174</td>
<td>S117</td>
<td>S152</td>
<td>Active Serine, Catalytic Triad</td>
</tr>
<tr>
<td>Y175</td>
<td>Y118</td>
<td>Y153</td>
<td>Oxyanion hole</td>
</tr>
<tr>
<td>D207</td>
<td>A149</td>
<td>E184</td>
<td></td>
</tr>
<tr>
<td>D208</td>
<td>P150</td>
<td>V185</td>
<td></td>
</tr>
<tr>
<td>W209</td>
<td>W151</td>
<td>F186</td>
<td></td>
</tr>
<tr>
<td>Y222</td>
<td>W166[^b]</td>
<td>W200[^b]</td>
<td>Larger aromatic</td>
</tr>
<tr>
<td>D307</td>
<td>D259</td>
<td>D291</td>
<td>Catalytic Triad</td>
</tr>
<tr>
<td>E309</td>
<td>- , L407[^c]</td>
<td>- , T470[^c]</td>
<td></td>
</tr>
<tr>
<td>D310</td>
<td>F261</td>
<td>L263</td>
<td></td>
</tr>
<tr>
<td>H340</td>
<td>H287</td>
<td>H336</td>
<td>Catalytic triad</td>
</tr>
</tbody>
</table>

[^a]: residue numbering includes the signal sequence of both RhCoCE and J1 acylase.
[^b]: The dual W sequence alignment indicates that D165 and S109 should align with Y222. However, in the 3D homology model (Supplemental Material), it is clear that the residues W188 and W220 (based on homology) properly align at this site.
[^c]: While there is a gap in the sequence alignment, 3D alignment using PyMol indicates that these residues fill the space occupied by I309 in X. campestris pv. campestris.

expected that some of these mutations would dramatically affect the enzyme activity; however, previous work with serine hydrolyses suggest that only the mutation on the catalytic serine, S174, should destroy the 100% activity of the enzyme. With the exception of Y175A and D310N, all variants exhibited zero ampicillin conversion after a 24 h incubation at 25 °C. To test if the lack of AEH thermostability masks specificity towards the variants listed above, an additional reaction was conducted at 17 °C; higher ampicillin conversions were achieved for both Y175A (59%) and D310N (40%) after 24 hours under these conditions. The other single variants still did not exhibit any conversion, thus indicating the critical nature of the active site residues, including carboxylate cluster residues D208 and E309, for the activity of this enzyme. Expression was confirmed via SDS-PAGE analysis for all of the variants, enzyme folding was not separately confirmed; however, it is not expected that the enzyme folding compromised since the variants were single point mutations.
Lastly, to investigate the role of the conserved active site residues in other enzymes, we compared the AEH enzyme to the homologous *Rhodococcus sp.* cocaine esterase (RhCocE) [96] and J1 glutaryl 7-ACA acylase (J1) [97]. Both the RhCocE and the J1 enzyme do not require an α-amino group in the substrate, Table 2.2. Both the sequence and crystal structures of AEH pdb:2B4K and pdb:RhCocE 1JU3 were aligned using Clustal W and superimposed using PyMol (DeLano Scientific). The PyMol alignment was anchored on the position of the catalytic triad; from this alignment the α carbon atoms aligned with an RMS of 0.6 Å. In both RhCocE and J1 the catalytic triad and oxyanion hole are conserved, while different amino acid residues are found at positions D207, D208, W209, Y222, E309, and D310. Positions W209 and Y222 are replaced by different aromatic residues suggesting that these changes compensate for size variability in the substrate. In the sequence alignment of *X. campestris pv. campestris* AEH and RhCocE, a gap appears at the E309 position. However, in the PyMol superposition of RhCocE on AEH, it appears that loop residue RhCocE L407 or the homologous J1 T470 residue fills the corresponding space in the binding pocket with a small non-polar amino acid (Fig. 2.9A).

At position D207, the RhCocE contains an alanine while J1 contains a glutamate. Our *X. campestris pv. campestris* AEH variants D207A, D207N, and D207L did not show significant activity toward ampicillin or cephalexin, suggesting that the entire group of acidic residues near the active site, D207, as well as D208, E309, and D310, are critical for protein function. While D207 does not directly interact with the α-amino substrate, as the γ-carbon of D207 is greater than 4 Å from the α-amino moiety on the substrate, the side chain of the residue has polar interactions with the backbone of several second shell residues (W312, G313, I201, D202). These backbone interactions are important in the positioning of its adjacent residue D208 that interacts with substrate as part of the carboxylate cluster. This was further confirmed using a RosettaBackrub model [114] of the D207N mutation. In this model
Figure 2.9: A. PyMol representation of the active-site residues from the AEH from *A. turbidans* shown with (R)-PG bound, pdb 2b4k [60] aligned with the RhcocE. The proteins were aligned using PyMol software using the alpha carbons of the catalytic triad (Xcc SER-174, Xcc HIS-340, and Xcc ASP-307) to an RMS of 0.6 Å. The residue number reflects the numbering for *X. campestris pv. campestris* (XCC) and for the RhcocE. B. Rosetta backrub model [114] for mutation Asp207Asn, the top ten structures are shown (blue) the wild-type structure is shown in green. The position of the Asp208 side chain is repositioned away from the ligand which impacts the polar interactions between (R)-PG and Asp208.

not only were the side-chain to backbone interactions of D207 eliminated, the position of the γ-carbon of the D208 residue was moved 0.8 Å–1.5 Å from the original position, which may disrupt the salt bridge formed with the substrate (Fig. 2.9B).

### 2.3 Conclusions

The *X. campestris pv. campestris* gaa protein is indeed a member of the AEH family of proteins. This protein is an excellent hydrolase with an ampicillin specific activity of >65 U·mg⁻¹. The low thermostability of this protein most likely hinders the ability for the protein to tolerate mutations, which tend to further decrease thermostability, as suggested by both Tokuriki *et al.* and Bloom *et al.* [91, 94, 95]. The current understanding of the carboxylate cluster focuses on its involvement in the recognition
of the $\alpha$-amino group on the substrate, thus positioning the substrate for catalysis. We argue that our results demonstrate that D207 should be regarded as part of the critical residues required for $\alpha$-amino group recognition, just as D208, E309, and D310. Surprisingly, a single alanine replacement of any of the critical residues totally suppressed or at least significantly decreased the wild-type activity of the enzyme. We hypothesize that this observation is due to the need for a highly electron-deficient substrate carbonyl carbon in addition to the requirement of the $\alpha$-amino group to be hydrogen-bonded to the carboxylate residues. Replacement of the negatively charged residues with alanine was insufficient to shift the substrate range to antibiotics that lack the $\alpha$-amino group. We therefore conclude that the carboxylate cluster is critical for both recognition and catalysis. We have successfully expanded the available repertoire of AEHs. The lack of mutability around the AEH active site suggest that these proteins are highly evolved for their substrate specificity toward $\alpha$-amino carboxylic acid electrophiles, as there are many residues associated with its substrate specificity. The enzyme has very good properties toward the substrates and would be useful for industrial applications.

2.4 Materials and Methods

2.4.1 Materials

6-aminopenicillanic acid (6-APA), (R)-phenylglycine (R-PG), (R)-ampicillin (R-AMP), (R)-phenylglycine methyl ester hydrochloride ((R)-PGME), (R)-phenylglycine methyl ester hydrochloride (S-PGME), 7-aminocephalosporanic acid (7-ACA), 7-desacetoxy cephalosporanic acid (7-ADCA), cepahlexin (CEX), penicillin G (PenG) all were procured from Sigma Aldrich (St. Louis, MO). Magic Media was from Invitrogen (Carlsbad, CA), Ni-NTA Superflow Resin was from Qiagen (Germantown, MD). The oligonucleotides for cloning of the gaa gene were provided by Eurofins-mwg—operon Biosciences (Huntsville, AL).
2.4.2 Bacterial Strains and Plasmids

The genomic DNA encoding the putative glutaryl 7-ACA acylase (gaa) of *Xanthomonas campestris pv. campestris* was obtained from the American Type Culture Collection (ATCC 33913D; Manassas, VA). *E. coli* strains BL21(DE3)pLysS (Promega; Madison, WI) and XL1Blue were used for expression and cloning, respectively. The plasmid pET28a (Novagen; Darmstadt, Germany) was utilized as a cloning and expression vector containing a histidine tag (His6) for purification. All PCR reactions were done with recombinant *Pfu* polymerase that was purified from *E. coli* cultivated in our lab.

2.4.3 DNA Sequencing

The DNA sequences were determined by Eurofins-mwg—operon Biosciences (Huntsville, AL).

2.4.4 Cloning of gaa into expression host

For expression of the gaa gene in *E. coli*, the vector pETXCC (gaa cloned in pET28) was constructed. Primers were first used to isolate the gaa gene from the ATCC 33913 DNA, Forward primer 5-CGCAGTGCTGGGAAGACATAT-3 and Reverse primer 5-ATCACCGCAACCACCGACCTTGGAC-3 were used. Forward primer 5-TGCACTGC ATGCCATGATTATGCGTGGCTTGGCCTCC-3 and reverse primer 5CGGC GGCCAAGCTTTTCAACCGACCCGGACAGACTGATG-3 with restriction sites Neo I and Hind III, underlined, were used to incorporate the gaa gene into a pET 28 vector system. A second reverse primer 5-CGGCCGGCCCAAGCTTACGCACGGCAGAC TGATGATG-3 was used to incorporate the C terminal 6X His Tag in the pET 28 vector for ease of purification. After denaturation of the DNA, the amplifications were completed in 30 cycles of 30 s at 98°C, 1 min at 52°C, and 4 min at 72°C. Products and vector were digested with NcoI and HindIII and ligated. The ligation mixture was used to transform chemically competent *E. coli* BL21(DE3)pLysS (Cm_r). The
construct was confirmed by sequencing.

2.4.5 Site-directed mutagenesis

Variants D208A, E309A, and D310A and double and triple variants E209A/D310A, 
D208A/E309A/D310A were generated using overlap PCR and ligated into the pET 28 
vector system. All constructs were confirmed by sequencing. Other active site variants 
were made using the QuikChange approach, primers are reported in Appendix B.

2.4.6 Recombinant overexpression of the AEH

A 5-mL culture was inoculated with a single colony, grown overnight (15 mL test 
tubes, 30 °C, 250 rpm) and subcultured into (1:50 [v:v]) in Invitrogen MagicMedia™E. 
coli expression medium supplemented with 30 µg·mL⁻¹ kanamycin(kan), 30 µg·mL⁻¹ 
chloramphenicol (Cm) (shake flask with baffles, 30°C, 200 rpm). After 6 hours, the 
cells were incubated for 24 hours at 25°C.

2.4.7 Isolation of recombinant gaa gene from E.coli

The cells were harvested by centrifugation and resuspended in 15 mL of cold lysis 
buffer (50 mM potassium phosphate (pH 8.0), 10 mM imidazole, and 300 mM sodium 
chloride) per 1 g of wet cell weight. The clarified cell lysate had a specific activity of 
1.6 U·mg⁻¹ (total 103 U). The mixture was then sonicated and batch purified as per 
the Ni-NTA His-Bind resin protocol under native conditions. Next, the pure protein 
was dialyzed using a Spectra/Por molecular porous membrane MWCO: 12-14,000 
in 50 mM phosphate buffer pH 7.0. The purified enzyme had a specific activity of 
59.7 U·mg⁻¹ (total 89.5 U), which resulted in an overall 87% yield of the desired 
protein. The purified enzyme was analyzed using 12.5% sodium dodecylsulfate-
polyacrylamide gel electrophoresis (SDS-PAGE) loaded with 15 µg of protein per 
lane.

47
2.4.8 Enzyme assays and determination of kinetic constants

Activity of the enzyme was assayed at 25°C by following the hydrolysis of 20 mM ampicillin in 100 mM phosphate pH 7.0 by high performance liquid chromatography (HPLC). Before analysis, the samples were quenched and diluted 10-fold by the addition of HPLC eluent (5 mM phosphate buffer (pH 3), 300 mg·L$^{-1}$ sodium dodecylsulfate (SDS), 30% acetonitrile). The initial rates (< 10% conversion) of hydrolysis of all substrates were determined by measuring product formation by HPLC. To determine kinetic parameters, the enzyme was incubated with varying substrate concentrations in the range of 0-30 mM (R)-ampicillin, cephalexin, and penicillin G each.

2.4.9 High Performance Liquid Chromatography

All analysis was conducted using high performance liquid chromatography complete with a Shimadzu-LC-20AT pump, Beckman Coulter Ultrasphere ODS 4.6 mm x 25 cm column, and SPD-M20A prominence diode array detector (PDA) monitored at 215 nm. The mobile phase is isocratic at 1.0 mL·min$^{-1}$ and contains 30% acetonitrile and 70% 5 mM phosphate buffer with 300 mg·L$^{-1}$ sodium dodecylsulfate (SDS) (pH 3), method adapted from Gabor [115]. All components, (R)-PG, (R)-PG, 6-APA, (R)-PGME, SPGME, (R)-ampicillin, (S)-ampicillin, analyzed on the HPLC with >95% mass balance closure.

2.4.10 Circular Dichromism (CD)

All analysis was conducted using a JASCO CD J-810 with quartz cuvettes. Thermal scans ranged between 5-95 °C and with a scan rate of 8-90 °C·h$^{-1}$. Data was analyzed using non-linear regression algorithm in MatLab (Appendix H) to determine $\Delta H$, $\Delta S$, $\Delta G$, and $T_M$ as described in Greenfield [116].
2.4.11 Publication Information

The work presented in Chapter II of this dissertation on the characterization of AEH is partially published in the Journal of Molecular Catalysis B - Enzymatic (Vol. 67, Issue 1-2, 2010, pp 21-28) with my thesis advisor, Andreas Bommarius as a co-author. The date of online publication was July 6, 2010.
CHAPTER III

SYNTHESIS OF AMPICILLIN USING AEH

3.1 Introduction

3.1.1 Kinetics of the Amidation reaction using AEH

The AEH from X. campestris pv. campestris was utilized to catalyze the kinetically-controlled synthesis of the β-lactam antibiotic ampicillin as shown in Chapter II, Fig. 2.1. Such a kinetically-controlled synthesis raises the issue of selectivity of synthesis versus the two competing hydrolysis reactions, primary hydrolysis of (R)-phenylglycine methyl ester ((R)-PGME) and secondary hydrolysis of ampicillin [29, 60, 117]. The optimal operating point is reached at the time point of maximum synthesis product concentration, $[P]_{\text{max}}$. We determine the synthesis parameters $\alpha$, $\beta_o$, and $\gamma$, which can be used to predict the maximum product yield as defined by equations (3.1 -3.3) [32, 117, 118].

$$
\frac{d[P_s]}{d[P_h]} = \frac{\beta_o[6APA][RPME] - \alpha[P_s](1 + \beta_o\gamma[6APA])}{(1 + \beta_o\gamma[6APA])([RPME] + \alpha[P_s])}
$$

(3.1)

$$
\alpha = \frac{(k_{\text{cat}}/K_M)P_s}{(k_{\text{cat}}/K_M)P_{\text{RPME}}}
$$

(3.2)

$$
\left(\frac{\nu_{Ps}}{\nu_{Ph}}\right)_{\text{init}} = \frac{1}{\gamma} \left(\frac{[6APA]}{1/\beta_o\gamma + [6APA]}\right)
$$

(3.3)

Here, $P_s$ is the synthesis product ampicillin, $P_h$ is the hydrolysis product (R)-phenylglycine, $\nu_{Ps}$ is the initial synthesis rate of ampicillin, and $\nu_{Ph}$ is the initial formation rate of the hydrolysis product (R)-phenylglycine. The differential equation (3.1) can be solved numerically to determine $[P]_{\text{max}}$. These equations (3.1 -3.3) have been successfully used to characterize the kinetically controlled synthesis of ampicillin.
3.1.2 One-pot synthesis using AEH and PGA

Cascade conversions, which combine multiple reactions without intermediate recovery steps, are increasingly studied to render syntheses more environmentally benign and economically advantageous. Replacing a multistage synthesis with a cascade process eliminates the need for isolation and purification of intermediates and therefore results in smaller reactor volumes, shorter cycle times, higher volumetric and space time yields, and decreased amount of waste produced [119–121]. Cascade conversions can combine multiple biocatalytic steps, multiple chemocatalytic steps, or can combine both biocatalytic and chemocatalytic steps. Typically it is easiest to combine multiple biocatalytic steps as most enzymes have similar operating conditions [120]. There have been several reports of utilizing cascade processes for semi-synthetic β-lactam antibiotic synthesis. Wegman et al. combined the synthesis of the acyl side chain (R)-phenylglycine amide from (R)-phenylglycine nitrile utilizing nitrile hydratase and the enzymatic coupling of (R)-phenylglycine amide with the β-lactam nucleus 7-aminodesacetoxy-cephalosporanic acid utilizing PGA to synthesize cephalaxin in a one-pot synthesis [122]. Fernández-Lafuente et al. reported a chemoenzymatic synthesis of cefazolin that started from the naturally occurring cephalosporin C and involved three biocatalytic transformations in fully aqueous medium [123,124]. Finally, Du et al. and Wu et al. employed PGA in partially organic media to catalyze both the hydrolysis of penG to the β-lactam nucleus 6-APA and the enzymatic coupling of 6-APA with (R)-phenylglycine methyl ester ((R)-PGME) or D-hydroxyphenylglycine methyl ester to synthesize ampicillin (AMP) [92] or amoxicillin [93], respectively, in a one-pot system.

We examined the feasibility of utilizing a cascade conversion with two biocatalytic reactions in fully aqueous medium to synthesize AMP as shown in Fig. 3.1. In the first reaction, 6-APA was produced from the thermodynamically-controlled hydrolysis of penG with immobilized penicillin G acylase (iPGA). The by-product from this
Figure 3.1: One-Pot, two-enzyme direct conversion of penicillin G (PENG) to ampicillin (AMP) using iPGA and AEH. Undesired side reactions, primary hydrolysis of (R)-phenylglycine methyl ester ((R)-PGME) to (R)-phenylglycine ((R)-PG) and secondary hydrolysis of AMP, are shown.

reaction, phenylacetic acid (PAA), is a known inhibitor of PGA with a $K_I = 70 \mu M$ [125]. In the second reaction, AMP was produced in a kinetically-controlled coupling of 6-APA with (R)-PGME using either iPGA or AEH [57]. As AEHs are unique in their specificity toward $\alpha$-amino groups on the acyl moiety, they cannot catalyze the hydrolysis of penG to yield 6-APA and are not inhibited by PAA, thus their advantage in this cascade [49,57]. In addition to the desired coupling reaction, both PGA and AEH catalyze the undesired primary hydrolysis of the activated acyl side chain, (R)-PGME, and the secondary hydrolysis of the antibiotic, AMP. These two side reactions negatively affect yield [13].
We investigated both a one-pot, one-step (1P1S) and one-pot, two-step (1P2S) scheme. In the 1P1S scheme, a batch process, we added (R)-PGME, penG, and either iPGA or both iPGA and AEH at the beginning of the experiment. In the 1P2S scheme, we first added penG with iPGA and allowed the reaction to proceed near completion to produce 6-APA. Next, we added (R)-PGME and either AEH or additional iPGA to the reaction mixture. We investigated the effect of different relative enzyme loadings on the overall yield of AMP for both schemes.

3.2 Results and Discussion

3.2.1 Synthesis using AEH alone

3.2.1.1 Synthesis of Ampicillin from (R)-PGME and pure 6-APA

(R)-Ampicillin was first synthesized through enzymatic amidation of (R)-PGME with 6-APA in purely aqueous system. This reaction is coupled with two enzyme-catalyzed side reactions, primary hydrolysis of (R)-PGME to (R)-PG and methanol and secondary hydrolysis of (R)-ampicillin to (R)-PG and 6-APA. A typical conversion profile for the synthesis of ampicillin using AEH is shown in Fig. 3.2A. In this case the electrophile ((R)-PGME) is in excess resulting large amounts of the (R)-PG primary hydrolysis product. The ampicillin synthesis reaction is rapid with an initial AEH turnover number of $\sim 230 \text{s}^{-1}$ for ampicillin production.

The nucleophile (6-APA) to electrophile ((R)-PGME) ratio has a large influence on the outcome of the synthesis reaction. Experiments were run over a range of [6-APA]:[(R)-PGME] ratios between 0.33–3.0. At high ratios the reaction proceeded quickly but has high degrees of secondary hydrolysis, while at lower ratios the reaction proceeded more slowly and interesting exhibited little to no secondary hydrolysis as summarized. The results of the synthesis reaction are shown in Table 3.1 and Fig. 3.3. At [6-APA] = 20 mM and [(R)-PGME] = 60 mM, the maximum product concentration $[P]_{max}$ achieved was $\sim 10$ mM, or 50% conversion with respect to 6-APA. The reaction
Table 3.1: Impact of varying 6APA:((R)-PGME) (same as DPGME) ratios on the maximum product yield, initial synthesis rate, and synthesis to hydrolysis ratios

<table>
<thead>
<tr>
<th>6-APA:DPGME Ratio</th>
<th>Pmax (mM)</th>
<th>Initial Ampicillin Rate (mM/sec)</th>
<th>Initial S:H ratio</th>
<th>Secondary hydrolysis (mM/sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60:20</td>
<td>3.00</td>
<td>10.1</td>
<td>0.023</td>
<td>1.60</td>
</tr>
<tr>
<td>45:20</td>
<td>2.25</td>
<td>9.4</td>
<td>0.020</td>
<td>1.34</td>
</tr>
<tr>
<td>33:20</td>
<td>1.65</td>
<td>8.0</td>
<td>0.016</td>
<td>1.10</td>
</tr>
<tr>
<td>40:35</td>
<td>1.14</td>
<td>13.5</td>
<td>0.017</td>
<td>0.88</td>
</tr>
<tr>
<td>20:20</td>
<td>1.00</td>
<td>6.5</td>
<td>0.014</td>
<td>0.68</td>
</tr>
<tr>
<td>20:33</td>
<td>0.61</td>
<td>8.4</td>
<td>0.015</td>
<td>0.59</td>
</tr>
<tr>
<td>20:45</td>
<td>0.44</td>
<td>10.3</td>
<td>0.016</td>
<td>0.47</td>
</tr>
<tr>
<td>20:60</td>
<td>0.33</td>
<td>10.4</td>
<td>0.016</td>
<td>0.43</td>
</tr>
</tbody>
</table>

stalled at 40 minutes with substrate still available: concentrations at 40 min are 10 mM (AMP), 10 mM (6-APA), 35 mM ((R)-PG), and 15 mM ((R)-PGME). Even after 24 hours under these conditions the product remained in solution at 10 mM and no secondary ampicillin hydrolysis was observed. Enzyme activity has been confirmed for the duration of the synthesis; thus, the lack of conversion is not due to poor enzyme stability. Furthermore, secondary hydrolysis can occur when using < 40 mM (R)-PGME as the substrate (Fig. 3.3). Similar results for secondary hydrolysis were obtained for cephalexin synthesis with partially pure enzyme isolated from X. citri (non-recombinant), when using a 20 mMmM (7-ADCA):60 mMmM ((R)-PGME) ratio.

3.2.1.2 Enantioselectivity

Employing X. campestris pv. campestris AEH (S)-ampicillin was synthesized analogously to the (R)-enantiomer (Fig. 3.2). The initial synthesis rate of (S)-ampicillin was 0.31 mM·min⁻¹ over the linear portion of the synthesis curve, only one seventh of the (R)-ampicillin synthesis rate of 2.0 mM·min⁻¹, indicating the enzyme’s preference for the (R)-substrate. Enantioselectivity was also tested in the competing synthesis of
Figure 3.2: A.) Kinetically-controlled synthesis of (R)-ampicillin with AEH from *X. campestris pv. campestris* starting with 20 mM 6-APA and 60 mM R-PGME B.) Kinetically-controlled synthesis of (R)-ampicillin employing AEH from *X. campestris pv. campestris* starting with 20 mM 6-APA and 60 mM S-PGME C.) Synthesis of (R/S)-ampicillin starting with 20 mM 6-APA and 90 mM racemic PGME. D.) Plot of the change in enantioselectivity of the reaction over time, when starting with racemic PGME.
both (R)- and (S)-ampicillin (Fig. 3.2C) employing a starting ratio of 90 mM ((R/S)-PGME): 20 mM (6-APA). While the enzyme has a preference for (R)-ampicillin, the resulting enantiomeric excess e.e.\textsubscript{p} of the ampicillin product was only 22% at steady state (Fig. 3.2D), resulting in an apparent enantiomeric ratio E\textsubscript{app} of only 2.4 (Equation 3.4). Higher overall yields were observed when starting with racemic PGME, with [P]\textsubscript{max} at 14 mM resulting in 70% conversion of the 6-APA substrate.

\[
E_{\text{app}} = \frac{\ln(1 - x(1 + \text{e.e.}\textsubscript{p}))}{\ln(1 - x(1 - \text{e.e.}\textsubscript{p}))}
\]  

(3.4)

3.2.1.3 Calculation of kinetic constants

Kinetic parameters \(\alpha\), \(\beta\), and \(\gamma\) were determined to fully characterize the synthetic properties of this enzyme. \(\alpha\)-values for the synthesis of ampicillin using the wild-type AEHs range from \(\alpha = 0.15\)-1.09 for \textit{A. turbidans} to \(\alpha = 2.2\) for \textit{X. citri}, calculated from
Figure 3.4: [6-APA] (M) vs. the inverse initial synthesis to hydrolysis ratio, [S:H]$_o$^{-1}. The data was fit to equation 3.3 using a least squares regression. Parameter $\beta_o = 42.8\, \text{M}^{-1}$ and parameter $\gamma = 0.23$ were calculated with an $R^2$ of 0.99. Initial synthesis to hydrolysis rates were determined using linear fits of rates during the first 5 minutes of synthesis reactions.

Previously reported data [46, 49, 50, 60]. Reported $\alpha$-values for penicillin G acylases range $\alpha = 7.6$-16.4 for (R)-PGME [115]. The $\alpha$-value for the AEH from *X. campestris pv. campestris* was calculated from the initial hydrolysis data, Chapter II Table 2.1 to be 0.25, which is at the low end of the calculated range. A low $\alpha$ is desirable as it is an indication for the enzyme’s specificity for the synthesis substrate ((R)-PGME) over that of the acyl donor, whereas a high $\alpha$ would indicate high secondary hydrolysis rate. The nucleophilic reactivity is defined by the $\beta_o$, and $\gamma$ values. $\beta_o$, and $\gamma$ were determined from a non-linear regression of a plot of [6-APA]$_o$, which ranged from 20 mM-60 mM, at constant [(R)-PGME]$_o$ concentration of 20 mM, versus the initial synthesis:hydrolysis-ratio observed during synthesis reactions. The parameters were calculated to be $\beta_o = 42.8\, \text{M}^{-1}$ and $\gamma = 0.23$ with an $R^2$ of 0.99 (Fig. 3.4). It is preferable that both $\beta_o$ and $1/\gamma (1/\gamma = 4.3$) values to be high [115, 118] to indicate low formation of the hydrolytic by-products from both primary and secondary hydrolysis. The $\beta_o$, and $\gamma$ values for the wild-type penicillin G acylases synthesis of ampicillin ranges from $\beta_o : 0.06$–0.50 mM$^{-1}$ and $1/\gamma$: 6–16 [31,115,126], with mutant penicillin G
Figure 3.5: Plot of \([\text{RPG}]\) versus \([\text{AMP}]\) results for six synthesis reactions at various \([\text{[6APA]}]_0:[\text{RPGME}]_o\) ratios and concentrations (20 mM:20 mM, 33 mM:20 mM, 45 mM:20 mM, 60 mM:20 mM, 20 mM:33 mM; 20 mM:45 mM; 20 mM:60 mM), fit to the model equation (equation 1) with parameters \(\alpha = 0.25, \beta_o = 42.8 \text{ M}^{-1}\), and \(\gamma = 0.23\). The experimental results are shown by the data points, while the model fits are represented by the lines.
Table 3.2: One-Pot, One-Step, Conversion results from the one-pot one-step (1P1S) reaction configuration.

<table>
<thead>
<tr>
<th>Enzyme loading (U)</th>
<th>$t^\text{min}$ [min]</th>
<th>Moles of $\text{PGME}$ per mole of $\text{AMP}_{\text{max conv.}}$</th>
<th>$n$ [mol/mol $^n$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPGA (U)</td>
<td>AEH (U)</td>
<td>24.8</td>
<td>11</td>
</tr>
<tr>
<td>99.2</td>
<td>1.1</td>
<td>360</td>
<td>8.7</td>
</tr>
<tr>
<td>99.2</td>
<td>2.2</td>
<td>300</td>
<td>6.3</td>
</tr>
<tr>
<td>99.2</td>
<td>4.4</td>
<td>60</td>
<td>7.5</td>
</tr>
<tr>
<td>99.2</td>
<td>5.5</td>
<td>60</td>
<td>11</td>
</tr>
<tr>
<td>99.2</td>
<td>none</td>
<td>360</td>
<td>31</td>
</tr>
<tr>
<td>114</td>
<td>none</td>
<td>1500</td>
<td>21</td>
</tr>
<tr>
<td>139</td>
<td>none</td>
<td>1500</td>
<td>20</td>
</tr>
<tr>
<td>136</td>
<td>none</td>
<td>360</td>
<td>25</td>
</tr>
</tbody>
</table>

(a) $\text{AMP}$ concentration is not limiting. (b) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (c) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (d) $\text{AMP}$ concentration is not limiting. (e) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (f) $\text{AMP}$ concentration is not limiting. (g) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (h) $\text{AMP}$ concentration is not limiting. (i) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (j) $\text{AMP}$ concentration is not limiting. (k) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (l) $\text{AMP}$ concentration is not limiting. (m) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (n) $\text{AMP}$ concentration is not limiting. (o) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (p) $\text{AMP}$ concentration is not limiting. (q) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (r) $\text{AMP}$ concentration is not limiting. (s) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (t) $\text{AMP}$ concentration is not limiting. (u) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (v) $\text{AMP}$ concentration is not limiting. (w) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (x) $\text{AMP}$ concentration is not limiting. (y) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (z) $\text{AMP}$ concentration is not limiting. (aa) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (bb) $\text{AMP}$ concentration is not limiting. (cc) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (dd) $\text{AMP}$ concentration is not limiting. (ee) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (ff) $\text{AMP}$ concentration is not limiting. (gg) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (hh) $\text{AMP}$ concentration is not limiting. (ii) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (jj) $\text{AMP}$ concentration is not limiting. (kk) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (ll) $\text{AMP}$ concentration is not limiting. (mm) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (nn) $\text{AMP}$ concentration is not limiting. (oo) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (pp) $\text{AMP}$ concentration is not limiting. (qq) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (rr) $\text{AMP}$ concentration is not limiting. (ss) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (tt) $\text{AMP}$ concentration is not limiting. (uu) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (vv) $\text{AMP}$ concentration is not limiting. (ww) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (xx) $\text{AMP}$ concentration is not limiting. (yy) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (zz) $\text{AMP}$ concentration is not limiting. (aaa) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (bbb) $\text{AMP}$ concentration is not limiting. (ccc) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (ddd) $\text{AMP}$ concentration is not limiting. (eee) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (ffe) $\text{AMP}$ concentration is not limiting. (gff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (hff) $\text{AMP}$ concentration is not limiting. (iff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (jff) $\text{AMP}$ concentration is not limiting. (kff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (lff) $\text{AMP}$ concentration is not limiting. (mff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (nff) $\text{AMP}$ concentration is not limiting. (off) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (pff) $\text{AMP}$ concentration is not limiting. (qff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (rff) $\text{AMP}$ concentration is not limiting. (sff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (tff) $\text{AMP}$ concentration is not limiting. (uff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (vff) $\text{AMP}$ concentration is not limiting. (wff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (xff) $\text{AMP}$ concentration is not limiting. (yff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (zff) $\text{AMP}$ concentration is not limiting. (aaff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (bbff) $\text{AMP}$ concentration is not limiting. (ccff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (ddff) $\text{AMP}$ concentration is not limiting. (eedf) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (fffe) $\text{AMP}$ concentration is not limiting. (gffe) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (hffe) $\text{AMP}$ concentration is not limiting. (ifff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (jfff) $\text{AMP}$ concentration is not limiting. (kfff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (lfff) $\text{AMP}$ concentration is not limiting. (mfff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (nfff) $\text{AMP}$ concentration is not limiting. (offf) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (pfff) $\text{AMP}$ concentration is not limiting. (qfff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (rfff) $\text{AMP}$ concentration is not limiting. (sfff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (tfff) $\text{AMP}$ concentration is not limiting. (ufff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (vfff) $\text{AMP}$ concentration is not limiting. (wfff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (xfff) $\text{AMP}$ concentration is not limiting. (yfff) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (zfff) $\text{AMP}$ concentration is not limiting. (aafffe) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (bbffe) $\text{AMP}$ concentration is not limiting. (ccffe) When $\text{AMP}$ concentration is limiting, $\text{AMP}$ is re-added. (ddffe) $\text{AMP}$ concentration is not limiting.

acetyl PAS2 (TM33) reported to have $\beta_o = 10.2 \text{ mM}^{-1}$ and $1/\gamma = 286$ for ampicillin synthesis [115]. To the best of our knowledge, $\beta_o$, and $\gamma$ values have not previously been calculated for this class of enzymes.

The kinetic parameters are able to closely estimate $P_{\text{max}}$ when fit to equations (3.1 -3.3) (Fig. 3.5). While the model appropriately tracks the raw data (Fig. 3.5), it should be noted that when (R)-PGME is in excess (20:45, 20:60) the model does not predict the lack of secondary hydrolysis when the (R)-PG concentration reaches $\sim$32 mM, as observed by the clusters of data points on the 20:45 and 20:60 curves. This finding suggests an inhibition effect of (R)-PG at this concentration.

3.2.2 Ampicillin Synthesis using a Two-Enzyme Cascade System with both $\alpha$-Amino Ester Hydrolase and Penicillin G Acylase

We evaluated both the 1P1S and 1P2S systems over a range of iPGA and AEH concentrations as shown in Tables 3.2 and 3.3. In this cascade, enzyme concentrations have a large effect on the overall yield achieved and the degree of secondary hydrolysis observed. Typical reaction profiles for both configurations are shown in Figs. 3.6.
Figure 3.6: Reaction profile the enzymatic conversion of penicillin using 99.2 UPenG of iPGA and 2.2 of UAmP AEH. Both the 1P1S (A) and the 1P2S (B) profiles are shown. The components are (R)-phenylglycine ((R)-PG), 6-APA, phenylacetic acid (PAA), ampicillin (AMP), (R)-phenylglycine methyl ester ((R)-PGME) and penicillin G (penG).
It has been previously shown that the initial ratio of (R)-PGME to 6-APA concentrations is an important parameter in optimizing the coupling reaction for semisynthetic antibiotics. In our experiments we targeted a (R)-PGME:6-APA ratio of 60 mM: 20 mM, which has been demonstrated as the optimal ratio for the AEH-catalyzed syntheses, as discussed in section 3.2.1.1.

3.2.2.1 One-pot, One-step (1P1S)

The two-enzyme 1P1S system resulted in AMP yields between 6% and 39%, as shown in Table 3.2 and Fig. 3.7A. The system performed poorly with low iPGA enzyme loading (22 UPenG) and high AEH enzyme loading (11 U Amp). AEHs have excellent (R)-PGME hydrolytic activity ($k_{cat} = 982 \text{ s}^{-1}$) [57], thus the majority of the (R)-PGME was hydrolyzed prior to the production 6-APA that is necessary for synthesis. Increased iPGA enzyme loading (99 UPenG) and decreased AEH enzyme loading (between 1.1 U Amp 5.5 U Amp) improved the AMP yields. The optimal configuration resulted in a 39% yield and was observed when 99 UPenG iPGA and 4.4 U Amp AEH was utilized. This configuration had a ratio of moles of (R)-PGME consumed per moles of AMP at the maximum product concentration (\(([(\text{R})-\text{PGME}]_{t=0} / [(\text{R})-\text{PGME}]_{t=\text{AMPmax}}) / [\text{AMP}]_{t=\text{AMPmax}}\) of 7.5. In the one-enzyme 1P1S system with iPGA the reactions only achieved a maximum conversion of 10% after 24 hours. The reduced reaction yield using iPGA alone was expected, due to the strong inhibition of \textit{E. coli} PGA with the intermediate PAA and the preference of \textit{E. coli} PGA for penG ($K_M=0.013 \text{ mM}$) over (R)-PGME ($K_M = 12.5 \text{ mM}$) [29,115,127].

3.2.2.2 One-pot, Two-step (1P2S)

The two-enzyme 1P2S system resulted in AMP yields between 27% and 47% as shown in Table 3.3 and Fig. 3.7B. Several configurations of enzyme loadings led to yields around 47%, which is equivalent to the observed yields when catalyzing the synthesis reaction with AEH directly from 6-APA and (R)-PGME [57]. In the IP2S system the
Figure 3.7: Ampicillin conversion profiles for both the 1P1S and 1P2S. In the 1P2S reaction profiles there is no observed ampicillin until the second reaction step that was initiated 60–140 minute into the reaction.

Enzyme loading of AEH mostly impacted the secondary hydrolysis and decreased AEH loadings (between 1.1 and 4.4 UAmp) reduced the amount of secondary hydrolysis observed. The optimal configuration resulted in a 46% yield with minimal secondary hydrolysis and was observed when 99 UPenG iPGA and 4.4 UAmp AEH was utilized. This configuration had a ratio of moles of (R)-PGME consumed per moles of AMP at the maximum product concentration of ∼6. Similar to the 1P1S configuration, the single enzyme systems using iPGA resulted in low yield with a maximum conversion of 15% after 23 hours.

To investigate the impact of the excess iPGA on the secondary hydrolysis in the system we conducted a one-pot, two-step, two-stage (1P2S-2S) scheme where iPGA was removed by filtration prior to the addition of AEH to the system in the second step. The removal of iPGA did not reduce the observed secondary hydrolysis of AMP, and therefore was not deemed beneficial to the 1P2S scheme.
Table 3.3: One-Pot, Two-Step, Two-Stage Conversion results from the one-pot two-step (1P2S) reaction configuration.

<table>
<thead>
<tr>
<th>iPGA [UPenG]</th>
<th>Step 2 Enzyme loading&lt;sup&gt;[a]&lt;/sup&gt;</th>
<th>AEH [UAmP]</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Total</th>
<th>Moles of α-PGME per mole of AMP at max conv.</th>
<th>Maximum conversion&lt;sup&gt;[c]&lt;/sup&gt; [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.8</td>
<td>none</td>
<td>11</td>
<td>145</td>
<td>15</td>
<td>160</td>
<td>6.0</td>
<td>47</td>
</tr>
<tr>
<td>99.2</td>
<td>none</td>
<td>1.1</td>
<td>60</td>
<td>300</td>
<td>360</td>
<td>6.9</td>
<td>27</td>
</tr>
<tr>
<td>99.2</td>
<td>none</td>
<td>2.2</td>
<td>60</td>
<td>180</td>
<td>240</td>
<td>6.3</td>
<td>35</td>
</tr>
<tr>
<td>99.2</td>
<td>none</td>
<td>4.4</td>
<td>60</td>
<td>90</td>
<td>150</td>
<td>6.2</td>
<td>46</td>
</tr>
<tr>
<td>99.2</td>
<td>none</td>
<td>5.5</td>
<td>60</td>
<td>30</td>
<td>90</td>
<td>6.1</td>
<td>47</td>
</tr>
<tr>
<td>24.8&lt;sup&gt;[a]&lt;/sup&gt;</td>
<td>none</td>
<td>11</td>
<td>130</td>
<td>20</td>
<td>150</td>
<td>6.1</td>
<td>45</td>
</tr>
<tr>
<td>24.8</td>
<td>none</td>
<td>74</td>
<td>130</td>
<td>410</td>
<td>540</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>99.2</td>
<td>15</td>
<td>none</td>
<td>60</td>
<td>1290</td>
<td>1350</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>99.2</td>
<td>30</td>
<td>none</td>
<td>60</td>
<td>1290</td>
<td>1350</td>
<td>15</td>
<td>14</td>
</tr>
</tbody>
</table>

[a] In ampicillin synthesis reactions starting from 6-APA and α-PGME, 1 UAmP of AEH≈6.8 UPenG of iPGA.  
[b] Time at which maximum conversion was obtained.  
[c] Conversions are based on the moles of ampicillin produced per moles of penicillin G starting material. All concentrations are based on analytical measurements, not isolated yields.  
[d] iPGA removed from the second step using filtration in the one-pot, two-step, two-stage process.

3.2.2.3 Discussion

The 1P1S system required fewer manipulations and has an overall faster cycle time but resulted in a lower overall yield when compared to the 1P2S system. The lower observed yields are likely due to the lower initial 6-APA nucleophile concentrations as 6-APA is generated at the same time it is consumed. The 1P2S step system required higher cycle times but results in higher overall yields and allowed for the most control of the system parameters, including the (R)-PGME: 6-APA ratio, when compared to the 1P1S system. One challenge for the cascade syntheses is that the ratio of moles of (R)-PGME consumed per moles of AMP at the maximum product concentration is elevated when compared to the ratio of the direct synthesis from 6-APA and (R)-PGME. For the 1P1S system this ratio is ~7.5 and for the 1P2S system this ratio is ~6. The direct syntheses with iPGA or AEH are <2 and ~4, respectively.
3.3 Conclusions

3.3.1 AEH Synthesis

The AEH from *X. campestris* *pv. campestris* strain ATCC 33913 displayed improved synthetic properties compared to AEH from *X. citri* and possibly also to AEH from *A. turbidans*, as evidenced by its lower $\alpha$-values. Furthermore, the AEH $\beta_0$ value is clearly favorable for synthesis compared to the PGA derived enzymes. The $\gamma$ values for AEH are comparable to the WT PGA but are significantly lower than mutated enzymes. The enzyme’s unique ability to catalyze the $\beta$-lactam synthesis with minimal secondary hydrolysis in purely aqueous solutions renders this enzyme an interesting target for $\beta$-lactam synthesis. The most commonly industrially used enzyme currently employed for this reaction, penicillin G acylase (E.C. 3.5.1.11), exhibits significantly more secondary hydrolysis, unless it is reacted in a biphasic system or highly concentrated solutions.

3.3.2 Cascade synthesis

We have demonstrated the first purely aqueous cascade system toward AMP using a two-enzyme system with both AEH and iPGA. The 1P1S scheme and the 1P2S scheme resulted in optimum AMP yields of 39 and 46%, respectively. At such conditions, the 1P1S configuration requires 7.5 moles of (R)-PGME per moles of AMP at the maximum product concentration, compared to only 6.2 for the 1P2S scheme. Maximum conversions were achieved in one to two hours, significantly reducing the reaction times previously observed in the systems that use iPGA and ethylene glycol [92,93]. In all cases, the two-enzyme system with iPGA and AEH out-performed the systems that used only iPGA, thus demonstrating the clear advantage of using AEH. While the 1P1S system resulted in slightly lower yields, it could be advantageous due to its operational ease and faster cycle times. In the 1P2S system, higher
conversion was achieved and secondary hydrolysis was minimized by adjusting the relative enzyme loadings. These reaction schemes could be scaled up and incorporated with enzyme reuse, which has been previously demonstrated for iPDA [121]. However, further optimization is still required to improve yields and reduce ester usage for these processes.

3.4 Materials and Methods

3.4.1 Materials

6-Aminopenicillinic acid (6-APA), (R)-phenylglycine ((R)-PG), ampicillin, (R)-phenylglycine methyl ester hydrochloride ((R)-PGME), penicillin G (penG), phenylacetic acid (PAA), and Eupergit-immobilized penicillin G acylase (iPGA) from *Escherichia coli* all were procured from Sigma Aldrich (St. Louis, MO). Soluble amino ester hydrolase (AEH) from *X. campestris pv. campestris* was prepared in our lab as described in Chapter II.

3.4.2 AEH Synthesis Reactions from (R)-PGME and 6-APA

Synthesis of (R)-ampicillin and (S)-ampicillin was conducted at 25°C with 20–60 mM (R)-PGME and 20–60 mM 6-APA in 100 mM phosphate buffer pH 6.2. A racemic mixture of 90 mM (R/S)-phenylglycine methyl ester hydrochloride was prepared to determine the enantioselectivity of the reaction. Lastly, synthesis of cephalexin was performed at 25°C with 60 mM (R)-PGME and 20 mM 7-ADCA. Before analysis the samples were quenched and diluted 10-fold by the addition of HPLC eluent. The analysis was conducted with HPLC.

3.4.3 One-Pot, One Step Synthesis (1P1S)

15 mL of 20 mM of penG and 60 mM (R)-PGME in 100 mM phosphate buffer (pH 7) were added to a round bottom flask along with iPDA or iPDA and purified *X. campestris pv. campestris* AEH, per Table 3.2. The reactions were stirred using a
magnetic stir plate and carried out at room temperature (22°C-25°C).

3.4.4 One-Pot, Two-Step Synthesis (1P2S)

7.5 mL of 40 mM penG in 100 mM phosphate buffer (pH 7) was added to a round bottom flask along with iPGA per Table 3.3, 124 UPenG per gram of carrier, where 1 UPenG is defined as one μmol of penicillin G hydrolyzed per minute. The reactions were stirred using a magnetic stir plate and carried out at room temperature (22°C-25°C). After the reaction reached near completion, as determined by HPLC, 7.5 mL of 120 mM (R)-PGME was added. The pH was adjusted with NaOH from ~6.4 to 7.0 and X. campestris pv. campestris AEH was added per Table 3.3, 79 UAmp·mg⁻¹ protein, where UAmp is defined as one μmol of AMP hydrolyzed per minute under saturation conditions. Additional experiments were conducted where pH was controlled between 7.0 ± 0.1, the pH control had no effect on the results of the experiment. In reactions where iPGA was used in both steps, we replaced the AEH with equivalent AMP synthesis units of iPGA based on initial synthesis rate data from 6-APA and (R)-PGME using only AEH and only iPGA where 1 UAmp of AEH ~1 UAmp of iPGA ~6.8 UPenG of iPGA.

3.4.5 One-Pot, Two-Step, Two-Stage Synthesis (1P2S-2S)

These experiments were conducted analogously to the 1P2S schemes, with the exception that after the completion of the first step the iPGA was removed from the reaction using filtration.

3.4.6 High Pressure Liquid Chromatography (HPLC) Assay

All analysis was conducted using high performance liquid chromatography complete with a Shimadzu-LC-20AT pump, Beckman Coulter Ultrasphere ODS 4.6 mm x 25 cm column, and SPD-M20A prominence diode array detector (PDA) monitored at 215 nm. Samples (100 μL) were diluted 10X into 900 μL of HPLC quench buffer (75%
Methanol, 25% 0.02 M potassium phosphate, pH 6.0). 2 µL of sample was loaded onto the column. The mobile phase operated under a gradient at 1.0 mL-min\(^{-1}\) and contains 20% methanol and 80% 0.02 phosphate buffer (pH 7) and increased to 50% methanol over a 25 min period. All components, (R)-PG, (R)-PG, 6-APA, (R)-PGME, SPGME, (R)-ampicillin, (S)-ampicillin, analyzed on the HPLC with >95% mass balance closure. The (R)-ampicillin and (S)-ampicillin diastereomers can be separated on the HPLC, however, the stereoisomers of RPG/SPG and (R)-PGME/(S)-PGME co-eluted. Enantiopurity of the product (ee\(_{p}\)) was determined using the (R)-Amp and (S)-Amp concentrations.

3.4.7 Publication Information

The work presented in Chapter III of this dissertation on the synthesis using AEH alone is partially published in the Journal of Molecular Catalysis B - Enzymatic (Vol. 67, Issue 1-2, pp 21-28, 2010) with my thesis advisor, Andreas Bommarius as a co-author. The initial date of online publication was July 6, 2010. The work presented on the one-pot synthesis of ampicillin using both AEH and PGA is published in ChemCatChem (Vol. 2, Issue 8, pp 987-991) with Andria Deagüero, Carolina Perez, and Andreas Bommarius as co-authors. The initial date of online publication was July 28, 2010.
CHAPTER IV

IMPROVED THERMOSTABILITY OF AEH USING A MODIFIED STRUCTURE-GUIDED CONSENSUS METHOD

4.1 Introduction

α-amino ester hydrolase (AEH) enzymes catalyze the synthesis and hydrolysis of α-amino β-lactam antibiotics, Fig. 2.1. These enzymes are alternatives to the more commonly used penicillin G acylase (PGA) [46, 49–52, 57, 60]. AEHs are particularly interesting for use in the one-pot synthesis of the semi-synthetic antibiotics ampicillin directly from Penicillin G, thus eliminating the isolation step of the intermediate 6-aminopenicillanic acid (6-APA), as described in Chapter III [56]. AEHs from three organisms (Acetobacter turbidans, Xanthomonas citri, and Xanthomonas campestris pv. campestris (this work)) have been isolated and cloned in E.coli [46, 49, 50, 57, 60]. Additionally, crystal structures for the AEHs from X. citri and A. turbidans have been solved. In comparison to the penicillin G acylases ($T_{opt} = 49^\circ C$ [128], $T_m = 64.5^\circ C$ [129]) these enzymes have low kinetic thermostability with optimal temperature ranging from $25^\circ C – 30^\circ C$ and observed half-life of 5 minutes at $30^\circ C$ [57].

To improve the industrial relevance of these enzymes the thermostability needs improvement [55–57, 130]. Improvement of the enzyme’s thermostability results in increased catalytic lifetime or total turnover number (TTN). Furthermore, improved thermostability has been linked to improved evolvability [89, 94]. Improved evolvability increases the applicability of the enzyme for protein engineering protocols designed to alter its reaction properties as was demonstrated in the ‘neutral’ drift evolution of TEM-1 β-lactamase toward the novel substrate cefotaxime [89]. Thus, it is desirable
to improve the thermostability of the AEH.

AEHs are composed of four identical 614 amino acid subunits that form a dimer of dimers. There are 81 residues that make up the subunit interfaces, Appendix F. The N-terminus (resi 24 - 42) of each subunit is buried in its adjacent subunit (A to C and B to D, and vice versa) resulting in many interactions at these positions. Additionally, AEHs have multiple intra-subunit interactions as each subunit is composed of three domains. The first domain, domain I (1-197, 285-410) contains the catalytic serine and $\alpha/\beta$ canonical fold. The cap domain II (198-284) contains mostly helixes and loop in a jellyroll fold, while domain III (411-637) makes up the C-terminal (Fig. 4.4A). Stabilization of domain I to domain II increased the thermostability of the structurally homologous *Rhodacoccus sp.* cocaine esterase 30-fold [131,132]. Lastly, the protein contains one inter-subunit disulfide bridge between C408 and C412. The monomer-monomer, dimer-dimer, and the intra-subunit interactions are of interest to best stabilize this protein.

Several different protein engineering techniques have been successfully used to improve protein thermostability, including error-prone PCR (epPCR) [65], DNA Shuffling via SCHEMA [66–69], temperature value (B-FIT) approach [70] and the structure-guided consensus concept [33,71] (SGC). SCHEMA, B-FIT and SCG also required structural information obtained from the crystal structure as described in Chapter I, Section 1.4.1. In this work, we focused on combining aspects of both the B-FIT and SGC approaches to develop a more comprehensive stabilization strategy. As the B-FIT and SGC approaches have been successful in stabilizing proteins, we were interested in developing a stabilization protocol that incorporates the information from both the approaches, thus, resulting in a modified structure-guided consensus (MSGC) approach (Fig. 4.1). The B-value, or temperature value, is determined for each atom in a protein structure during high-resolution X-ray crystallography and has been linked to increased amino acid flexibility [133]. The average B-value for an amino
acid is calculated by averaging the B-values for each atom of an amino acid position, this can be calculated using the B-FITTER program (http://www.mpi-muelheim.mpg.de/kofo/institut/arbeitsbereiche/reetz(evolution-tools.zip). [74, 134] The incorporation of B-values is a straightforward addition as the B-values are another piece of structural data that can be incorporated in the heuristic rules that were previously determined by Polizzi, et al. [33]. In the standard B-FIT approach, residues with the top twenty average B-values in the top 20 of all B-values as determined by the B-FITTER program are targeted.

The positions targeted by the B-FIT tend to be biased toward Lys, Arg, and Glu; and are usually located in loops and on the enzyme surface [74, 133, 134]. While the B-value approach does identify target residues, it does not predict what to mutate that residue to; instead NNK or similar libraries are introduced into the site(s) and evaluated iteratively. [70] Typically, this protocol results in four to five individual combinatorial active site saturation testing (CAST) [70] libraries, each with sizes of 20-8,000(20^3) amino acid mutations. Depending on how many sites are in each library the resulting screening requirements can be on the order of 10^5 per library. Alternatively, restricted libraries can be utilized at these positions to reduce screening requirements to \( \leq 10^4 \) per library [134]. In SGC, residues are targeted based on mutation toward the consensus residues (usually >50% consensus) as identified by alignment with functional homologous sequences along with the structural guidelines defined in Vazquez-Figueroa, et al. [71] (Fig. 4.1). The advantage of SGC was that this approach has low screening requirements; \( \sim 20-40 \) site-directed mutations per protein. Furthermore, Vazquez-Figueroa demonstrated that SGC was successfully able to target positions at the subunit interfaces, which drastically increased the thermostability of the enzyme [71]. There is little overlap in the residues selected for mutation when using the two separate approaches, see Chapter II. It is desired to determine if these two approaches can be synergistically applied to stabilize proteins.
Figure 4.1: Modified structure guided consensus (MSGC) decision tree. The questions in the dashed box are the heuristics applied by Vazquez et al. in the thermostabilization of glucose dehydrogenase. [71]
MSCG includes both positions with high degree of consensus and positions in the top 20 average B-values as described by Fig. 4.1. As an initial target, amino acid positions with high average B-values were replaced with the consensus residue at those positions; thus, maintaining a small rational library. Alternatively, high B-value positions were also individually NNK saturated and screened.

In this work, we apply a MSGC approach to improve the thermostability of the AEH from X. *campestris pv. campestris*. Four rounds of protein engineering were completed to improve the thermostability of AEH. The first round included all of the single variants identified by the MSCG approach. This round was followed by evaluation of double, triple, quadruple and a sextuple variants of the first round variant(s). In the third round, two high average B-value positions were targeted for NNK randomization. Lastly in the fourth round, the impact of an intersubunit disulfide-bridge was investigated.

4.2 Results and Discussion

4.2.1 Round 1: Single Variants

The genomic sequence from the X. *campestris pv. campestris* AEH (AEH_Xcc) was aligned with seven homologous sequences that included the known AEHs from *Aerotobacter turbidans* (A.turbidans AEH) and *X.citri* (AEH_Xcitri) in addition to the structurally homologous *Rhodococcus sp*. cocaine esterase (RhCocE) [96], *Brevibacillus laterosporus* glutaryl 7-ACA acylase (J1 acylase) [97], putative *Acidobacterium capsulatum* AEH(pAEH_ATCC 51196), putative *Gluconobacter oxydans* AEH (pAEH_- goxidans), lastly a *Stenotrophomonas maltophilia R551-3* X-Pro peptidase. These sequences range from 23% to 93% identical on an amino acid level, Table 4.1. The sequences are all members of the α/β hydrolase fold family [96, 97]. The consensus sequence was determined with a 37% absolute consensus as the cutoff, Appendix D.

The subunit averaged B-values were determined by averaging the average B-values
calculated using B-FITTER for each of the four AEH subunits in the *X. citri* crystal structure.pdb 1MPX, Appendix D. The overall subunit average B-value for this protein was 21.0. For the high B-value residues, we selected the consensus residue for mutation, if there was a consensus residue that was different than in the wild-type template. Positions 143, 435, and 436 were identified as interesting targets to include in our MSGC because they featured high average B-values of 39.3, 44.6, and 48.6 and consensus residues differed from the wild-type residue. Not surprisingly, given that B-FIT method targets highly flexible positions, we observed that the majority of the top 20 B-value positions were either weakly conserved (~37% of the residues were conserved) or had no clearly conserved residue according to our alignment.

The eighteen first-round mutations (Table 4.2) were carried out with the Quik-Change® site-directed mutagenesis protocol and were confirmed by sequencing. Of the eighteen residues selected for MSGC, three residues (A435G, A436G, and E143K) were in the list of the top 20 average B-values, the other high B-value sites contained the consensus residue in the wild-type sequence, thus these were not mutated in the first round of evolution at these sites. These variantss were expressed in

**Table 4.1**: Percent amino acid identity matrix of homologous proteins to AEH from *X. campestris pv. campestris*

<table>
<thead>
<tr>
<th></th>
<th>AEH_Xcc</th>
<th>AEH_Xcitri</th>
<th>X-pro peptidase</th>
<th>A. turbidans AEH</th>
<th>pAEH_goxidans</th>
<th>pAEH_ATCC51196</th>
<th>RhCoCE</th>
<th>J1 acylase</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEH_Xcc</td>
<td>100</td>
<td>93</td>
<td>77</td>
<td>58</td>
<td>56</td>
<td>40</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td>AEH_Xcitri</td>
<td>93</td>
<td>100</td>
<td>77</td>
<td>59</td>
<td>55</td>
<td>40</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td>X-pro peptidase</td>
<td>77</td>
<td>77</td>
<td>100</td>
<td>58</td>
<td>57</td>
<td>40</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>A. turbidans AEH</td>
<td>58</td>
<td>59</td>
<td>58</td>
<td>100</td>
<td>68</td>
<td>40</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>pAEH_goxidans</td>
<td>56</td>
<td>55</td>
<td>57</td>
<td>68</td>
<td>100</td>
<td>38</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>pAEH_ATCC51196</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>38</td>
<td>100</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>RhCoCE</td>
<td>27</td>
<td>27</td>
<td>23</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>100</td>
<td>35</td>
</tr>
<tr>
<td>J1 acylase</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>35</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 4.2: MSGC Round 1: List of site-directed mutations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>143</td>
<td>E</td>
<td>K</td>
<td>37%</td>
<td>L</td>
<td>39.3[^c]</td>
<td>26.8</td>
<td>1.0</td>
</tr>
<tr>
<td>186</td>
<td>N</td>
<td>D</td>
<td>50%</td>
<td>L</td>
<td>27.7</td>
<td>28.7</td>
<td>0.8</td>
</tr>
<tr>
<td>275</td>
<td>A</td>
<td>P</td>
<td>75%</td>
<td>L</td>
<td>20.1</td>
<td>28.6</td>
<td>1.4</td>
</tr>
<tr>
<td>322</td>
<td>E</td>
<td>K[^d]</td>
<td>37%</td>
<td>H</td>
<td>20.2</td>
<td>n.d.</td>
<td>n.d</td>
</tr>
<tr>
<td>377</td>
<td>V</td>
<td>K[^d]</td>
<td>50%</td>
<td>H</td>
<td>24.1</td>
<td>33.3</td>
<td>0.1</td>
</tr>
<tr>
<td>416</td>
<td>S</td>
<td>L</td>
<td>50%</td>
<td>L</td>
<td>28.3</td>
<td>n.d</td>
<td>n.d</td>
</tr>
<tr>
<td>426</td>
<td>K</td>
<td>G</td>
<td>63%</td>
<td>S</td>
<td>28.3</td>
<td>33.6</td>
<td>0.3</td>
</tr>
<tr>
<td>435</td>
<td>A</td>
<td>G</td>
<td>25%</td>
<td>L</td>
<td>44.6[^e]</td>
<td>28.1</td>
<td>0.5</td>
</tr>
<tr>
<td>436</td>
<td>Q</td>
<td>G</td>
<td>37%</td>
<td>L</td>
<td>48.6[^e]</td>
<td>31.4</td>
<td>1.0</td>
</tr>
<tr>
<td>485</td>
<td>F</td>
<td>Y</td>
<td>50%</td>
<td>S</td>
<td>17.5</td>
<td>n.d</td>
<td>n.d</td>
</tr>
<tr>
<td>489</td>
<td>P</td>
<td>V</td>
<td>50%</td>
<td>L</td>
<td>23.2</td>
<td>25.8</td>
<td>0.4</td>
</tr>
<tr>
<td>496</td>
<td>I</td>
<td>V</td>
<td>50%</td>
<td>S</td>
<td>18.9</td>
<td>n.d</td>
<td>n.d</td>
</tr>
<tr>
<td>502</td>
<td>V</td>
<td>A</td>
<td>50%</td>
<td>S</td>
<td>19.7</td>
<td>n.d</td>
<td>0.1</td>
</tr>
<tr>
<td>567</td>
<td>F</td>
<td>Y</td>
<td>50%</td>
<td>S</td>
<td>19.5</td>
<td>n.d</td>
<td>n.d</td>
</tr>
<tr>
<td>577</td>
<td>T</td>
<td>V</td>
<td>50%</td>
<td>S</td>
<td>17.2</td>
<td>n.d</td>
<td>n.d</td>
</tr>
<tr>
<td>580</td>
<td>P</td>
<td>K</td>
<td>50%</td>
<td>L</td>
<td>20.5</td>
<td>n.d</td>
<td>n.d</td>
</tr>
<tr>
<td>584</td>
<td>V</td>
<td>I</td>
<td>63%</td>
<td>S</td>
<td>16.1</td>
<td>34.0</td>
<td>0.4</td>
</tr>
<tr>
<td>622</td>
<td>V</td>
<td>I</td>
<td>63%</td>
<td>S</td>
<td>18.7</td>
<td>34.4</td>
<td>0.6</td>
</tr>
</tbody>
</table>

[^a] = L=Loop, H = α helix, S = β sheet
[^b] calculated using B-FITTER
[^c] indicates a B-value in the Top 20
[^d] Lys(K) is a known helix stabilizer
[^e] RA is the relative initial ampicillin activity measured at 25°C
n.d. not detected

Figure 4.2: Experimental design of T[^30] experiments
Figure 4.3: Results of round 1 – 4 variants with Relative Activity values in black (left axis) and $T_{50}^{30}$ values in gray (right axis). The dashed reference line the wild-type $T_{50}^{30}$ value.

BL21(DE3)pLysS cells as confirmed by SDS-PAGE, and purified using Ni-NTA affinity chromatography, see Materials and Methods.

The purified variants were then analyzed for thermostability by evaluating their $T_{50}^{30}$ value, the temperature at which the observed half-life is 30 minutes (Fig. 4.2). The $T_{50}^{30}$ value of the wild-type protein is 27±2°C. In the first-round, we obtained nine variants with improved or equivalent thermostability to the wild-type, a success rate of 50% (Fig. 4.3, Tab. 4.2). Additionally, there were seven variants that had no activity; hence, it is assumed that the enzyme, while expressed as evidenced by SDS-PAGE, did not fold correctly. The best single variant V622I had the highest $T_{50}^{30}$ of 34°C, resulting in a $\Delta T_{50}^{30}$ of 7°C compared to wild-type, though the specific activity was reduced by half. This type of improvement in $T_{50}^{30}$ values was comparable
to first-round improvements observed using the CASTing approach [74], which has a much higher screening requirement. The overall first-round hit rate of \(~50\%\) is excellent for a molecular biology protocol and is comparable to hit rates observed with the SGC approach for thermostabilizations [33, 71].

As in the 2006 work by Vazquez-Figueroa, the effect of the mutation on the specific activity is quite variable (Fig. 4.3). With the exception of A275P, which showed of 1.4-fold improved specific activity, all other single variants negatively affected the activity. The mutation A275P occurs on a loop between two helixes, thus we hypothesized that proline at this position is better than alanine in making this turn; proline is a better helix breaker than alanine. Additionally, A275P is located at the interface between domains II and III. The slight increase in bulk by proline compared to alanine may improve the van der Waals interactions between residue 275 in domain II and the domain III residues Y549, N606, and F608 (Fig. 4.4B). Stabilization of domain II has been previously shown to be important in the overall stabilization of cocaine esterase [131, 132]. Otherwise, trends in the location of the residues that resulted in first-round hits are not evident; they are evenly distributed across of the protein structure and various secondary structure elements (Fig. 4.4A).

In the first-round, the selected residues with high B-values did not show any significant improvement in thermostability, thus indicating that mutating toward the consensus residue at these positions was not adequate to improve the thermostability of the protein. However, the mutations at these positions only had minimal impact on the activity. This will be further addressed in the Round three mutations. To investigate relationships between consensus, average B-value, specific activity, and thermostability, the covariances were calculated (Table 4.3). The covariances indicated that degree of consensus did not correlate with either \(T_{50}^{30}\) or relative activity; however, the B-value was slightly positively correlated to relative activity, but was
Figure 4.4: A.) Locations of stabilizing residues shown with red spheres. Single subunit of AEH protein colored by domain, α/β-hydrolase fold domain (I), grey, cap-domain (II), green, and C-terminal Jelly Roll domain (III), magenta. B.) Model of the A275P mutation in the X. campestris pv. campestris AEH protein. Position of the proline is based on A. turbidans, pdb 2B4K and X. citri, pdb 1MPX, homology models. P306 A. turbidans is homologous to A275. A275P (yellow) is at the interface of domain II (residues 275,204,210) and domain III (residues 584,606,608). The dashed lines show all interactions within 4 Å of A275P, the A275P mutation adds two additional interactions between the γ and δ carbons of the P275 and N606.
negatively correlated with $\Delta T_{50}^{30}$. Therefore, mutating towards the high B-value position’s consensus residue was not sufficient to improve enzyme thermostability.

### 4.2.2 Round 2: Multiple Variants

To test for property additivity, combinations of first round variants were tested, results shown in Fig. 4.3. We analyzed the double variants A275P/N186D, A275P/V622I, N186D/V622I, triple variant A275P/N186D/V622I and the sextuple variant N186D/-A275P/V377K/K426G/V584I/V622I. The sextuple variant includes the top five most stabilizing single mutations in addition to A275P. The multi-variants were purified and analyzed for both the $T_{50}^{30}$ and relative activity as described previously, the results are in Table 4.4.

In all cases, the activity recovered when combining the thermostabilizing variant with A275P. However, the most active variant was N186D/V622I, which had a 1.6-fold increase in specific activity. The triple variant (TM), N186D/A275P/V622I, had a specific activity of 80 U.mg$^{-1}$ and $T_{50}^{30}$ of 32.0°C resulting in a $\Delta T_{50}^{30}$ of 5.2°C. The triple variant was selected as the best variant from the second round because additional mutations to the variant did not significantly improve the either the $T_{50}^{30}$ or specific activity, suggesting that the additional mutations did not contribute to overall stabilizing network of the triple variant.

### 4.2.3 Round 3: NNK libraries

As the substitution of the consensus residue into the high B-value residues did not result in significant improvements in thermostability, two locations on the structure
were identified for individual site-saturation (NNK) libraries. Positions 34 and 143 were selected for randomization into the consensus triple variant (TM), N186D/A275P/V622I, from the second round. Position 34 featured the highest average B-value (49.5) in the AEH structure and is also involved in the N-terminus subunit – subunit interactions. Lysine (K34) is the consensus residue at this position, albeit low at only 37% but is represented in the wild-type. Another top 20 B-value position, E143 (B-value: 39.3) was evaluated in the Round one library (E143K). The E143K variant of Round one had similar activity and thermostability as the wild-type enzyme. Two libraries, K34NNK-TM and E143NNK-TM, were created to determine the best amino acid for each of these positions.

### 4.2.3.1 High Throughput Screening Assay

To evaluate the NNK libraries a new screening method was required. To achieve full library coverage, 95 colonies would need to be screened, as determined according

---

**Table 4.4: Results from Rounds 2–4**

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Round</th>
<th>T$_{50}^{M}[^{\circ}C]$</th>
<th>Residual Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>-</td>
<td>26.8</td>
<td>1.0</td>
</tr>
<tr>
<td>A275P/N186D</td>
<td>2</td>
<td>33.5</td>
<td>0.7</td>
</tr>
<tr>
<td>A275P/V622I</td>
<td>2</td>
<td>31.6</td>
<td>1.0</td>
</tr>
<tr>
<td>N186D/V622I</td>
<td>2</td>
<td>28.1</td>
<td>1.6</td>
</tr>
<tr>
<td>N186D/A275P/V622I</td>
<td>2</td>
<td>32.0</td>
<td>1.2</td>
</tr>
<tr>
<td>N186D/A275P/V622I/V584I</td>
<td>2</td>
<td>32.9</td>
<td>0.6</td>
</tr>
<tr>
<td>N186D/A275P/V622I/V76I</td>
<td>2</td>
<td>29.9</td>
<td>0.3</td>
</tr>
<tr>
<td>N186D/A275P/V622I/V607I</td>
<td>2</td>
<td>31.5</td>
<td>1.1</td>
</tr>
<tr>
<td>N186D/A275P/V622I/V377K/K426G/V584I</td>
<td>2</td>
<td>32.3</td>
<td>1.2</td>
</tr>
<tr>
<td>N186D/A275P/V622I/K34L</td>
<td>3</td>
<td>30.8</td>
<td>0.6</td>
</tr>
<tr>
<td>N186D/A275P/V622I/K34C</td>
<td>3</td>
<td>31.3</td>
<td>0.7</td>
</tr>
<tr>
<td>N186D/A275P/V622I/E143H</td>
<td>3</td>
<td>34.7</td>
<td>1.3</td>
</tr>
<tr>
<td>N186D/A275P/V622I/E143G</td>
<td>3</td>
<td>33.1</td>
<td>1.7</td>
</tr>
<tr>
<td>N186D/A275P/V622I/G357C</td>
<td>4</td>
<td>37.0</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Figure 4.5: p-dimethyl-amino benzaldehyde (PDAB) reacts with 6-APA to form a chromogenic product.

Figure 4.6: Experimental design of small-scale T_{30}^{50} experiments; a control experiment with non-incubated protein was conducted to calculate residual activity (not shown)
to $N = [\ln(1 - P)]/[\ln(1 - \frac{1}{n})]$, which determines the number of colonies ($N$) that have to be picked to have a 95% probability ($P$) of selecting one of each codon ($n$) is $\frac{1}{32}$. A 96-well plate assay was developed to monitor both the non-incubated initial activity and residual activity after a short incubation between 27-34°C (Fig. 4.6) as described in the Materials and Methods section. Cell lysate from expression cultures was used to hydrolyze ampicillin over 30 minutes and conversion was measured via a colorimetric assay utilizing para-dimethyl amino benzaldehyde (PDAB). In the chromogenic assay, PDAB forms a Schiff base, which absorbs at 405 nm with a yellow color, when reacted with the newly formed amino group in the hydrolytic product, 6-APA (Fig. 4.5) [135]. The non-incubated activity (NIA) was measured for each well by measuring the absorbance after the ampicillin reaction using enzyme that was not exposed to thermal stress over that of the background reaction with empty pET28a vector expressed cell lysate (eq. 4.1). Similarly the incubated activity (IA) was measured after thermal incubation for 15 min. Residual activity (RA) values were then determined by the ratio of IA to NIA equation 4.3. The variants were then ranked according to their residual activity (RA) multiplied by the absorbance of the incubated wells over background (RAI), equation 4.4 and Appendix G. The RAI ranking eliminated the variants that have high RAs but were only negligibly active following incubation. The 96-well plate assay can evaluate variants rapidly compared to using HPLC.

\[
NIA = (A_{405\text{variant}} - A_{405\text{background}})_{\text{not incubated}} \tag{4.1}
\]

\[
IA = (A_{405\text{variant}} - A_{405\text{background}})_{\text{incubated}} \tag{4.2}
\]

\[
RA = \frac{IA}{NIA} \tag{4.3}
\]

\[
RAI = RA \times IA \tag{4.4}
\]
Figure 4.7: Screening results for libraries K34NNK-TM (A) and E143NNK-TM (B) when incubated at 34°C; the shaded area is the region containing the hits. The triple variant has an NIA of 0.17 AU and RA of 0.3 AU. Mutations K34C and E143G, E143H, E143I are in this region. Other mutations for the K34NNK library were selected at lower temperatures.

4.2.3.2 Library Results

Both the K34NNK-TM and E143NNK-TM libraries featured good diversity, as indicated by the sequencing of non-selected colonies and sequencing of the library plasmid mix after PCR (Appendix G). The K34NNK-TM library was screened at incubation temperatures of 29°C, 32°C, and 34°C. Sequencing of the top nine hits from the K34 library resulted in the variants K34C (2), K34L (3), K34W (1), and K34K(3). At 34°C, only one variant, K34C, was significantly active as indicated by the assay(Fig. 4.7). K34C and K34L were purified for analysis of the $T_{50}^{30}$ and activity. Both K34C and K34L had $T_{50}^{30}$ values around 32°C; thus, they were as stable the triple variant template. Interestingly, the K34C variant impacted the tertiary structure; an aggregate of ~100 kDa was identified by SDS-PAGE analysis when run without β-mercaptoethanol unlike with the triple variant (TM) as shown in Fig. 4.8. The E143NNK-TM library was screened at 34°C because the results from the K34 library indicated that 29 and 32°C were not stringent enough to find improved variants compared to the triple variant. The top five hits were sequenced, resulting in E143H, E143I, E143G, and E143E(2) (Fig. 4.7). Purified variants E143H and E143G
Figure 4.8: SDS-PAGE analysis of WT(1), K34L(2), K34C(3), and TM(7) reduced using β-mercaptoethanol and WT(4), K34L(5), K34C(6), and TM(8) not reduced. The protein marker (L) ranges between 25-250 kDa, the single subunit of AEH is observed at 68 kDa. The K34C variant formed aggregates when not reduced indicating that this mutation altered the tertiary and/or quaternary structures.

yielded $T_{30}^{50}$ values of 34.7°C and 33.1°C, respectively. Thus, the quadruple variant E143H/N186D/A275P/V622I showed an 8°C increase in $T_{30}^{50}$ over the wild-type and, in addition, a specific activity of 83 U·mg$^{-1}$, which is 1.3-fold the wild-type activity.

One of the libraries selected by B-FIT analysis resulted in an improved enzyme compared to the template triple variant enzyme. In the K34 library, the wild-type and consensus residue, K, was screened as the best variant. However, at position E143, the consensus residue, K, was not one of the improved enzymes; thus, the “hit” could not have been predicted from the consensus analysis alone. Therefore, the NNK library analysis was advantageous at the high B-value positions.

4.2.4 Round 4: Introduction of a disulfide-bond: G357C

Subunit interactions have been previously been shown to have large influence on thermostability [108,136,137]. A disulfide-bond was introduced into the X. campestris pv. campestris AEH by mutating G357 of the triple variant to cysteine creating the quadruple variant, N186D/A275P/G357C/V622I (G357C-TM). G357 is located at the interface of the A and B chains of the AEH protein; the residues of both chains
**Figure 4.9:** Design of disulfide bridge between chain A and chain B of AEH at position G357 (pdb 1MPX). The $\alpha$ carbons of the G357 are 4.6 Å apart, as indicated with the dashed line. The wild-type disulfide bridge between C408 and C412 is shown in spheres and subunit interacting residues within 5 Å are shown in blue (chain A) and red (chain B).

are just $\sim$4 Å apart. This position was considered to be ideal for the creation of a disulfide bond because it required the change of just one amino acid and the mutation should not interfere with the enzyme’s secondary structure, as G357 is located in a loop. The single mutation should create two disulfide bridges, between the chains A and B, and similarly one between chains C and D, (Fig. 4.9); thus covalently binding the dimers and restricting dimer dissociation.

To investigate formation of the disulfide bond, SDS-PAGE analysis of the expressed and purified G357C-TM variants was completed under both reducing and non-reducing conditions. The additional cysteine impacted the quaternary structure because bands indicating increased molecular mass between $\sim$120-200 kDa were observed on the non-reduced lanes of the cysteine variant, while no aggregates are observed on the triple variant controls (Fig 4.10). The G357C-TM variant was characterized for both initial activity and $T_{50}$. The initial activity of the variant was
reduced 5-fold to a specific activity of 14 U·mg$^{-1}$; however, the $T_{50}^{30}$ was increased to 37.0°C.

### 4.2.5 Total Turnover Number

All variants were analyzed for total turnover number (TTN) which can be approximated for first-order deactivation by the ratio of $k_{cat,obs}/k_{d,obs}$ at 25°C [112]. The deactivation constant $k_{d,obs}$ was approximated based on first-order deactivation at 25°C. This data is obtainable by determining the residual activity at 25°C from the raw data used to obtain the $T_{50}^{30}$ values as shown in Table 4.5 and Fig. 4.11. While the kinetic deactivation mechanism of AEH is still not clear (Chapter II), this approach allows us to easily compare the variants on both activity and thermostability.

Surprisingly, we again observed no indication of mutation additivity or synergy with respect to thermostability. Additivity or synergy has been observed applying similar thermostabilization protocols on other proteins such as *Bacillus subtilis* Lipase A [74] and glucose dehydrogenase (GDH) [71], as discussed in Chapter I. The expected additivity compared to the actual temperature improvement is shown in Table 4.6.
### Table 4.5: Total Turnover Number Calculation

<table>
<thead>
<tr>
<th>Mutation</th>
<th>RA after 30 min at 25°C</th>
<th>$k_{cat,obs}$ [s$^{-1}$]</th>
<th>$k_{cat,obs}$ [s$^{-1}$]</th>
<th>Estimated TTN</th>
<th>Fold Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>0.65</td>
<td>2.1E-04</td>
<td>72.5</td>
<td>3.0E+05</td>
<td>1.0</td>
</tr>
<tr>
<td>E143K</td>
<td>0.72</td>
<td>1.8E-04</td>
<td>55.6</td>
<td>3.0E+05</td>
<td>1.0</td>
</tr>
<tr>
<td>N186D</td>
<td>0.81</td>
<td>1.2E-04</td>
<td>30.5</td>
<td>2.6E+05</td>
<td>0.9</td>
</tr>
<tr>
<td>A275P</td>
<td>0.75</td>
<td>1.6E-04</td>
<td>100.5</td>
<td>6.3E+05</td>
<td>2.1</td>
</tr>
<tr>
<td>V377K</td>
<td>0.96</td>
<td>2.1E-04</td>
<td>8.0</td>
<td>3.8E+05</td>
<td>1.3</td>
</tr>
<tr>
<td>K426G</td>
<td>0.75</td>
<td>1.6E-04</td>
<td>21.3</td>
<td>1.3E+05</td>
<td>0.4</td>
</tr>
<tr>
<td>A435G</td>
<td>0.69</td>
<td>2.0E-04</td>
<td>38.2</td>
<td>1.9E+05</td>
<td>0.6</td>
</tr>
<tr>
<td>Q436G</td>
<td>0.85</td>
<td>9.1E-05</td>
<td>69.9</td>
<td>7.7E+05</td>
<td>2.6</td>
</tr>
<tr>
<td>T577V</td>
<td>0.48</td>
<td>4.1E-04</td>
<td>71.1</td>
<td>1.7E+05</td>
<td>0.6</td>
</tr>
<tr>
<td>V584I</td>
<td>0.89</td>
<td>6.2E-05</td>
<td>30.3</td>
<td>4.9E+05</td>
<td>1.6</td>
</tr>
<tr>
<td>V622I</td>
<td>0.98</td>
<td>9.8E-06</td>
<td>38.4</td>
<td>3.9E+06</td>
<td>13.1</td>
</tr>
<tr>
<td>A275P/N186D</td>
<td>0.94</td>
<td>3.3E-05</td>
<td>47.7</td>
<td>1.4E+06</td>
<td>4.8</td>
</tr>
<tr>
<td>A275P/V622I</td>
<td>0.88</td>
<td>7.0E-05</td>
<td>73.9</td>
<td>1.1E+06</td>
<td>3.5</td>
</tr>
<tr>
<td>N186D/V622I</td>
<td>0.67</td>
<td>2.2E-04</td>
<td>114.7</td>
<td>5.2E+05</td>
<td>1.8</td>
</tr>
<tr>
<td>N186D/A275P/V622I</td>
<td>0.93</td>
<td>4.1E-05</td>
<td>87.0</td>
<td>2.1E+06</td>
<td>7.1</td>
</tr>
<tr>
<td>N186D/A275P/V622I/V584I</td>
<td>0.86</td>
<td>8.4E-05</td>
<td>43.5</td>
<td>5.2E+05</td>
<td>1.7</td>
</tr>
<tr>
<td>N186D/A275P/V622I/V377K/K426G/V584I</td>
<td>0.88</td>
<td>7.1E-05</td>
<td>87.0</td>
<td>1.2E+06</td>
<td>4.1</td>
</tr>
<tr>
<td>N186D/A275P/V622I/V76I</td>
<td>0.94</td>
<td>3.4E-05</td>
<td>21.8</td>
<td>6.3E+05</td>
<td>2.1</td>
</tr>
<tr>
<td>N186D/A275P/V622I/V607I</td>
<td>0.90</td>
<td>5.9E-05</td>
<td>79.8</td>
<td>1.4E+06</td>
<td>4.6</td>
</tr>
<tr>
<td>N186D/A275P/V622I/K34L</td>
<td>0.85</td>
<td>9.0E-05</td>
<td>43.5</td>
<td>4.8E+05</td>
<td>1.6</td>
</tr>
<tr>
<td>N186D/A275P/V622I/K34C</td>
<td>0.82</td>
<td>1.1E-04</td>
<td>50.8</td>
<td>4.6E+05</td>
<td>1.5</td>
</tr>
<tr>
<td>N186D/A275P/V622I/E143H</td>
<td>0.89</td>
<td>6.5E-05</td>
<td>94.2</td>
<td>1.5E+06</td>
<td>4.9</td>
</tr>
<tr>
<td>N186D/A275P/V622I/E143G</td>
<td>0.85</td>
<td>9.0E-05</td>
<td>123.3</td>
<td>1.4E+06</td>
<td>4.6</td>
</tr>
<tr>
<td>N186D/A275P/V622I/G357C</td>
<td>0.99</td>
<td>5.6E-05</td>
<td>14.5</td>
<td>2.6E+06</td>
<td>8.7</td>
</tr>
</tbody>
</table>

Figure 4.11: $T_{50}^{30}$ profiles for top variants from Rounds 1–4. A275P is shown with TTN calculation example.
Table 4.6: Test for Additivity Effects of Variants

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Expected $\Delta T_{50}^{30}$ [°C]</th>
<th>Observed $\Delta T_{50}$ [°C]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A275P/N186D</td>
<td>9.3</td>
<td>6.9</td>
</tr>
<tr>
<td>A275P/V622I</td>
<td>8.9</td>
<td>4.7</td>
</tr>
<tr>
<td>N186D/V622I</td>
<td>14.5</td>
<td>1.3</td>
</tr>
<tr>
<td>N186D/A275P/V622I</td>
<td>16.4</td>
<td>5.2</td>
</tr>
<tr>
<td>N186D/A275P/V622I/V584I</td>
<td>21.6</td>
<td>4.8</td>
</tr>
<tr>
<td>N186D/A275P/V622I/V377K/K426G/V584I</td>
<td>33.0</td>
<td>4.8</td>
</tr>
</tbody>
</table>

The lack of additivity may be attributed to the enzyme’s uncooperative unfolding and the apparent disconnect between the thermodynamic stability and the kinetic thermostability, Chapter II.

4.2.6 Thermodynamic evaluation of variants with circular dichroism

As discussed for the wild-type protein in Chapter II, Section 2.2.3, the apparent melting temperature of the triple variant was also scan rate dependent (Fig. 4.12). Over the range of scan rates between $30^\circ$C·hr$^{-1}$–$90^\circ$C·hr$^{-1}$, the wild-type protein has a difference of $3.8^\circ$C in melting temperature while the difference for the triple variant was $2.7^\circ$C. The average $T_m$ for the triple variant was $41.7^\circ$C. The kinetically thermostabilized triple variant, as indicated by the $5^\circ$C improvement in $T_{50}^{30}$ had an average melting temperature improvement of $2^\circ$C averaged over the three different scan rates. The most thermostable variant, G357C-TM, was additionally evaluated using CD spectroscopy at both scan rates of $30^\circ$C·h$^{-1}$ and $60^\circ$C·h$^{-1}$. The G357C-TM melting temperature was around $41^\circ$C, the same as the triple variant. This result may indicate that dissociation of the dimer plays a large role in the loss of activity prior to the melting of the secondary structure; thus, explaining the differences between the improvements in $T_{50}^{30}$ and TTN when compared to the melting temperature improvements. The mutations that were made improved the kinetic stability more than thermodynamic stability.
Figure 4.12: Circular Dichroism evaluation of Triple Variant (black) compared to wild-type (blue) at thermal scan rates between 30°C·hr⁻¹ - 90°C·hr⁻¹. The observed melting temperature for 30°C·hr⁻¹= 37.3°C (WT), 40.4°C (TM); 60°C·hr⁻¹= 40.4°C (WT), 41.8°C (TM); and 90°C·hr⁻¹= 41.0(WT), 43.5°C (TM).

4.3 Conclusions

The modified structure-guided consensus approach led to variants with improved thermostability, resulting in an AEH with increased operational stability that can be employed in both in synthesis applications and mutagenic studies toward novel activities. The randomization the E143 position that had high B-value further thermostabilized the consensus triple variant demonstrating the value of combining both the B-FIT and consensus approach, as the E143 position would not have been analyzed using the SGC method. The best resulting AEH quadruple variant (N186D/A275P/E143H/-V622I) improved the T₃₀°₅₀ by 8°C with increased activity and total turnover number compared to wild-type. ‘Hits’ were obtained by the screening of only 186 variants, making the MSGC an efficient method. The addition of the disulfide bond led to the most improvement in T₃₀°₅₀; however, the reduction in catalytic activity is not desirable. Overall, a MSGC with individual site-saturation libraries of high B-value
residues is an effective and efficient way to improve the properties of the protein.

4.4 Materials and Methods

4.4.1 Protein

Genomic DNA from *Xanthomonas campestris pv. campestris* ATCC 38835 was obtained from the American Type Culture collections and cloned in *E.coli* as described in Chapter II.

4.4.2 Mutations

Mutations were introduced with the aid of primers per the protocol described by the QuickChange™ Site-Directed Mutagenesis Kit. A list of primers are in Appendix B. Variants were cloned into pET28a (Novagen, Madison, WI) (DE3) pLysS (Invitrogen, Carlsbad, California, USA).

4.4.3 Expression

Overnight cultures from a single colony containing the plasmid were used to inoculate 100 mL cultures in Magic Media (Invitrogen, Carlsbad, California, USA). Cultures were then pelleted by centrifugation and lysed by sonication. Clarified cell extracts were purified by immobilized-metal affinity chromatography (IMAC) with use of a Ni-nitrilotriacetic acid matrix (Qiagen). Purified variants were measure for initial hydrolysis activity of ampicillin which was monitored at 25°C by using HPLC as described in Chapter II.

4.4.4 $T_{50}^{30}$ measurements

4.4.4.1 Large scale activity assay

$T_{50}^{30}$ values were determined by incubating the purified enzyme over a range of temperature from 16°C-36°C for 30 minutes using a gradient method in a PCR thermocycler. After incubation the reaction was quenched in ice for 30-60 minutes prior to residual
ampicillin hydrolysis activity measurement at 25°C (Fig. 4.2). Conversions and residual activities were measured using HPLC. T° were determined using a MATLAB 7.8.0 (R2009a) interpolation algorithm to determine the temperature corresponding to 0.5 residual activity. Residual Activity (RA) after 30 minutes was similarly calculated using an interpolation to determine the RA corresponding to 25°C.

4.4.4.2 Well-Plate Colorimetric activity assay

Colonies were grown on agar plates (Kan/Cm) at 37°C and picked into 96-well microtiter plates, well plate rows A-G contained library variants and row H, columns 01-06 contain empty pET28a vector background colonies and 07-12 contain the N186D/-A275P/V622I(TM) AEH variant as a positive control. The colonies were grown overnight at 30°C in 200 µL LB media containing 30 µg·mL⁻¹ of Kan and 30 µg·mL⁻¹ of Cm and agitated at 180 RPM. After 16 h, the transformants were replicate plated into 200 µL of induction media (Invitrogen MagicMedia™) for expression. Sterile 70% glycerol was added to the master cell plate and stored at -80°C. The expression plate was incubated for 5 h at 30°C and then moved to room temperature for 24 hours. The cells were centrifuged for 15 min at 1500 RPM (1230 x g) and stored at -80°C for a minimum of 16 h. The frozen cell pellets were lysed using 55 µL B-PER detergent (Thermo Scientific, Rockford, IL) for 60 minutes on ice and centrifuged at 15 min at 1500 RPM (1230 x g) to obtain clarified lysate. Effective lysis was confirmed used SDS-PAGE, not shown. 25 µL of the clarified lysate was transferred to pcr tubes for incubation in a thermocycler for 15 min at 34°C, the remaining 30 µL of each well, termed “not incubated,” were kept on ice. Following incubation the samples were stored on ice for up to 1 h. 20 µL of sample was then added to the “not incubated” and “incubated” reaction plates that contained 140 µL of 23 mM ampicillin in 100 mM potassium phosphate buffer, pH 7. The hydrolysis reaction was allowed to proceed at 25°C for 30 minutes. Finally, 25 µL of each well of from the reaction plates was added
to the assay plates that contained 175 µL of 0.05 wt% paradimethyl-amino benzaldehyde (Sigma Aldrich, St. Louis, MO) (PDAB). The PDAB assay was adapted to well plate assay from the PDAB filter lift assay used by del Rio and co-worker to evaluate mutant penicillin G acylases. [135] PDAB forms a Schiff Base with the 6-APA hydrolytic product from the ampicillin hydrolysis reaction [62,135]. The ampicillin hydrolysis reaction is quenched when added to PDAB assay substrate as the Schiff base reaction requires a low pH to occur which denatures the enzyme. The assay plates were incubated for ~15 min at room temperature followed by 5 min centrifugation at 1500 RPM (1230 x g) to clear any cell debris. The 30 min conversion data was measured using SPECTRAMAX Plus 384 spectrophotometer at 405 nm.

0.05 wt% PDAB was prepared by first preparing 10 mL of 0.5 wt% (5 mg·mL⁻¹) PDAB in methanol. 4.7 mL of the 0.5 wt% PDAB was added to 25 mL of 20% acetic acid, followed by 12.5 mL of 50 mM NaOH.

### 4.4.5 High-Performance Chromatography (HPLC)

High-performance liquid chromatography analysis was complete with a Shimadzu-LC-20AT pump, Beckman Coulter Ultrasphere ODS 4.6 mm x 25 cm column, and SPD-M20A prominence diode array detector (PDA) monitored at 215 nm. The mobile phase was a gradient method at 1.0 mL·min⁻¹ that initially contains 20% methanol and 80% 20 mM phosphate buffer (pH 7), under a gradient for 5 min to 45% methanol and returning to 20% method over the next 5 min. All components, (R)-PG, 6-APA, and (R)-ampicillin, were analyzed on the HPLC with >95% mass balance closure. See Appendix A for more information on the HPLC methods.

### 4.4.6 CD

All analysis was conducted using a JASCO CD J-810 with quartz cuvettes. Thermal scans ranged between 8-70 °C and with a scan rate of 30-90 °C·hr⁻¹. Data fit to a two-state van’t Hoff model and analyzed using non-linear regression algorithm
in MATLAB (Appendix H) to determine $\Delta H$, $\Delta S$, $\Delta G$, and $T_M$ as described in Greenfield [116].

4.4.7 Publication Information

The work presented in Chapter IV of this dissertation on the improved thermostability of AEH is in preparation for submission with Thanh Ha, M. Daniel Ricketts, and my advisor, Andreas Bommarius as co-authors.
CHAPTER V

RECOMMENDATIONS AND CONCLUSIONS

5.1 Recommendations

5.1.1 Further Improvements in AEH Thermostability

While significant improvements in kinetic thermostability have been obtained via the Modified Structure Guided Consensus (MSGC) approach, it is of interest to continue to stabilize the enzyme both kinetically and thermodynamically. A melting temperature of 42°C is still low compared to industrially used enzymes such as PGA with a melting temperature ($T_m$) of 64.5°C [129]. The most stabilized variant, N186D/-A275P/V622I/E143H, may be at a local minimum in the sequence-function landscape rendering it difficult to gain further improvements via simple site-directed mutations.

Three approaches are proposed as a path forward to further stabilize the AEH enzyme. 1.) More MSGC (B-FIT NNK and consensus libraries) 2.) Directed evolution via SCHEMA Recombination or Error Prone PCR (epPCR) and 3.) Rational site mutations identified through mechanistic studies.

1. The first approach is to continue with additional site-directed libraries selected according to high B-values. This approach is promising based on the successes of the Round three mutations, as detailed in Chapter IV, a library that focuses on the positions 433–436, which are in the top 10 average B-value rankings, is of interest. Because of library size constraints due to the medium-throughput nature of the screening assay, these positions would have to be individually site saturated or have severely restricted degenerate codons. The site-directed replacement of highly conserved positions (>50% conservation) have been exhausted for the *X. campestris pv. campestris* AEH.
2. Directed evolution protocols such as epPCR and DNA shuffling could be used to dramatically navigate the sequence-function landscape. As epPCR would require a high error rate, resulting in an unreasonable library size, a better approach would be to apply SCHEMA recombination. SCHEMA is a computational algorithm that estimates the amount of disruption ($<E>$) caused upon DNA recombination; thus, selecting crossover sites that are more likely to lead to a folded protein. Shuffling with the known AEHs from *X. citri* and *A. turbidans*, neither of which are significantly more thermostable compared to the *X. campestris pv. campestris* AEH, could lead to an improved AEH, as the SCHEMA protocol has been used to obtain increased thermostabilities above that of the parents [66].

The SCHEMA RASPP (Recombination as a Shortest Path Problem) algorithm [69] was applied to determine crossover sites for AEH with minimum disruption of the structure. We applied the RASPP algorithm (http://www.che.caltech.edu/groups/fha) to predict the crossover sites at positions, 44, 166, 290, 373, 437, 494, 604, resulting in a $<E>$ of 31 and $<m>$, the average number of mutations, of 96. A library with seven crossovers and three parents would consist of 6561 ($3^8$) chimeras but could be reduced to 256 ($2^8$) chimeras with two only parents, *A. turbidans* and *X. campestris pv. campestris* AEHs. A higher risk aim would be to include the thermostabilized *Rhodacoccus sp.* cocaine esterase (RhCocE) double variant (T172R/G173Q) in the SCHEMA algorithm. RhCocE only has 27% amino acid identity and altered activity when compared to the parental AEH proteins, which may result in low numbers of active chimeras. However, SCHEMA has been successfully demonstrated for $\beta$-lactamases with 34-42% identity [69]. The RhCocE double variant has the same $T_m$ value as the thermostabilized AEHs; although kinetic measures of the RhCocE thermostability, $\tau_{1/2} = 370$ min at $37^\circ$C, are improved compared to
the thermostabilized AEH, $\tau_{1/2} \leq 30\text{ min at } 37^\circ\text{C}$. SCHEMA recombination has been used to improve the thermostability of multiple enzymes and would be a useful tool in the development of more stable AEHs.

3. Mechanistic understanding of the AEH unfolding process will help target residues for randomization or site-directed studies. Foremost, the residues 225 and 231 are of interest for randomization to improve the interaction between the $\alpha/\beta$-hydrolase fold domain (I) and the cap-domain(II), proven as an important interaction when cocaine esterase was stabilized [131,132](Fig. 4.4). Additional biophysical characterization would lead to a better understanding of the mechanism of unfolding and perhaps points of instability; thus, a rational protein engineering design can be applied. The three-state Lumry-Eyring model (Eq. 5.1) is the most commonly applied deactivation model [138]; however, more complex four and five state mechanisms have been observed (Eq. 5.2-5.3) [113].

$$N \rightleftharpoons U \rightarrow D$$  \hspace{1cm} (5.1)

where $N =$ native state, $U =$ reversibly unfolded state and $D =$ irreversibly denatured state.

$$N \rightleftharpoons M \rightleftharpoons U \rightarrow D$$  \hspace{1cm} (5.2)

where $N =$ native state, $M =$ reversibly molten globule state that is active or inactive, $U =$ reversible inactive unfolded state and $D =$ irreversibly denatured state.

$$N \rightleftharpoons I_1 \rightleftharpoons I_2,\ldots \rightleftharpoons I_n \rightarrow D$$  \hspace{1cm} (5.3)

where $N =$ native state, $I_n =$ multiple reversible states, and $D =$ irreversibly denatured state.

The deactivation of AEH may be caused by dimer-dimer, subunit-subunit, and/or intra-subunit dissociations, which are first-order processes, followed by
unfolding and irreversible aggregation, the later of which is a second-order process. Alternatively, the AEH’s deactivation may be caused by slight changes around the active site leading to enzyme deactivation prior to unfolding. Mechanistic studies can be aided by energy minimization and molecular dynamics (MD) studies using both Rosetta and AMBER algorithms [131]. These techniques were used to obtain the RhCocE mutations leading to improved half-lives, although the hit rate using this approach was lower than that of the consensus models with only 3/27 mutations demonstrating stability improvements [131]. Elucidating the unfolding mechanism along with computer modeling may help target inter- and intra-facial residues for stabilization.

Further B-FIT randomization and/or two-parent SCHEMA recombination are of interest and can be evaluated using current assay techniques. SCHEMA has the disadvantage that shuffling results in so many changes that it difficult to determine what led to the improvements [63]. Of the three proposed methods, the mechanistic studies followed by rationally-designed libraries is the most accessible route toward improved thermostability and would significantly increase the body of knowledge about this class of enzymes.

### 5.1.2 Altered Substrate Specificities

Penicillin G acylase (PGA) and AEH are promising template enzymes for protein engineering of novel biocatalysis toward the synthesis of complex β-lactam antibiotics. PGA is used commercially in both the production of 6-amino-penicillanic acid (6-APA) from penicillin G and the semi-synthesis of antibiotics such as ampicillin and amoxicillin, which are β-lactamase-sensitive. The binding pocket of PGA is delimited by the protein backbone; thus, the pocket has minimal expansion room to be gained by amino acid side-chain substitutions. AEH’s alternative enzyme fold of homotetrameric α/β-hydrolase versus PGA’s heterodimeric N-terminal serine hydrolase fold
renders the former an interesting candidate for protein engineering of improved and altered activity over PGA.

We began to explore how to change the substrate specificity of the AEHs. Because the crystal structure of AEH has been solved, protein engineering methods should be focused on both rational and semi-rational, such as site-directed degenerate libraries, approaches [63]. Because of the link between thermostability and mutability, the most thermostabilized AEH, currently the quadruple variant (N186D/A275P/-E143H/V622I) designed in Chapter IV, should be used as the template for protein engineering. In addition to the α-amino specificity, AEHs are highly specific to the size of the acyl-side chain moiety of the substrate but are not specific to the β-lactam side of the pocket. The largest known side-chain accepted by AEH is amoxicillin that has an additional hydroxyl group on the benzyl ring when compared to ampicillin. A wide-range of structures can be accepted on the β-lactam side of the pocket including a single methyl, benzoic acid, and a range of β-lactams including 7-ACA and 7-ADCA. The approach to our library design is to use a homology model based on the holo A. turbidans crystal structure (pdb 2B4K) solved with the ampicillin hydrolysis product, (R)-phenylglycine, along with knowledge about previously identified residues in literature (Appendix E) and in this work (Chapter II).

In general, our design strategy will be as follows: 1.) keep the catalytic triad and the area around the 6-APA ring the same, 2.) mutate the residues associated with the carboxylate cluster (D208, E309, D310), because these residues are associated with the enzymes affinity for α-amino side chains, and 3.) open up the binding pocket around the side chain using targeted mutations (Fig. 5.1). Due to the magnitude of change required for the enzyme to accept the side-chain of oxacillin compared to ampicillin’s side chain, multiple targeted amino acid libraries around the enzyme’s active site will be required to evolve the protein.

A “substrate walking” approach, in which the target substrate is modified to
Figure 5.1: Amino Ester Hydrolase library design strategy. The enzyme is high specific for the acyl side-chain moiety of the $\beta$-lactam antibiotics, the phenylglycl group in ampicillin; but it is largely flexible on the $\beta$-lactam side of the pocket, the 6-APA group in ampicillin and oxacillin

structural intermediates between the wild-type and the ultimate target substrate, should be employed. A truncated substrate was used as an intermediate in the Merck and Codexis’s successful redesign of the transaminase, ATA-117 [78]. Zhao et al. similarly applied a substrate walking approach for the coevolution of the human estrogen receptor (hER) to have binding affinity toward corticosterone, which is a novel function for hER [139]. For the AEH redesign, multiple intermediate substrates may be required. Penicillin G is an interesting target substrate as it lacks the $\alpha$-amino group and is commercially available. Other intermediate substrates, such as a truncated oxacillin with only the isoxazole ring, would be of interest. It is a critical barrier in protein engineering to obtain an enzyme-substrate pair with some activity that can be evolved and improved [77]. “Substrate walking” will increase the success rate of the libraries and increase the rational understanding of the mutations made.

5.1.2.1 Rational Active Site Replacement by Homology

Understanding the role of the carboxylate cluster (D208, E309, D310) is critical to the design of targeted libraries to alter AEH specificity. Homologous proteins RhCocE and J1 Acylase do not contain acidic residues at the carboxylate cluster positions and as such the RhCocE and the J1 Acylase enzymes do not require an $\alpha$-amino
Figure 5.2: Table of homologous residues between AEH, RhCocE, and J1 Acylase. Structure of glutaryl-7-ACA, the wild-type J1 Acylase substrate, and cocaine, the wild-type RhCocE substrate, are shown.

group in their wild-type substrate, as shown in Chapter II, Section 2.2.4 and Fig. 5.2. The active sites of both proteins are very similar to AEH’s active site, the RhCocE active site can be aligned to Cα of 0.9 Å. The catalytic triad and oxyanion hole is conserved, while different amino acid residues are found at positions D207, D208, W209, Y222, E309, and D310 (Fig. 5.2). Positions W209 and Y222 are replaced by different aromatic residues suggesting that these changes compensate for size variability in the substrates. In the sequence alignment of X. campestris pv. campestris AEH and RhCocE, a gap appears at the E309 position. However, in the PyMol superposition of the RhCocE structure onto the AEH structure, it appears that loop residue RhCocE L407 or the homologous J1 acylase T470 residue fills the corresponding space in the binding pocket with a small non-polar amino acid (Fig. 2.9). The information gained from the replacement of the four acidic residues in the AEH binding pocket is useful to elucidate the role of the carboxylate cluster and alter the enzyme substrate specificity.
Figure 5.3: PDB: 2b4K A. *turbidans* AEH complexed with phenylglycine. Numbering reflects *X. campestris pv. campestris* enzyme due to clustal W alignment.

5.1.2.2 Semi-Rational Design

To evolve the substrate specificity of AEH toward the β-lactamase resistant antibiotics, such as oxacillin, targeted mutations around the enzymes active-site should be made and tested with either a truncated substrate or oxacillin. Combinatorial active-site saturation testing (CASTing) in combination with iterative saturation mutagenesis (ISM) is a technique that breaks the active site up in to multiple libraries of two-three residues based on the secondary structure and then iteratively evaluating the libraries. This technique keeps the library sizes at a minima while maintaining interaction of residues. This technique led to enhanced enantioselectivity of an epoxide hydrolase [79] and a variety of the method was used with B-FIT [74] to improve thermostability as discussed in Chapter IV.

A series of libraries based on this design was designed in Fig. 5.3. Knowledge from previous experiments should be used as a starting point for these libraries. Libraries 1 and 2 are focused on the carboxylate moiety, while library 3 and 4 are focused on the
space requirement. Restricted libraries, which are degenerate codon libraries that do not code all 20 amino acids, are used to vastly decrease the screening requirements, while maintaining enough diversity and interaction. The NDT library is a well-rounded restricted codon library that has 12 amino acids encoded by 12 codons, the amino acids are diverse with a variety of polar, non-polar, small, large, and aromatic residues. RNT and GNT are smaller libraries that were selected to have a mix of charged and small residues. As the goal is to increase the size of the binding pocket, the TNS libraries has a range of aromatics and small residues. Restricted libraries have been demonstrated to be more efficient and as effective as NNK libraries, which encode all 20 amino acids, in protein evolution [140, 141]. The restricted libraries should be able to radically redesign the AEH binding pocket.

The ability to test for positive variants is a challenge for any directed evolution protocol. For example a library containing three site-saturated (NNK, where N=A,T,C,G and K=T,G) residues will have $32^3 = 32,768$ nucleotide variants or

<table>
<thead>
<tr>
<th>Position</th>
<th>Degenerate Codons</th>
<th>Amino Acids</th>
<th>Size</th>
<th>Screening Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q308</td>
<td>RNT</td>
<td>A,N,D,G,I,S,T,V</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>E309</td>
<td>NDT</td>
<td>R,N,D,C,G,H,I,L,F,S,Y,V</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>D310</td>
<td>GNT</td>
<td>A,D,G,V</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>-</td>
<td>384</td>
<td>1149</td>
</tr>
<tr>
<td>M200</td>
<td>NDT</td>
<td>R,N,D,C,G,H,I,L,F,S,Y,V</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>D208</td>
<td>RNT</td>
<td>A,N,D,G,I,S,T,V</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>W209</td>
<td>TNS</td>
<td>F,L,S(2),Y,C,W,stop</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>-</td>
<td>768</td>
<td>2299</td>
</tr>
<tr>
<td>W465</td>
<td>TNS</td>
<td>F,L,S(2),Y,C,W,stop</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>-</td>
<td>64</td>
<td>190</td>
</tr>
<tr>
<td>V222</td>
<td>NDT</td>
<td>R,N,D,C,G,H,I,L,F,S,Y,V</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>Q226</td>
<td>NDT</td>
<td>R,N,D,C,G,H,I,L,F,S,Y,V</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>-</td>
<td>144</td>
<td>430</td>
</tr>
</tbody>
</table>
20^3 = 8,000 protein variants. To ensure screening of the entire library \( \sim 3 \) times the size of the library must be screened according to 

\[ N = \frac{\ln(1 - P)}{\ln(1 - \frac{1}{n})} \]

which determines the number of colonies (\( N \)) that have to be picked to have a 95% probability (\( P \)) of selecting one of each codon in a library (\( n \)) i.e. \( \frac{1}{384} \) for Library 1 as described in Chapter IV, Section 4.2.3.1. The screening requirements for libraries 1–4 are shown in Table 5.1. Even with the use of restricted codons, testing this many mutants via traditional HPLC would be an extremely arduous task, because HPLC requires the colonies to be transferred to 96 well-plates, reacted with the oxacillin, and analyzed on HPLC (\( \sim 15 \) min per sample), one 96 well-plate would take 32 hours.

Screening using the PDAB assay, as described in Chapter IV, would be useful to avoid the HPLC requirement. The PDAB assay employs pardimethyl amino benzaldehyde (PDAB) that reacts with the hydrolytic product 6-APA to form a colorimetric product that can be detected as a yellow color or by spectrometric analysis at 405 nm. The PDAB assay can be used as a high-throughput filter-lift assay using visual detection of yellow colonies that have been transferred to nitrocellulose filter paper, lysed, reacted with substrate, followed by the PDAB reaction [135]. Similarly, colonies can be picked and transferred to a 96-well plate for analysis of absorbance. Our lab has used the PDAB filter-lift assay to screen \( \sim 30,000 \) transformants for screening a PGA library [62]. However, the sensitivity of this method is too low to detect slow hydrolysis. A better method would be to screen on 96 well-plates, as the spectrophotometer is significantly more sensitive compared to the human eye. The CAST library sizes are still relatively large with screening requirements of 200-2300 colonies, in 96 well plates, which is feasible, though still laborious.

Lastly, different template enzymes could be considered as alternative starting points for the evolution of an enzyme capable of synthesizing \( \beta \)-lactamase resistant antibiotics. For example, the already stabilized double variant of RhCocE, \( (T_m = 41^\circ C \ [131]) \), or J1 Acylase \( (T_{opt} = 40^\circ C \ [142]) \) are both interesting templates. As
previously discussed RhCocE and J1 Acylase enzymes are both members of the $\alpha/\beta$ hydrolase fold family and do not have the $\alpha$-amino specificity, perhaps rendering them easier to evolve. These enzymes feature the advantage as AEH as that they have homo‘multimeric’ tertiary structures and do not undergo the heterodimeric maturation process of PGA. The disadvantage is that neither RhCocE nor J1 Acylase showed ampicillin activity and RhCocE would have to be first evolved into an acylase from an esterase, a change in substrate specificities similar to that previously evolved in the $\alpha/\beta$ hydrolase fold family [143,144]. As a variation of the consensus approach, ancestral sequences may be created from known homologues and can be used as a template for protein engineering. Often, ancestral sequences are known to be more promiscuous than their present day homologues, though preliminary analysis did not indicate that any of the homologues were thermophiles. Creation of an ancestral sequence also requires codon optimized gene synthesis of the entire gene and may be complex to express. [89,145,146]. Obtaining initial novel activity is one of the most difficult evolutionary barriers to overcome, the use of alternative template enzymes may be useful in obtaining a more promiscuous enzyme starting point.

Overall, the combination of structural studies around the enzymes binding pocket, ‘substrate walking’, and targeted restricted codon CAST libraries should successfully redesign the AEH binding pocket, enabling the enzyme to catalyze reactions of novel substrates. Development of a high-throughput screening or selection method still represents the best solution to analyze large libraries and is further discussed in Chapter V, Section 5.1.3. Alternatively, a different starting protein scaffold could be utilized in case a more promiscuous enzyme starting point is required. Ultimately an enzyme that catalyzes the synthesis of $\beta$-lactamase resistant antibiotics can be developed using protein engineering techniques.
Table 5.2: AEH site-directed mutations with Cocaine Esterase and J1 Acylase like active-sites  

<table>
<thead>
<tr>
<th>Name</th>
<th>Mutations</th>
<th>24 h Conversion 20mM Ampicillin</th>
<th>24 h Conversion 20mM Pen G</th>
<th>24 h Conversion 60mM PGME</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2</td>
<td>D207A/D208P-WT</td>
<td>n.d</td>
<td>n.d</td>
<td>n.d</td>
</tr>
<tr>
<td>T1</td>
<td>E309L/D310F-WT</td>
<td>n.d</td>
<td>n.d</td>
<td>n.d</td>
</tr>
<tr>
<td>T3</td>
<td>D207E/D208V-WT</td>
<td>n.d</td>
<td>n.d</td>
<td>n.d</td>
</tr>
<tr>
<td>Positive Control</td>
<td>WT</td>
<td>100%</td>
<td>n.d</td>
<td>100%</td>
</tr>
<tr>
<td>Negative Control</td>
<td>pet28</td>
<td>n.d</td>
<td>n.d</td>
<td>n.d</td>
</tr>
</tbody>
</table>

n.d = not detected above the negative control

5.1.2.3 Preliminary Data

Twelve variants were constructed to mimic the active sites of both the RhCocE and the J1 Acylase as described in Table 5.2. These mutations were incorporated both into the wild-type AEH and into a thermostabilized AEH (N186D/A275P/V622I) that was described in Chapter IV. The thermostabilized variant is hypothesized to be more tolerant of mutations. All of the constructed variants were overexpressed in E. coli BL21(DE3)pLys cells and analyzed for protein expression using protein gels (SDS-PAGE). The sequence of all variants were confirmed by MWG Operon (Huntsville, AL). The variants were expressed and lysed. The soluble lysate was used to hydrolyze phenylglycine methyl ester, ampicillin, and penicillin G; none of the variants exhibited activity above background, Table 5.2. The four variants RhCocE-WT, RhCocE-TM, J1Acylase-WT, and J1Acylase-TM were purified and tested for activity; however, no activity was detected.

Two methods were applied to test for protein folding, dynamic light scattering
Figure 5.4: Mean molar ellipticity per residue, mdeg cm$^{-1}$ dmol$^{-1}$, from CD wavelength scans of homologous variants, RhCocE WT, RhCocE TM, J1Acylase WT, J1Acylase TM, indicated that the RhCocE mutations did not disrupt the AEH secondary structure fold compared to the triple variant (TM) and the wild-type (WT) proteins. The $\alpha$ helix minimum is at 222 nm and $\beta$ sheet minimum is at 216 nm.

(DLS) [147] and circular dichroism (CD) [116] spectroscopy. The hydrodynamic diameter of the wild-type AEH and the four variants was measured by DLS. The hydrodynamic diameter is an indication of the tertiary structure, the hydrodynamic diameter of the wild-type (after filtration with a 20 nm filter) was shown to be $\sim$9.5 nm. The DLS results for the variants gave mixed results, due to the low concentration of the protein in the samples after filtration. Secondly, CD spectroscopy was used to investigate folding of secondary structure elements, the wild-type AEH has 23% $\alpha$ helical and 27% $\beta$ sheet content. Both of the homologous RhCocE variants were found to be similarly folded to the wild-type and triple variant proteins; however, the negative 222 nm peak ($\alpha$ helices) was reduced in this variant. The two J1 acylase AEH variants had significantly reduced molar ellipticities compared to the wild-type, indicating a significant increase in random coil structure. The CD spectroscopy data provided a good indication that the lack of activity for the J1 acylase AEH variants was due to lack of proper folding, although the results do not explain the lack of
Figure 5.5: Enzyme catalyzed hydrolysis of 3-phenyl-5-methyl-4-isoxazolyl-L-leucine (Oxacillin side chain - Leucine) creates free L-Leucine leading to cell grown in the leucine auxotrophic HB-101 cell line.

activity for the RhCocE AEH variants. The RhCocE AEH variants could be further evaluated through the analysis of small libraries around the active site, in addition to the analysis of alternative substrates.

5.1.3 Improved High-Throughput Selection Method

The inability to evaluate large numbers of variants of AEH (as well as PGA) continues to be a challenge in the development of an AEH enzyme with novel activities or substrate specificities. The two current methods of evaluating variants include i.) an HPLC assay that is slow (15 minutes for hydrolysis and 45 minutes for synthesis per sample) and ii.) the colorimetric para-dimethyl amino benzoic acid (PDAB) well-plate assay that lacks sensitivity toward low conversions and is medium-throughput (∼10^3 variants could be evaluated at a time), are not adequate. Currently, neither the HPLC nor the PDAB assays are efficient enough to evaluate large libraries (∼10^3), which would be required even for a two position site-saturated library.

The development of a high-throughput selection (HTS) assay continues to be a promising alternative to the previously described screening methods (see Chapter IV and Chapter V, Section 5.1.2). Selection assays are linked to the survival of an organism under selection pressure. In our case, a positive selection system linked to the enzyme’s function is desirable such that the organism will survive and hence grow, if the enzyme can produce the desired product.
**Figure 5.6:** Schematic mechanism for ligand activation of transcription. The nuclear receptor LBD is fused to the Gal4 DNA-binding domain which binds the Gal4 DNA response element, regulating transcription of an essential gene. Mammalian nuclear receptor coactivator (NR CoAct) is fused to the Gal4 activation domain (Gal4 AD). The LBD binds the ligand and undergoes a conformational change forming a binding site for the coactivator. When LBD and coactivator associate, the Gal4 AD and Gal4 DBD are in close proximity, activating transcription of the regulated gene. (For clarity, only one NR CoAct:Gal4 AD is shown.)

One potential selection assay utilizes a leucine auxotrophic *E. coli* cell line (HB101) with 3-phenyl-5-methyl-4-isoxazolyl-L-leucine (*L*-leucine coupled with the oxacillin side chain) as a substrate (Fig. 5.1.3). Only those cells that express AEHs can hydrolyze the dipeptide thus allow the organism to grow when the leucine is liberated (‘hits’). Gabor *et al.* applied this approach using phenylacetyl-*L*-leucine (PAL) to engineer novel PGA mutants, the screen successfully reduced the screening effort by factor of 5.6 [115]; in our case, the selection should eliminate 100% of the undesirable mutants. To succeed with this assay we would require synthesis of the 3-phenyl-5-methyl-4-isoxazolyl-L-leucine and successful expression into the HB101 cell line. This assay would alleviate nearly 100% of the screening effort and could allow for fast analysis of large libraries such as those in Section 5.1.2.

A second option, is to develop a high throughput assay based on the synthesis of oxacillin. Chemical complementation is a novel yeast selection system that links the binding of a target small molecule ligand to a nuclear receptor; which turns on a gene (ADE2) that allows the yeast to survive on selection media that lacks adenosine
Figure 5.7: Enzyme-linked activation of transcription. A nuclear receptor that has been previously engineered to bind the desired small molecule product of an enzyme is expressed in the microbe. From substrates that are endogenous metabolites or supplied in the media, the enzyme produces the small molecule product. The product (ligand) binds to the nuclear receptor activating transcription of the regulated gene.

(Fig. 5.6) [148]. When the nuclear receptor ligand and the desired enzyme product are the same enzyme-activated growth can be achieved (Fig. 5.7) [148,149]. Chemical complementation has been successfully applied to evaluate large libraries leading to novel nuclear receptor activities [150]. The library size that can be evaluated using chemical complementation is only limited by the transformation efficiency, which is \( \geq 10^6 \) variants, far exceeding our current capabilities. However, enzyme-activated growth has yet to be demonstrated for \( \beta \)-lactam antibiotics. A second challenge is that the chemical complementation system is expressed in yeast, while our target enzymes are expressed in \( E.\text{coli} \). Efforts to move chemical complementation into bacterial systems have proved unsuccessful [151]; thus, the AEH enzyme system must be moved into yeast. Chemical complementation is advantageous in its use of synthesis over hydrolysis, although, it has many hurdles to overcome to be used for selection of enzyme’s capable of synthesizing \( \beta \)-lactamase resistant antibiotics.
5.1.4 Further improvements in Protein Engineering Protocols

The trend of improving proteins with targeted libraries by incorporating structural and mechanistic knowledge as well as algorithms for interpreting data points on the function-sequence map likely will continue and even accelerate. Tools such as Rosetta and ProSAR need to be made more accessible to the general research community so that they can be used in a wider-range of applications. Less dependence on large-scale screening equipment and ultimately a higher hit rate per unit of resource spent drive this development. Several of the described tools, such as targeted libraries, restricted codons, or structure-guided consensus, have contributed to improve biocatalysts to levels useful in large-scale applications. Furthermore, more rational protocols such as the B-FIT method, SGC, SCHEMA, 3DM, CASTING, and ‘neutral’ drift need to be synergized into a single algorithm to guide researchers on the appropriate methodology for stabilizing or developing novel activities for their proteins. Computational methods for improved reactivity based on structures, especially the transition state, seem to require further optimization to yield fast catalysts, however, they already have yielded new catalysts [87, 88].

It is of interest to further challenge the MSGC consensus protocol using a less complex enzyme that unfolds more cooperatively when compared with the AEH. The enoate reductases from the Old Yellow Enzyme (OYE) family of enzymes may be good candidates for this analysis [152–155]. OYEs catalyze the asymmetric catalytic reduction of carbon-carbon double bonds. Stabilization of the OYE from *Kluyveromyces lactis* (KYE1) is an interesting target as it is significantly less stable than the other characterized OYEs. [152] The crystal structure of the homologous OYE from *Sachromyces carlbergensis* (OYE1, 70% identity to KYE1) has been solved. In 2010, the crystal structure of a thermostable OYE from *Thermus scotoductus* SA-01 has also been solved adding structural insight into the thermostabilization mechanism [156]. Further challenging the MSGC protocol with new enzymes will lead to a
more comprehensive stabilization protocol.

5.2 Conclusions

A kinetically stabilized α-amino ester hydrolase has been developed for the use in the synthesis and hydrolysis of β-lactam antibiotics. This AEH was characterized and shown to be a very efficient enzyme with a $k_{cat}$ of 73 s$^{-1}$ and $K_M$ of 1.1 mM for the hydrolysis of ampicillin and has excellent synthesis properties. The enzyme was successfully employed for the first time in a cascade synthesis of ampicillin directly from Pencillin G in a one-pot reaction, thus eliminating the expensive 6-APA isolation step [56]. Lastly, the modified structure-guided consensus approach was able to incorporate stabilization from positions with both high degree of consensus and high B-values with minimal screening effort of < 200 mutants, resulting in a significantly kinetically stabilized enzyme with an 8°C improvement in $T_{50}^{30}$ and 1.3× the wild-type activity.

Efforts are still required to alter this enzyme’s substrate specificity toward the synthesis of β-lactamase-resistant antibiotics. The AEH’s affinity for the α-amino group currently renders it difficult to evolve toward non-natural substrates. Preliminary results indicate replacement of the carboxylate cluster with homologous residues from cocaine esterase and J1 acylase is not sufficient to alter this specificity without destroying the enzyme’s activity. Further improvements in assay and screening techniques will allow for the analysis of larger libraries around the enzyme’s active site leading to the evolution of new and novel activities.
APPENDIX A

ANALYTICAL METHODS- HPLC AND SPECTROPHOTOMETRIC ASSAYS
HPLC Analytical Assay

Column Information: Reversed Phase Ultrasphere® 5 µm Spherical 80 Å Pore, C-18 Analytical

HPLC Information: Two different HPLC systems were used for this work, both HPLC gave identical results. Beckman Coulter System Gold® reverse phase HPLC (Fullerton, CA) with a 168 Detector monitored at 215 nm. Shimadzu-LC-20AT pump, and SPD-M20A prominence diode array detector (PDA) monitored at 215 nm.

Sample Chromatogram (Hydrolysis: Acetonitrile Method)
Method: The reaction was quenched 10X into a solution of the mobile phase. 2uL of sample was injected onto the column. The mobile phase was isocratic at 1.0 mL/min and contains 30% acetonitrile and 70% 5 mM phosphate buffer with 300 mg/L sodium dodecylsulfate (SDS) (pH 3). Method was run for 15 min per sample.
Sample Chromatogram (Hydrolysis: Methanol Method)
Method: The reaction was quenched 10X into a solution of the mobile phase. 2uL of sample was injected on the column. A gradient method was used with the aqueous phase of 20 mM phosphate buffer, pH 7 and an organic phase of methanol. Flowrate was held constant at 1 mL/min

At t= 0 min to 2.5 min, isocratic 20% Methanol: 80% Phosphate

at=2.5 min to 5.5 min, linear gradient 20% to 45% Methanol

t=5.5 min to 9.0 min, linear gradient 45% to 20% Methanol

t = 9.0 to 15.0 min, isocratic 20% Methanol: 80% Phosphate

D-phenylglycine/ RT= 3.4
6-APA/ RT =3.9
Ampicillin/ RT= 7.6
Sample Chromatogram (Synthesis: Methanol Method)
Method: The reaction was quenched 10X into a solution of the mobile phase. 2uL of sample was injected on the column. A step change method was used with the aqueous phase of 20 mM phosphate buffer, pH 7 and an organic phase of methanol. Flowrate was held constant at 1 mL/min

At t= 0 min to 5.5 min, isocratic 20% Methanol: 80% Phosphate
At t= 5.5 min to 25 min, isocratic 35% Methanol
At t = 25 to 35 min, isocratic 20% Methanol: 80% Phosphate

D-phenylglycine/ RT= 2.9
6-APA/ RT =3.6
Ampicillin/ RT= 9.0
D-phenylglycine methyl ester/ RT =21.0
Sample Chromatogram (One-Pot)
Method: The reaction was quenched 10X into a solution of the mobile phase. 2uL of sample was injected on the column. A step change method was used with the aqueous phase of 20 mM phosphate buffer, pH 7 and an organic phase of methanol. Flowrate was held constant at 1 mL/min

At t= 0 min to 5.5 min, isocratic 20% Methanol: 80% Phosphate
At t=5.5 min to 25 min, isocratic 35% Methanol
At t = 25 to 35 min, isocratic 20% Methanol: 80% Phosphate

D-phenylglycine/ RT= 2.933
6-APA/ RT =3.604
Phenylacetic Acid/ RT = 6.789
Ampicillin/ RT= 13.537
D-phenylglycine methyl ester/ RT =15.367
Penicillin G/ RT = 26.785
APPENDIX B

PRIMERS
### Cloning and Sequencing Primers

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaa_gene For</td>
<td>CGC AGT GCC TGG AAG ACA TAT</td>
</tr>
<tr>
<td>gaa_gene Rev</td>
<td>ATC ACC GCA ACC ACC TTT GAC</td>
</tr>
<tr>
<td>gaa_NCOI_For</td>
<td>TGC ATG CAT GCC ATG GTG ATG CTT GCC GCC TGC</td>
</tr>
<tr>
<td>gaa_HINDIII_Rev</td>
<td>C GCC GCC CCA TTT CAA CAC ACC GCC AGA CTG ATG TAG</td>
</tr>
<tr>
<td>gaa_HINDIII_Rev</td>
<td>CGG CCA CTC TGG AGC TCG ACC CCG CGG ACT GAG</td>
</tr>
<tr>
<td>gaa500</td>
<td>GCA TGA TGG CTT GCT CGT ACC AGC</td>
</tr>
<tr>
<td>gaa1000</td>
<td>TTG TGG ATG GGG GCC GGG GGT G</td>
</tr>
<tr>
<td>gaa1500</td>
<td>GCC CAG ATG CAC ATC CGG C</td>
</tr>
<tr>
<td>AT-gene For</td>
<td>ATG GTG GGA GAT ATT ACC CTT TCA</td>
</tr>
<tr>
<td>AT_gene Rev</td>
<td>TTA CTG TTT CAC AAC CGG GAG GCA G</td>
</tr>
<tr>
<td>DelNcoI For</td>
<td>GAA CTG CCT GTG TCT ATG GAT ATT TTC CGG GCC</td>
</tr>
<tr>
<td>AT_NCOI  For</td>
<td>CGG GCC ACA CCA TGG TGG GAC AGA TTA</td>
</tr>
<tr>
<td>AT_HindIII_Rev</td>
<td>GAT ACC GCC AGA CTG TGG TGG ACC CAG GCC GAG</td>
</tr>
<tr>
<td>AT-HindIII_Rev</td>
<td>TAA TGC GGG TGG ATG GCC GGC TGG GCC TGG GCC GAG</td>
</tr>
<tr>
<td>AT-HindIII_Rev</td>
<td>GAT ACC GCC AGA CTG TGG TGG ACC CAG GCC GAG</td>
</tr>
<tr>
<td>AT-HindIII_Rev</td>
<td>TAA TGC GGG TGG ATG GCC GGC TGG GCC TGG GCC GAG</td>
</tr>
<tr>
<td>T7 for</td>
<td>GAA AGC AAC TGC ACT GGT GG</td>
</tr>
<tr>
<td>T7_rev</td>
<td>CTA GGT ATT GCT CAG CGG T</td>
</tr>
</tbody>
</table>

### Site Directed Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>WT</th>
<th>Mutation</th>
<th>Forward Primer 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>resi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76 V</td>
<td>L</td>
<td></td>
<td>GTG GCC CAA ATC CTT CTC ACC GGC AGC AGG CGG TAC</td>
</tr>
<tr>
<td>82 N</td>
<td>A</td>
<td></td>
<td>GTG GCC CAA ATC CTT CTC ACC GGC AGC AGG CGG TAC</td>
</tr>
<tr>
<td>143 F</td>
<td>K</td>
<td></td>
<td>TGC ACC GCC AGA CTG TGG TGG ACC CAG GCC GAG</td>
</tr>
<tr>
<td>148 S</td>
<td>A</td>
<td></td>
<td>GTG GCC CAA ATC CTT CTC ACC GGC AGC AGG CGG TAC</td>
</tr>
<tr>
<td>174 S</td>
<td>A</td>
<td></td>
<td>GTG GCC CAA ATC CTT CTC ACC GGC AGC AGG CGG TAC</td>
</tr>
<tr>
<td>175 V</td>
<td>A</td>
<td></td>
<td>GTG GCC CAA ATC CTT CTC ACC GGC AGC AGG CGG TAC</td>
</tr>
<tr>
<td>175 V</td>
<td>F</td>
<td></td>
<td>GTG GCC CAA ATC CTT CTC ACC GGC AGC AGG CGG TAC</td>
</tr>
<tr>
<td>186 N</td>
<td>D</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>187 T</td>
<td>A</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>207 D</td>
<td>A</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>207 D</td>
<td>N</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>207 D</td>
<td>L</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>208 D</td>
<td>A</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>208 D</td>
<td>N</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>208 D</td>
<td>L</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>209 N</td>
<td>A</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>209 N</td>
<td>Y</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>209 N</td>
<td>F</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>212 F</td>
<td>A</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>212 F</td>
<td>N</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>275 A</td>
<td>P</td>
<td></td>
<td>CAG GGC CAG</td>
</tr>
<tr>
<td>307 D</td>
<td>A</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>307 D</td>
<td>N</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>307 D</td>
<td>L</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>309 E</td>
<td>A</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>309 E</td>
<td>N</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>310 D</td>
<td>A</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>310 D</td>
<td>N</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>310 D</td>
<td>L</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>322 E</td>
<td>K</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>357 C</td>
<td>K</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>377 V</td>
<td>K</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>415 S</td>
<td>A</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>426 K</td>
<td>G</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>433 T</td>
<td>A</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>435 A</td>
<td>G</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>436 G</td>
<td>G</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>485 P</td>
<td>V</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
<tr>
<td>489 P</td>
<td>V</td>
<td></td>
<td>CAG GGC CTG TGG ATG GCC GGC TGG TTC ATT TAG GGT GGC TGC</td>
</tr>
</tbody>
</table>
Site Directed Primers, Continued

<table>
<thead>
<tr>
<th>resi</th>
<th>WT</th>
<th>Mutation</th>
<th>Forward Primer 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>496</td>
<td>I</td>
<td>V</td>
<td>TG ACC GCA CCG CTG CAG GTC GGC GAA GGT GAT GTG CAT</td>
</tr>
<tr>
<td>502</td>
<td>V</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>567</td>
<td>F</td>
<td>V</td>
<td>GCA GCC AAC CAG CCG CTT GCA TAT CAG TGC GCC GTC CAC</td>
</tr>
<tr>
<td>577</td>
<td>P</td>
<td>V</td>
<td>C CTC GCC ACT GCC AAT CAG GTG TTC CAG CCG GGC CAC C</td>
</tr>
<tr>
<td>580</td>
<td>P</td>
<td>K</td>
<td>C ACT GCC AAT CAC ACC TTC CAG AAA GCC CAC CCG GTG ATG GT</td>
</tr>
<tr>
<td>584</td>
<td>V</td>
<td>I</td>
<td>AG CCG GCC CAC CCG ATA ATG GTG CAG GTG CAG TCC AG</td>
</tr>
<tr>
<td>607</td>
<td>V</td>
<td>I</td>
<td>CG CAG ACC TAC GTG CCC AAT ATG TTC TTC GCC AAG CCC G</td>
</tr>
<tr>
<td>622</td>
<td>V</td>
<td>I</td>
<td>G AAG GCG ACC CAG CAC ATA TAC GTG GCC CGG GGG</td>
</tr>
<tr>
<td>207208</td>
<td>DD</td>
<td>AP</td>
<td>C</td>
</tr>
<tr>
<td>207208</td>
<td>DD</td>
<td>EV</td>
<td>GAT GCC TGG ATG GCC GAA GTG TGG TTC ATG TAC GTG GGC TTC</td>
</tr>
<tr>
<td>309310</td>
<td>ED</td>
<td>AA</td>
<td>GCC CTC TGG GAC CAG GCT GCT ATG TGG GCC GCC AEC</td>
</tr>
<tr>
<td>309310</td>
<td>ED</td>
<td>LF</td>
<td>CAG GGC CTC TGG GAC CAG TCT TAT ATG TGG GCC GCC AEC C</td>
</tr>
<tr>
<td>309310</td>
<td>ED</td>
<td>TL</td>
<td>CAG GGC CTC TGG GAC CAG ACG TTA ATG TGG GCC GCC AEC C</td>
</tr>
</tbody>
</table>

Library Primers

<table>
<thead>
<tr>
<th>resi</th>
<th>WT</th>
<th>Mutation</th>
<th>Forward Primer 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>K</td>
<td>NNK</td>
<td>ACC CCG GAT ATC ACC GGC NNK CCG TTC GTG GCA GGC G</td>
</tr>
<tr>
<td>143</td>
<td>E</td>
<td>NNK</td>
<td>GGG CCC CTC AAT CCC AGC NNK GTC GAC CAG GCC ACC GAT G</td>
</tr>
<tr>
<td>175</td>
<td>Y</td>
<td>NDT</td>
<td>C AAG GTC GCC ATG ATC GCA TCC TTC NDT GAA GTT TTC ACC GTG</td>
</tr>
<tr>
<td>225</td>
<td>S</td>
<td>NNK</td>
<td>GCC AAG</td>
</tr>
<tr>
<td>231</td>
<td>F</td>
<td>NNK</td>
<td>C GCC CAG CTC AGC AAG CTT NNK AAG GCC GCC GGC ATC</td>
</tr>
<tr>
<td>310</td>
<td>F</td>
<td>GNT</td>
<td>CAG GCC CTC TGG GAC CAG CTC GNT ATG TGG GCC GCC AEC C</td>
</tr>
</tbody>
</table>
APPENDIX C

CLUSTAL PROTEIN ALIGNMENT

CLUSTAL W (1.83) multiple sequence alignment

<table>
<thead>
<tr>
<th></th>
<th>ALIQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD</th>
<th>ALIQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD</th>
<th>ALIQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAEH_Xcc</td>
<td>---------------------LAACLLATAVAAATSALQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD---------------------</td>
<td>---------------------LAACLLATAVAAATSALQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD---------------------</td>
<td>---------------------LAACLLATAVAAATSALQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD---------------------</td>
</tr>
<tr>
<td>AEH_Xcitri</td>
<td>---------------------LAACLLATAVAAATSALQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD---------------------</td>
<td>---------------------LAACLLATAVAAATSALQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD---------------------</td>
<td>---------------------LAACLLATAVAAATSALQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD---------------------</td>
</tr>
<tr>
<td>AEH_AT</td>
<td>---------------------LAACLLATAVAAATSALQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD---------------------</td>
<td>---------------------LAACLLATAVAAATSALQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD---------------------</td>
<td>---------------------LAACLLATAVAAATSALQHMDLQASAGGYNVYGFHDQTVDRGKYSQDVYMTPLRGLNPSEVHDATD---------------------</td>
</tr>
<tr>
<td>pAEH_Xcc</td>
<td>ITGKPFAQGAADNYVYKFMDQVYTFQVQVMDQYDQYVMTPLRGLNPSEVHDATD</td>
<td>ITGKPFAQGAADNYVYKFMDQVYTFQVQVMDQYDQYVMTPLRGLNPSEVHDATD</td>
<td>ITGKPFAQGAADNYVYKFMDQVYTFQVQVMDQYDQYVMTPLRGLNPSEVHDATD</td>
</tr>
<tr>
<td>AEH_Xcitri</td>
<td>ITGKPFAQGAADNYVYKFMDQVYTFQVQVMDQYDQYVMTPLRGLNPSEVHDATD</td>
<td>ITGKPFAQGAADNYVYKFMDQVYTFQVQVMDQYDQYVMTPLRGLNPSEVHDATD</td>
<td>ITGKPFAQGAADNYVYKFMDQVYTFQVQVMDQYDQYVMTPLRGLNPSEVHDATD</td>
</tr>
<tr>
<td>AEH_AT</td>
<td>ITGKPFAQGAADNYVYKFMDQVYTFQVQVMDQYDQYVMTPLRGLNPSEVHDATD</td>
<td>ITGKPFAQGAADNYVYKFMDQVYTFQVQVMDQYDQYVMTPLRGLNPSEVHDATD</td>
<td>ITGKPFAQGAADNYVYKFMDQVYTFQVQVMDQYDQYVMTPLRGLNPSEVHDATD</td>
</tr>
</tbody>
</table>

120
Red: Catalytic Triad, Blue: Carboxylate Cluster, Green: Oxyanion Hole
APPENDIX D

MODIFIED STRUCTURE GUIDE CONSENSUS
ALIGNMENT AND B-FIT ANALYSIS
<table>
<thead>
<tr>
<th>Name</th>
<th>Resi</th>
<th>Subunit Averaged B-Value</th>
<th>Overall Rank</th>
<th>Selected</th>
<th>Name</th>
<th>Resi</th>
<th>Subunit Average</th>
<th>Overall Rank</th>
<th>Selected</th>
<th>Name</th>
<th>Resi</th>
<th>Subunit Average</th>
<th>Overall Rank</th>
<th>Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>THR</td>
<td>24</td>
<td>34.2</td>
<td>30</td>
<td></td>
<td>VAL</td>
<td>63</td>
<td>17.3</td>
<td>423</td>
<td></td>
<td>ALA</td>
<td>102</td>
<td>15.4</td>
<td>541</td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>25</td>
<td>35.5</td>
<td>20</td>
<td></td>
<td>ILE</td>
<td>64</td>
<td>18.1</td>
<td>375</td>
<td></td>
<td>GLY</td>
<td>103</td>
<td>14.6</td>
<td>573</td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>26</td>
<td>36.6</td>
<td>17</td>
<td></td>
<td>VAL</td>
<td>65</td>
<td>17.0</td>
<td>438</td>
<td></td>
<td>ASP</td>
<td>104</td>
<td>19.9</td>
<td>285</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>27</td>
<td>34.6</td>
<td>25</td>
<td></td>
<td>LEU</td>
<td>66</td>
<td>19.3</td>
<td>311</td>
<td></td>
<td>ASP</td>
<td>105</td>
<td>16.2</td>
<td>499</td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>28</td>
<td>32.5</td>
<td>40</td>
<td></td>
<td>PRO</td>
<td>67</td>
<td>20.6</td>
<td>251</td>
<td></td>
<td>VAL</td>
<td>106</td>
<td>16.9</td>
<td>445</td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>29</td>
<td>28.3</td>
<td>74</td>
<td></td>
<td>LYS</td>
<td>68</td>
<td>23.2</td>
<td>169</td>
<td></td>
<td>PHE</td>
<td>107</td>
<td>20.6</td>
<td>252</td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>30</td>
<td>23.4</td>
<td>157</td>
<td></td>
<td>GLY</td>
<td>69</td>
<td>26.5</td>
<td>96</td>
<td></td>
<td>VAL</td>
<td>108</td>
<td>18.0</td>
<td>385</td>
<td></td>
</tr>
<tr>
<td>ILE</td>
<td>31</td>
<td>23.4</td>
<td>156</td>
<td></td>
<td>ALA</td>
<td>70</td>
<td>25.9</td>
<td>104</td>
<td></td>
<td>GLU</td>
<td>109</td>
<td>24.4</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>32</td>
<td>30.3</td>
<td>59</td>
<td></td>
<td>ARG*</td>
<td>71</td>
<td>37.4</td>
<td>15</td>
<td></td>
<td>GLY</td>
<td>110</td>
<td>20.9</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>33</td>
<td>39.4</td>
<td>10</td>
<td></td>
<td>ASN</td>
<td>72</td>
<td>31.8</td>
<td>45</td>
<td></td>
<td>GLY</td>
<td>111</td>
<td>20.1</td>
<td>277</td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>34</td>
<td>49.5</td>
<td>1 x</td>
<td></td>
<td>ALA</td>
<td>73</td>
<td>21.4</td>
<td>217</td>
<td></td>
<td>TYR</td>
<td>112</td>
<td>19.2</td>
<td>319</td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>35</td>
<td>42.0</td>
<td>6</td>
<td></td>
<td>ILE</td>
<td>74</td>
<td>19.6</td>
<td>301</td>
<td></td>
<td>ILE</td>
<td>113</td>
<td>17.6</td>
<td>405</td>
<td></td>
</tr>
<tr>
<td>PHE</td>
<td>36</td>
<td>34.4</td>
<td>28</td>
<td></td>
<td>ILE</td>
<td>75</td>
<td>18.7</td>
<td>347</td>
<td></td>
<td>ARG</td>
<td>114</td>
<td>18.3</td>
<td>360</td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>37</td>
<td>33.2</td>
<td>35</td>
<td></td>
<td>VAL</td>
<td>76</td>
<td>17.4</td>
<td>417</td>
<td></td>
<td>VAL</td>
<td>115</td>
<td>16.6</td>
<td>465</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>38</td>
<td>33.2</td>
<td>34</td>
<td></td>
<td>LEU</td>
<td>77</td>
<td>17.0</td>
<td>442</td>
<td></td>
<td>PHE</td>
<td>116</td>
<td>16.9</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>39</td>
<td>34.6</td>
<td>27</td>
<td></td>
<td>THR</td>
<td>78</td>
<td>15.6</td>
<td>531</td>
<td></td>
<td>GLN</td>
<td>117</td>
<td>17.9</td>
<td>387</td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>40</td>
<td>32.1</td>
<td>42</td>
<td></td>
<td>ARG</td>
<td>79</td>
<td>17.2</td>
<td>431</td>
<td></td>
<td>ASP</td>
<td>118</td>
<td>17.4</td>
<td>414</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>41</td>
<td>31.6</td>
<td>47</td>
<td></td>
<td>THR</td>
<td>80</td>
<td>16.8</td>
<td>458</td>
<td></td>
<td>VAL</td>
<td>119</td>
<td>17.1</td>
<td>436</td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>42</td>
<td>24.5</td>
<td>134</td>
<td></td>
<td>PRO</td>
<td>81</td>
<td>16.7</td>
<td>461</td>
<td></td>
<td>ARG</td>
<td>120</td>
<td>20.9</td>
<td>241</td>
<td></td>
</tr>
<tr>
<td>ASN</td>
<td>43</td>
<td>20.0</td>
<td>282</td>
<td></td>
<td>TYR</td>
<td>82</td>
<td>18.0</td>
<td>383</td>
<td></td>
<td>GLY</td>
<td>121</td>
<td>18.2</td>
<td>368</td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>44</td>
<td>19.3</td>
<td>312</td>
<td></td>
<td>ASP</td>
<td>83</td>
<td>19.0</td>
<td>329</td>
<td></td>
<td>LYS</td>
<td>122</td>
<td>18.0</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>TYR</td>
<td>45</td>
<td>16.9</td>
<td>447</td>
<td></td>
<td>ALA</td>
<td>84</td>
<td>16.5</td>
<td>469</td>
<td></td>
<td>TYR</td>
<td>123</td>
<td>18.0</td>
<td>384</td>
<td></td>
</tr>
<tr>
<td>ILE</td>
<td>46</td>
<td>23.5</td>
<td>155</td>
<td></td>
<td>SER</td>
<td>85</td>
<td>18.1</td>
<td>379</td>
<td></td>
<td>GLY</td>
<td>124</td>
<td>19.9</td>
<td>286</td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>47</td>
<td>20.0</td>
<td>284</td>
<td></td>
<td>GLY</td>
<td>86</td>
<td>18.5</td>
<td>356</td>
<td></td>
<td>SER</td>
<td>125</td>
<td>19.1</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td>ARG</td>
<td>48</td>
<td>31.8</td>
<td>44</td>
<td></td>
<td>ARG</td>
<td>87</td>
<td>23.3</td>
<td>165</td>
<td></td>
<td>GLU</td>
<td>126</td>
<td>24.2</td>
<td>143</td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>49</td>
<td>24.9</td>
<td>123</td>
<td></td>
<td>THR</td>
<td>88</td>
<td>17.5</td>
<td>408</td>
<td></td>
<td>GLY</td>
<td>127</td>
<td>21.4</td>
<td>214</td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>50</td>
<td>23.2</td>
<td>171</td>
<td></td>
<td>GLU</td>
<td>89</td>
<td>25.2</td>
<td>119</td>
<td></td>
<td>ASP</td>
<td>128</td>
<td>28.3</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>51</td>
<td>23.4</td>
<td>161</td>
<td></td>
<td>ARG</td>
<td>90</td>
<td>18.1</td>
<td>377</td>
<td></td>
<td>TYR</td>
<td>129</td>
<td>20.4</td>
<td>265</td>
<td></td>
</tr>
<tr>
<td>ILE</td>
<td>52</td>
<td>20.9</td>
<td>235</td>
<td></td>
<td>LEU</td>
<td>91</td>
<td>16.5</td>
<td>472</td>
<td></td>
<td>VAL</td>
<td>130</td>
<td>20.8</td>
<td>245</td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>53</td>
<td>23.9</td>
<td>146</td>
<td></td>
<td>ALA</td>
<td>92</td>
<td>18.2</td>
<td>369</td>
<td></td>
<td>MSE</td>
<td>131</td>
<td>21.2</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>54</td>
<td>22.7</td>
<td>180</td>
<td></td>
<td>SER</td>
<td>93</td>
<td>16.3</td>
<td>491</td>
<td></td>
<td>THR</td>
<td>132</td>
<td>18.9</td>
<td>335</td>
<td></td>
</tr>
<tr>
<td>ARG</td>
<td>55</td>
<td>26.8</td>
<td>92</td>
<td></td>
<td>PRO</td>
<td>94</td>
<td>18.7</td>
<td>344</td>
<td></td>
<td>ARG</td>
<td>133</td>
<td>22.0</td>
<td>199</td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>56</td>
<td>26.0</td>
<td>102</td>
<td></td>
<td>HIS</td>
<td>95</td>
<td>19.0</td>
<td>328</td>
<td></td>
<td>PRO</td>
<td>134</td>
<td>22.0</td>
<td>197</td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>57</td>
<td>26.1</td>
<td>101</td>
<td></td>
<td>MSE</td>
<td>96</td>
<td>16.9</td>
<td>451</td>
<td></td>
<td>LEU</td>
<td>135</td>
<td>25.3</td>
<td>113</td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>58</td>
<td>24.3</td>
<td>137</td>
<td></td>
<td>LYS</td>
<td>97</td>
<td>24.5</td>
<td>129</td>
<td></td>
<td>ARG</td>
<td>136</td>
<td>36.4</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>59</td>
<td>26.9</td>
<td>91</td>
<td></td>
<td>ASP</td>
<td>98</td>
<td>16.7</td>
<td>462</td>
<td></td>
<td>GLY</td>
<td>137</td>
<td>28.1</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>60</td>
<td>20.1</td>
<td>274</td>
<td></td>
<td>LEU</td>
<td>99</td>
<td>15.3</td>
<td>546</td>
<td></td>
<td>PRO</td>
<td>138</td>
<td>31.3</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>HIS</td>
<td>61</td>
<td>20.8</td>
<td>244</td>
<td></td>
<td>LEU</td>
<td>100</td>
<td>16.2</td>
<td>498</td>
<td></td>
<td>LEU</td>
<td>139</td>
<td>29.5</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>62</td>
<td>17.9</td>
<td>391</td>
<td></td>
<td>SER</td>
<td>101</td>
<td>16.4</td>
<td>485</td>
<td></td>
<td>ASN</td>
<td>140</td>
<td>25.9</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Resi</td>
<td>Subunit Average</td>
<td>Overall Rank</td>
<td>Selected</td>
<td>Name</td>
<td>Resi</td>
<td>Subunit Average</td>
<td>Overall Rank</td>
<td>Selected</td>
<td>Name</td>
<td>Resi</td>
<td>Subunit Average</td>
<td>Overall Rank</td>
<td>Selected</td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>-----------------</td>
<td>--------------</td>
<td>-----------</td>
<td>------</td>
<td>-------</td>
<td>-----------------</td>
<td>--------------</td>
<td>-----------</td>
<td>------</td>
<td>-------</td>
<td>-----------------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PRO</td>
<td>141</td>
<td>32.3</td>
<td>41</td>
<td></td>
<td>VAL</td>
<td>180</td>
<td>17.0</td>
<td>441</td>
<td></td>
<td>ASN</td>
<td>219</td>
<td>17.4</td>
<td>418</td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>142</td>
<td>29.3</td>
<td>68</td>
<td></td>
<td>VAL</td>
<td>181</td>
<td>17.2</td>
<td>428</td>
<td></td>
<td>PHE</td>
<td>220</td>
<td>17.8</td>
<td>396</td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>143</td>
<td>39.3</td>
<td>11x</td>
<td>MSE</td>
<td>182</td>
<td>20.5</td>
<td>262</td>
<td>ASP</td>
<td>221</td>
<td>18.3</td>
<td>362</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>144</td>
<td>24.5</td>
<td>130</td>
<td>ALA</td>
<td>183</td>
<td>20.7</td>
<td>248</td>
<td>TYR</td>
<td>222</td>
<td>17.9</td>
<td>390</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>145</td>
<td>21.6</td>
<td>210</td>
<td>LEU</td>
<td>184</td>
<td>20.5</td>
<td>260</td>
<td>PHE</td>
<td>223</td>
<td>18.5</td>
<td>353</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIS</td>
<td>146</td>
<td>18.8</td>
<td>340</td>
<td>THR</td>
<td>185</td>
<td>23.4</td>
<td>159</td>
<td>THR</td>
<td>224</td>
<td>18.5</td>
<td>352</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>147</td>
<td>19.2</td>
<td>313</td>
<td>ASN</td>
<td>186</td>
<td>27.7</td>
<td>80x</td>
<td>GLY</td>
<td>225</td>
<td>17.5</td>
<td>409</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>148</td>
<td>21.3</td>
<td>219</td>
<td>PRO</td>
<td>187</td>
<td>23.4</td>
<td>162</td>
<td>GLN</td>
<td>226</td>
<td>17.5</td>
<td>410</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>149</td>
<td>19.7</td>
<td>293</td>
<td>HIS</td>
<td>188</td>
<td>22.6</td>
<td>181</td>
<td>LEU</td>
<td>227</td>
<td>20.0</td>
<td>283</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>150</td>
<td>18.6</td>
<td>349</td>
<td>PRO</td>
<td>189</td>
<td>25.2</td>
<td>116</td>
<td>SER</td>
<td>228</td>
<td>24.7</td>
<td>127</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRP</td>
<td>151</td>
<td>30.7</td>
<td>55</td>
<td>ALA</td>
<td>190</td>
<td>21.3</td>
<td>221</td>
<td>LYS</td>
<td>229</td>
<td>34.6</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>152</td>
<td>20.5</td>
<td>256</td>
<td>LEU</td>
<td>191</td>
<td>19.6</td>
<td>300</td>
<td>ARG</td>
<td>230</td>
<td>26.7</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>153</td>
<td>18.2</td>
<td>369</td>
<td>LYS</td>
<td>192</td>
<td>20.8</td>
<td>247</td>
<td>GLY</td>
<td>231</td>
<td>20.9</td>
<td>236</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILE</td>
<td>154</td>
<td>19.7</td>
<td>294</td>
<td>VAL</td>
<td>193</td>
<td>17.8</td>
<td>398</td>
<td>LYS</td>
<td>232</td>
<td>34.6</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>155</td>
<td>23.1</td>
<td>175</td>
<td>ALA</td>
<td>194</td>
<td>17.2</td>
<td>432</td>
<td>GLY</td>
<td>233</td>
<td>24.5</td>
<td>132</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRP</td>
<td>156</td>
<td>23.4</td>
<td>159</td>
<td>VAL</td>
<td>195</td>
<td>17.5</td>
<td>412</td>
<td>ALA</td>
<td>234</td>
<td>22.6</td>
<td>183</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>157</td>
<td>21.0</td>
<td>233</td>
<td>PRO</td>
<td>196</td>
<td>16.4</td>
<td>487</td>
<td>GLY</td>
<td>235</td>
<td>19.7</td>
<td>296</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>158</td>
<td>27.0</td>
<td>89</td>
<td>GLU</td>
<td>197</td>
<td>16.5</td>
<td>475</td>
<td>ILE</td>
<td>236</td>
<td>17.5</td>
<td>410</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>159</td>
<td>28.0</td>
<td>78</td>
<td>SER</td>
<td>198</td>
<td>15.7</td>
<td>524</td>
<td>ALA</td>
<td>237</td>
<td>15.7</td>
<td>529</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASN</td>
<td>160</td>
<td>27.4</td>
<td>85</td>
<td>PRO</td>
<td>199</td>
<td>16.4</td>
<td>480</td>
<td>ARG</td>
<td>238</td>
<td>18.3</td>
<td>362</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>161</td>
<td>23.8</td>
<td>147</td>
<td>MSE</td>
<td>200</td>
<td>20.0</td>
<td>281</td>
<td>GLN</td>
<td>239</td>
<td>23.6</td>
<td>152</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>162</td>
<td>27.4</td>
<td>84</td>
<td>ILE</td>
<td>201</td>
<td>18.2</td>
<td>369</td>
<td>GLY</td>
<td>240</td>
<td>15.9</td>
<td>517</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>163</td>
<td>24.8</td>
<td>125</td>
<td>ASP</td>
<td>202</td>
<td>17.8</td>
<td>392</td>
<td>HIS</td>
<td>241</td>
<td>17.6</td>
<td>404</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>164</td>
<td>27.7</td>
<td>83</td>
<td>GLY</td>
<td>203</td>
<td>16.8</td>
<td>453</td>
<td>ASP</td>
<td>242</td>
<td>14.1</td>
<td>591</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASN</td>
<td>165</td>
<td>23.1</td>
<td>173</td>
<td>TRP</td>
<td>204</td>
<td>16.4</td>
<td>479</td>
<td>ASP</td>
<td>243</td>
<td>16.4</td>
<td>481</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>166</td>
<td>24.7</td>
<td>128</td>
<td>MSE</td>
<td>205</td>
<td>19.0</td>
<td>325</td>
<td>TYR</td>
<td>244</td>
<td>14.7</td>
<td>570</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>167</td>
<td>27.3</td>
<td>86</td>
<td>GLY</td>
<td>206</td>
<td>16.5</td>
<td>469</td>
<td>SER</td>
<td>245</td>
<td>16.1</td>
<td>502</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>168</td>
<td>19.1</td>
<td>323</td>
<td>ASP</td>
<td>207</td>
<td>17.8</td>
<td>396</td>
<td>ASN</td>
<td>246</td>
<td>14.8</td>
<td>566</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>169</td>
<td>16.9</td>
<td>443</td>
<td>ASP</td>
<td>208</td>
<td>17.9</td>
<td>386</td>
<td>PHE</td>
<td>247</td>
<td>16.6</td>
<td>464</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>170</td>
<td>16.3</td>
<td>492</td>
<td>TRP</td>
<td>209</td>
<td>17.5</td>
<td>412</td>
<td>LEU</td>
<td>248</td>
<td>19.9</td>
<td>290</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILE</td>
<td>171</td>
<td>15.2</td>
<td>549</td>
<td>PHE</td>
<td>210</td>
<td>15.6</td>
<td>532</td>
<td>GLN</td>
<td>249</td>
<td>27.7</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>172</td>
<td>16.2</td>
<td>495</td>
<td>ASN</td>
<td>211</td>
<td>15.7</td>
<td>528</td>
<td>ALA</td>
<td>250</td>
<td>20.1</td>
<td>280</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>173</td>
<td>18.3</td>
<td>367</td>
<td>TYR</td>
<td>212</td>
<td>20.6</td>
<td>253</td>
<td>GLY</td>
<td>251</td>
<td>18.9</td>
<td>333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>174</td>
<td>20.4</td>
<td>264</td>
<td>GLY</td>
<td>213</td>
<td>16.4</td>
<td>478</td>
<td>SER</td>
<td>252</td>
<td>16.2</td>
<td>493</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYR</td>
<td>175</td>
<td>17.9</td>
<td>389</td>
<td>ALA</td>
<td>214</td>
<td>15.4</td>
<td>542</td>
<td>ALA</td>
<td>253</td>
<td>15.2</td>
<td>551</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>176</td>
<td>18.4</td>
<td>359</td>
<td>PHE</td>
<td>215</td>
<td>15.2</td>
<td>547</td>
<td>GLY</td>
<td>254</td>
<td>17.4</td>
<td>416</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>177</td>
<td>16.5</td>
<td>476</td>
<td>ARG</td>
<td>216</td>
<td>15.4</td>
<td>538</td>
<td>ASP</td>
<td>255</td>
<td>21.0</td>
<td>229</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHE</td>
<td>178</td>
<td>17.4</td>
<td>419</td>
<td>GLN</td>
<td>217</td>
<td>15.7</td>
<td>525</td>
<td>PHE</td>
<td>256</td>
<td>17.2</td>
<td>429</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>179</td>
<td>16.9</td>
<td>443</td>
<td>VAL</td>
<td>218</td>
<td>16.1</td>
<td>503</td>
<td>ALA</td>
<td>257</td>
<td>18.7</td>
<td>346</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Resi</td>
<td>Subunit Average</td>
<td>Overall Rank</td>
<td>Selected</td>
<td>Name</td>
<td>Resi</td>
<td>Subunit Average</td>
<td>Overall Rank</td>
<td>Selected</td>
<td>Name</td>
<td>Resi</td>
<td>Subunit Average</td>
<td>Overall Rank</td>
<td>Selected</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>-----------------</td>
<td>--------------</td>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>-----------------</td>
<td>--------------</td>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>-----------------</td>
<td>--------------</td>
<td>----------</td>
</tr>
<tr>
<td>LYS</td>
<td>258</td>
<td>29.6</td>
<td>66</td>
<td></td>
<td>VAL</td>
<td>297</td>
<td>20.5</td>
<td>257</td>
<td></td>
<td>GLY</td>
<td>336</td>
<td>14.2</td>
<td>586</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>259</td>
<td>23.6</td>
<td>153</td>
<td></td>
<td>PRO</td>
<td>298</td>
<td>19.6</td>
<td>299</td>
<td></td>
<td>PRO</td>
<td>337</td>
<td>14.1</td>
<td>590</td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>260</td>
<td>22.0</td>
<td>198</td>
<td></td>
<td>THR</td>
<td>299</td>
<td>18.2</td>
<td>372</td>
<td></td>
<td>TRP</td>
<td>338</td>
<td>13.2</td>
<td>611</td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>261</td>
<td>22.3</td>
<td>188</td>
<td></td>
<td>MSE</td>
<td>300</td>
<td>16.8</td>
<td>454</td>
<td></td>
<td>ARG</td>
<td>339</td>
<td>17.4</td>
<td>420</td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>262</td>
<td>21.1</td>
<td>228</td>
<td></td>
<td>TRP</td>
<td>301</td>
<td>15.9</td>
<td>516</td>
<td></td>
<td>HIS</td>
<td>340</td>
<td>18.1</td>
<td>376</td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>263</td>
<td>31.3</td>
<td>50</td>
<td></td>
<td>LEU</td>
<td>302</td>
<td>15.5</td>
<td>535</td>
<td></td>
<td>SER</td>
<td>341</td>
<td>16.4</td>
<td>477</td>
<td></td>
</tr>
<tr>
<td>GLN</td>
<td>264</td>
<td>25.3</td>
<td>111</td>
<td></td>
<td>GLU</td>
<td>303</td>
<td>15.6</td>
<td>530</td>
<td></td>
<td>GLN</td>
<td>342</td>
<td>13.2</td>
<td>612</td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>265</td>
<td>22.2</td>
<td>192</td>
<td></td>
<td>GLY</td>
<td>304</td>
<td>14.5</td>
<td>576</td>
<td></td>
<td>VAL</td>
<td>343</td>
<td>14.5</td>
<td>577</td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>266</td>
<td>21.1</td>
<td>226</td>
<td></td>
<td>LEU</td>
<td>305</td>
<td>14.1</td>
<td>593</td>
<td></td>
<td>ASN</td>
<td>344</td>
<td>16.6</td>
<td>468</td>
<td></td>
</tr>
<tr>
<td>TRP</td>
<td>267</td>
<td>21.0</td>
<td>232</td>
<td></td>
<td>TRP</td>
<td>306</td>
<td>13.8</td>
<td>603</td>
<td></td>
<td>TYR</td>
<td>345</td>
<td>19.1</td>
<td>322</td>
<td></td>
</tr>
<tr>
<td>TRP</td>
<td>268</td>
<td>20.5</td>
<td>255</td>
<td></td>
<td>ASP</td>
<td>307</td>
<td>16.0</td>
<td>513</td>
<td></td>
<td>ASP</td>
<td>346</td>
<td>16.9</td>
<td>448</td>
<td></td>
</tr>
<tr>
<td>HIS</td>
<td>269</td>
<td>22.3</td>
<td>190</td>
<td></td>
<td>GLN</td>
<td>308</td>
<td>14.8</td>
<td>568</td>
<td></td>
<td>GLY</td>
<td>347</td>
<td>13.7</td>
<td>605</td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>270</td>
<td>21.0</td>
<td>231</td>
<td></td>
<td>GLU</td>
<td>309</td>
<td>17.5</td>
<td>407</td>
<td></td>
<td>SER</td>
<td>348</td>
<td>14.6</td>
<td>574</td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>271</td>
<td>21.2</td>
<td>224</td>
<td></td>
<td>ASP</td>
<td>310</td>
<td>16.1</td>
<td>504</td>
<td></td>
<td>ALA</td>
<td>349</td>
<td>14.4</td>
<td>580</td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>272</td>
<td>19.9</td>
<td>288</td>
<td></td>
<td>MSE</td>
<td>311</td>
<td>16.4</td>
<td>482</td>
<td></td>
<td>LEU</td>
<td>350</td>
<td>14.9</td>
<td>561</td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>273</td>
<td>23.8</td>
<td>148</td>
<td></td>
<td>TRP</td>
<td>312</td>
<td>19.6</td>
<td>298</td>
<td></td>
<td>GLY</td>
<td>351</td>
<td>16.1</td>
<td>507</td>
<td></td>
</tr>
<tr>
<td>HIS</td>
<td>274</td>
<td>22.9</td>
<td>176</td>
<td></td>
<td>GLY</td>
<td>313</td>
<td>16.8</td>
<td>456</td>
<td></td>
<td>ALA</td>
<td>352</td>
<td>17.2</td>
<td>434</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>275</td>
<td>20.1 x</td>
<td></td>
<td></td>
<td>ALA</td>
<td>276</td>
<td>19.0</td>
<td>326</td>
<td></td>
<td>ILE</td>
<td>315</td>
<td>18.0</td>
<td>381</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>277</td>
<td>19.0</td>
<td>327</td>
<td></td>
<td>HIS</td>
<td>316</td>
<td>18.6</td>
<td>350</td>
<td></td>
<td>PHE</td>
<td>355</td>
<td>15.1</td>
<td>554</td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>278</td>
<td>22.2</td>
<td>191</td>
<td></td>
<td>SER</td>
<td>317</td>
<td>16.2</td>
<td>495</td>
<td></td>
<td>GLU</td>
<td>356</td>
<td>16.9</td>
<td>445</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>279</td>
<td>23.4</td>
<td>158</td>
<td></td>
<td>TYR</td>
<td>318</td>
<td>17.8</td>
<td>399</td>
<td></td>
<td>GLY</td>
<td>357</td>
<td>14.0</td>
<td>595</td>
<td></td>
</tr>
<tr>
<td>PHE</td>
<td>280</td>
<td>21.5</td>
<td>213</td>
<td></td>
<td>ALA</td>
<td>319</td>
<td>19.2</td>
<td>314</td>
<td></td>
<td>ASP</td>
<td>358</td>
<td>14.0</td>
<td>594</td>
<td></td>
</tr>
<tr>
<td>PHE</td>
<td>281</td>
<td>19.1</td>
<td>321</td>
<td></td>
<td>ALA</td>
<td>320</td>
<td>20.1</td>
<td>279</td>
<td></td>
<td>THR</td>
<td>359</td>
<td>13.9</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>282</td>
<td>25.3</td>
<td>111</td>
<td></td>
<td>MSE</td>
<td>321</td>
<td>21.1</td>
<td>227</td>
<td></td>
<td>ALA</td>
<td>360</td>
<td>14.0</td>
<td>597</td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>283</td>
<td>33.0</td>
<td>37</td>
<td></td>
<td>GLU</td>
<td>322</td>
<td>20.2</td>
<td>272 x</td>
<td></td>
<td>ARG</td>
<td>361</td>
<td>25.7</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>GLN</td>
<td>284</td>
<td>25.3</td>
<td>114</td>
<td></td>
<td>PRO</td>
<td>323</td>
<td>25.0</td>
<td>122</td>
<td></td>
<td>GLN</td>
<td>362</td>
<td>15.8</td>
<td>523</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>285</td>
<td>20.2</td>
<td>270</td>
<td></td>
<td>ARG</td>
<td>324</td>
<td>26.5</td>
<td>95</td>
<td></td>
<td>PHE</td>
<td>363</td>
<td>15.3</td>
<td>544</td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>286</td>
<td>20.3</td>
<td>268</td>
<td></td>
<td>ASP</td>
<td>325</td>
<td>21.6</td>
<td>212</td>
<td></td>
<td>ARG</td>
<td>364</td>
<td>14.9</td>
<td>564</td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>287</td>
<td>21.7</td>
<td>207</td>
<td></td>
<td>ALA</td>
<td>326</td>
<td>37.7</td>
<td>14</td>
<td></td>
<td>HIS</td>
<td>365</td>
<td>29.2</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>288</td>
<td>31.8</td>
<td>46</td>
<td></td>
<td>ARG</td>
<td>327</td>
<td>35.1</td>
<td>23</td>
<td></td>
<td>ASP</td>
<td>366</td>
<td>20.6</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>289</td>
<td>21.6</td>
<td>211</td>
<td></td>
<td>ASN</td>
<td>328</td>
<td>19.6</td>
<td>302</td>
<td></td>
<td>VAL</td>
<td>367</td>
<td>18.2</td>
<td>373</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>290</td>
<td>20.5</td>
<td>259</td>
<td></td>
<td>THR</td>
<td>329</td>
<td>22.5</td>
<td>185</td>
<td></td>
<td>LEU</td>
<td>368</td>
<td>17.4</td>
<td>415</td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>291</td>
<td>23.7</td>
<td>149</td>
<td></td>
<td>LEU</td>
<td>330</td>
<td>21.8</td>
<td>204</td>
<td></td>
<td>ARG</td>
<td>369</td>
<td>24.5</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>ARG</td>
<td>292</td>
<td>35.4</td>
<td>21</td>
<td></td>
<td>ASN</td>
<td>331</td>
<td>17.2</td>
<td>429</td>
<td></td>
<td>PRO</td>
<td>370</td>
<td>19.4</td>
<td>308</td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>293</td>
<td>24.3</td>
<td>138</td>
<td></td>
<td>TYR</td>
<td>332</td>
<td>17.3</td>
<td>426</td>
<td></td>
<td>PHE</td>
<td>371</td>
<td>18.4</td>
<td>357</td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>294</td>
<td>27.8</td>
<td>79</td>
<td></td>
<td>LEU</td>
<td>333</td>
<td>16.2</td>
<td>494</td>
<td></td>
<td>PHE</td>
<td>372</td>
<td>19.9</td>
<td>289</td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>295</td>
<td>25.8</td>
<td>106</td>
<td></td>
<td>VAL</td>
<td>334</td>
<td>15.1</td>
<td>553</td>
<td></td>
<td>ASP</td>
<td>373</td>
<td>20.2</td>
<td>269</td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>296</td>
<td>34.3</td>
<td>29</td>
<td></td>
<td>MSE</td>
<td>335</td>
<td>15.7</td>
<td>526</td>
<td></td>
<td>GLN</td>
<td>374</td>
<td>22.4</td>
<td>186</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Resi</td>
<td>Subunit Average</td>
<td>Overall Rank</td>
<td>Selected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYR</td>
<td>375</td>
<td>22.1</td>
<td>195</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>376</td>
<td>20.6</td>
<td>254</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>377</td>
<td>24.1</td>
<td>144</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>378</td>
<td>30.6</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>379</td>
<td>27.7</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>380</td>
<td>25.7</td>
<td>108</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>381</td>
<td>27.0</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>382</td>
<td>33.6</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>383</td>
<td>24.2</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>384</td>
<td>32.7</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>385</td>
<td>19.4</td>
<td>306</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>386</td>
<td>20.4</td>
<td>266</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>387</td>
<td>18.0</td>
<td>382</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>388</td>
<td>16.9</td>
<td>452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHE</td>
<td>389</td>
<td>15.9</td>
<td>518</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILE</td>
<td>390</td>
<td>16.0</td>
<td>509</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYR</td>
<td>391</td>
<td>15.2</td>
<td>552</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASN</td>
<td>392</td>
<td>15.8</td>
<td>522</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>393</td>
<td>16.0</td>
<td>510</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>394</td>
<td>18.9</td>
<td>332</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>395</td>
<td>19.0</td>
<td>324</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASN</td>
<td>396</td>
<td>22.6</td>
<td>182</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIS</td>
<td>397</td>
<td>20.3</td>
<td>267</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRP</td>
<td>398</td>
<td>16.4</td>
<td>484</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>399</td>
<td>18.6</td>
<td>351</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARG</td>
<td>400</td>
<td>28.9</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>401</td>
<td>19.6</td>
<td>303</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>402</td>
<td>29.7</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>403</td>
<td>19.0</td>
<td>330</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRP</td>
<td>404</td>
<td>16.9</td>
<td>449</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>405</td>
<td>21.4</td>
<td>218</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARG</td>
<td>406</td>
<td>23.7</td>
<td>150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>407</td>
<td>23.3</td>
<td>163</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYS</td>
<td>408</td>
<td>26.4</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>409</td>
<td>29.2</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>410</td>
<td>42.4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>411</td>
<td>33.0</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYS</td>
<td>412</td>
<td>31.1</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>413</td>
<td>28.4</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Resi</td>
<td>Subunit Average</td>
<td>Overall Rank</td>
<td>Selected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>492</td>
<td>35.1</td>
<td>22</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>493</td>
<td>23.7</td>
<td>151</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>494</td>
<td>20.9</td>
<td>238</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLN</td>
<td>495</td>
<td>30.8</td>
<td>54</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILE</td>
<td>496</td>
<td>18.9</td>
<td>334 X</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>497</td>
<td>18.4</td>
<td>358</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>498</td>
<td>18.3</td>
<td>366</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>499</td>
<td>18.8</td>
<td>342</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>500</td>
<td>18.8</td>
<td>341</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>501</td>
<td>25.2</td>
<td>118</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>502</td>
<td>19.7</td>
<td>295 X</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIS</td>
<td>503</td>
<td>18.7</td>
<td>121</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>504</td>
<td>22.3</td>
<td>189</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLN</td>
<td>505</td>
<td>26.6</td>
<td>94</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>506</td>
<td>19.6</td>
<td>297</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>507</td>
<td>19.3</td>
<td>310</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>508</td>
<td>19.4</td>
<td>307</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>509</td>
<td>19.9</td>
<td>287</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>510</td>
<td>18.7</td>
<td>345</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>511</td>
<td>18.5</td>
<td>355</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>512</td>
<td>17.8</td>
<td>393</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>513</td>
<td>17.8</td>
<td>394</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>514</td>
<td>16.6</td>
<td>467</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRP</td>
<td>515</td>
<td>18.5</td>
<td>354</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>516</td>
<td>15.4</td>
<td>540</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>517</td>
<td>14.8</td>
<td>567</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>518</td>
<td>14.3</td>
<td>585</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>519</td>
<td>14.3</td>
<td>584</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILE</td>
<td>520</td>
<td>13.3</td>
<td>610</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>521</td>
<td>15.1</td>
<td>557</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>522</td>
<td>14.5</td>
<td>576</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYR</td>
<td>523</td>
<td>16.6</td>
<td>463</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>524</td>
<td>16.5</td>
<td>471</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>525</td>
<td>21.9</td>
<td>201</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLU</td>
<td>526</td>
<td>16.5</td>
<td>474</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>527</td>
<td>17.6</td>
<td>395</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>528</td>
<td>16.4</td>
<td>486</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>529</td>
<td>20.8</td>
<td>242</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASN</td>
<td>530</td>
<td>21.0</td>
<td>230</td>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Resi</td>
<td>Subunit Average</td>
<td>Overall Rank</td>
<td>Selected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHE</td>
<td>609</td>
<td>21.9</td>
<td>202</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>610</td>
<td>19.2</td>
<td>315</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>611</td>
<td>30.7</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>612</td>
<td>22.1</td>
<td>194</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY</td>
<td>613</td>
<td>21.1</td>
<td>225</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>614</td>
<td>20.1</td>
<td>276</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYR</td>
<td>615</td>
<td>18.3</td>
<td>365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLN</td>
<td>616</td>
<td>21.6</td>
<td>209</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS</td>
<td>617</td>
<td>20.8</td>
<td>243</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>618</td>
<td>18.8</td>
<td>337</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>619</td>
<td>18.3</td>
<td>361</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLN</td>
<td>620</td>
<td>17.3</td>
<td>422</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARG</td>
<td>621</td>
<td>25.2</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>622</td>
<td>18.7</td>
<td>343</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYR</td>
<td>623</td>
<td>24.8</td>
<td>124</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>624</td>
<td>20.5</td>
<td>238</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>625</td>
<td>26.2</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>626</td>
<td>30.1</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY*</td>
<td>627</td>
<td>40.6</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLN</td>
<td>628</td>
<td>33.4</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>629</td>
<td>26.3</td>
<td>98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>630</td>
<td>19.8</td>
<td>291</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYR</td>
<td>631</td>
<td>22.8</td>
<td>178</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILE</td>
<td>632</td>
<td>21.6</td>
<td>208</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>633</td>
<td>20.8</td>
<td>246</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEU</td>
<td>634</td>
<td>21.4</td>
<td>216</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>635</td>
<td>20.1</td>
<td>278</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAL</td>
<td>636</td>
<td>22.9</td>
<td>177</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARG</td>
<td>637</td>
<td>36.5</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# APPENDIX E

## TABLE OF IMPORTANT AEH RESIDUES

<table>
<thead>
<tr>
<th>X. citri residue</th>
<th>X. campestris pv. campestris residue</th>
<th>A. turbidans residue</th>
<th>Function</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y82, Y175</td>
<td>Y82*, Y175*</td>
<td>Y112, Y206</td>
<td>Oxyanion hole – stabilizes negative oxygen ion of the tetrahedral intermediate</td>
<td>[50]</td>
</tr>
<tr>
<td>D310, H340</td>
<td>D310*, H340*</td>
<td>D341*, H370*</td>
<td>Proposed to stabilize imidazolinium ion formed during catalysis through Coulomb interactions</td>
<td>[46]</td>
</tr>
<tr>
<td>W465</td>
<td>W465*</td>
<td>W493*</td>
<td>Proposed to stabilize the carboxylate cluster. Activity was reduced (65% residual activity) in the presence of dinitrobenzenesulfenyl chloride.</td>
<td>[46]</td>
</tr>
<tr>
<td>Y222, D208</td>
<td>Y222*, D208*</td>
<td>Y253, D239</td>
<td>Manual docking of ampicillin with short energy minimization showed stacking interactions with the phenylglycine ring.</td>
<td>[46]</td>
</tr>
<tr>
<td>M200,W209,D219</td>
<td>M200*,W209*,D219*</td>
<td>M231*,W240*, A250*</td>
<td>Manual docking of ampicillin with short energy minimization showed that these residues delimit the pocket binding of the phenyl ring.</td>
<td>[46]</td>
</tr>
<tr>
<td>n.a.</td>
<td>n.a.</td>
<td>Y206A</td>
<td>Increased β-lactam transferase activity and Synthesis to Hydrolysis Ratio</td>
<td>[60]</td>
</tr>
<tr>
<td>n.a.</td>
<td>n.a.</td>
<td>S205A, D388A, H370A</td>
<td>Mutations rendered protein inactive, identified as active site residues.</td>
<td>[46]</td>
</tr>
</tbody>
</table>

*Assumed due to homology
# APPENDIX F

## SUBUNIT INTERACTING RESIDUES

<table>
<thead>
<tr>
<th>Chain</th>
<th>Residue numbers (Preserving PDB numeration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90+91+93+95+97+99+102+355+356+357+358+362+524+525+526+528+529</td>
</tr>
<tr>
<td>B</td>
<td>90+91+93+95+97+99+102+355+356+357+358+362+524+525+526+528+529</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chain</th>
<th>Residue numbers (Preserving PDB numeration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>24+26+28+29+30+31+32+35+36+37+39+42+92+93+94+128+129+130+133+134+139+140+141+224+227+228+229+230+237+238+239+240+259+260+261+264+265+266+267+269+270+273+274+280</td>
</tr>
<tr>
<td>B</td>
<td>24+26+28+29+30+31+32+35+36+37+39+42+92+93+94+128+129+130+133+134+139+140+141+224+227+228+229+230+237+238+239+240+259+260+261+264+265+266+267+269+270+273+274+280</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chain</th>
<th>Residue numbers (Preserving PDB numeration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>239+240+241+242+244+249+448+455+456+472+475+477+478+479+480+481+482+483</td>
</tr>
<tr>
<td>B</td>
<td>239+240+241+242+244+249+448+455+456+472+475+477+478+479+480+481+482+483</td>
</tr>
</tbody>
</table>
Amino acids at the subunit-subunit interface of AEH were determined. Interfacial amino acids are shown in red (acidic), blue (basic), yellow (polar) and green (nonpolar)(top) and the corresponding number is shown to the right (bottom). This was done using a program that integrates enhanced relative solvent accessibility (RSA) with high resolution data—SPPIDER. SPPIDER uses machine learning to combine protein interaction data, sequence-based data and structural features for enhanced recognition of protein-protein interaction sites. The program can be found at http://sppider.cchmc.org/
APPENDIX G

EXAMPLE LIBRARY DIVERSITY AND RANKING
K34NNK – Library Diversity

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>N</th>
<th>K</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>K</td>
</tr>
<tr>
<td>1</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>W</td>
</tr>
<tr>
<td>2</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>W</td>
</tr>
<tr>
<td>3</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>S</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>R</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>L</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>R</td>
</tr>
</tbody>
</table>

E143NNK-Library Diversity

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>N</th>
<th>K</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>R</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>L</td>
</tr>
<tr>
<td>wt</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>E</td>
</tr>
</tbody>
</table>
Figure G.1: Example of the results from the high throughput screening assay for the E143NNK library. Library Results: NI = Not Incubated (Ave=0.140.08), I = Incubated (Average = 0.030.03). Triple Mutant Results:NI (Ave =0.110.06), I (Average = 0.010.03): N=6
APPENDIX H

MATLAB CODE FOR ANALYSIS OF KINETIC AND CD DATA

CD Data Analysis

CDmodel.m

function yfit = cdmodel(para, xdata)
    %UNTITLED2 Summary of this function goes here
    % Detailed explanation goes here
    yfit = (1/para(1))*xdata./(xdata+1/(para(1)*para(2)));
    yfit = exp((1./xdata)*(para(1)/8.314472).*xdata/para(2)-1))./(1+exp((1./xdata)*(para(1)/8.314472).*xdata/para(2)-1));
end

CDanalysis.m

load /users/janna/Documents/AEM_Figures/XccCDscanraw.txt;
tempdegC=XccCDscanraw(:,1);
thetaexp=XccCDscanraw(:,2);
tempkelvin = tempdegC+273.15;
a1 = -6;
a2 = -33.4;
alphaexp = (thetaexp-a2)/(a1-a2);
MeltTemp = interp1(alphaexp,tempdegC,0.5,'linear')+273.15;
DeltaH = 174842;
R=8.314472;
\( \alpha = \exp((1./\text{tempkelvin})*(\text{DeltaH}/R).*(\text{tempkelvin/MeltTemp-1}.)) \)
ypred = \exp((1./\text{tempkelvin})*(\text{DeltaH}/R).*(\text{tempkelvin/MeltTemp-1}).\... \)
\( (1+\exp((1./\text{tempkelvin})*(\text{DeltaH}/R).*(\text{tempkelvin/MeltTemp-1}).)) \);
\%nonlinear parameter estimation
[parameters, R, J] = NLINFIT(tempkelvin, alphaexp, @cdmodel, [DeltaH MeltTemp]);
\%now R2 in the actual space
ypred = \exp((1./\text{tempkelvin})*(\text{parameters(1)/8.314472}).\... \)
(\text{tempkelvin/parameters(2)-1}).\... \)
\( (1+\exp((1./\text{tempkelvin})*(\text{parameters(1)/8.314472}).*(\text{tempkelvin/parameters(2)-1})).) \);
MeltTempCal = parameters(2)-273.15
DeltaHcalc = parameters(1)

dev=alphaexp-mean(alphaexp);
resid=alphaexp-ypred;
SST=sum(dev.^2);
SSE=sum(resid.^2);

138
Rsq=1-SSE/SST

covparam = inv(J'*J)*sum(R.'2)/(3);

xfit = [273.15:0.1:373.15];
yfit = exp((1./xfit)*(parameters(1)/8.314472).*xfit/parameters(2)-1))./(1+exp((1./xfit)*(parameters(1)/8.314472).*xfit/parameters(2)-1));

plot (tempkelvin-273.15,alpah
e,tempkelvin-273.15,ypred,-k')
xlabel '\circn'
ylabel '\alpha'
AXIS([0 100 -.1 1.1])

plot (XccCDscanraw(:,1), XccCDscanraw(:,2))

CDconvert.m

%use to convert CD data into mean molar ellipticity per residue
%load raw data to be converted
[labels, x0, y0] = readColData('Users/janna/Desktop/08122010/...WT08122010_run1.txt',3,19);
[labels, x1, y1] = readColData('Users/janna/Desktop/08122010/...WT08122010_run2.txt',3,19);
[labels, x2, y2] = readColData('Users/janna/Desktop/08122010/...TM08122010-50nguL.txt',3,19);
[labels, x3, y3] = readColData('Users/janna/Desktop/08122010/...TM8408122010-50nguL.txt',3,19);
[labels, x4, y4] = readColData('Users/janna/Desktop/08122010/...TM8448608122010-50nguL.txt',3,19);
[labels, x5, y5] = readColData('Users/janna/Desktop/08122010/...TM3775848608122010-50nguL.txt',3,19);
[labels, x6, y6] = readColData('Users/janna/Desktop/08122010/...CocTM_50nguL.txt',3,19);
[labels, x7, y7] = readColData('Users/janna/Desktop/08122010/...CocWT_45nguL.txt',3,19);
[labels, x8, y8] = readColData('Users/janna/Desktop/08122010/...J1TM_50nguL.txt',3,19);
[labels, x9, y9] = readColData('Users/janna/Desktop/08122010/...J1WT_verydilute.txt',3,19);
[labels, x10, y10] = readColData('Users/janna/Desktop/08122010/...08122010/WT_50nguL.txt',3,19);

%load background data
[labels, x1, bg] = readColData('Users/janna/Desktop/08122010/...50mMphosphate08122010.txt',3,19);
yd0 = y0(:,1) - bg(:,1);
yd1 = y1(:,1) - bg(:,1);
yd2 = y2(:,1) - bg(:,1);
yd3 = y3(:,1) - bg(:,1);
yd4 = y4(:,1) - bg(:,1);
yd5 = y5(:,1) - bg(:,1);
yd6 = y6(:,1) - bg(:,1);
yd7 = y7(:,1) - bg(:,1);
yd8 = y8(:,1) - bg(:,1);
yd9 = y9(:,1) - bg(:,1);
yd10 = y10(:,1) - bg(:,1);

M = 68000; %g/mol, enzyme molecular weight
l = 0.5; %cm, pathlength
nr = 615; % number of residues

c0 = 50; %ng/uL, enzyme concentration
c1 = 50; %ng/uL, enzyme concentration
c2 = 50; %ng/uL, enzyme concentration
c3 = 50; %ng/uL, enzyme concentration
c4 = 50; %ng/uL, enzyme concentration
c5 = 50; %ng/uL, enzyme concentration
c6 = 50; %ng/uL, enzyme concentration
c7 = 45; %ng/uL, enzyme concentration
c8 = 50; %ng/uL, enzyme concentration
c9 = 11; %ng/uL, enzyme concentration
c10 = 50; %ng/uL, enzyme concentration

ymr0 = (0.1*yd0*M)/(c0*1*nr);
ymr1 = (0.1*yd1*M)/(c1*1*nr);
ymr2 = (0.1*yd2*M)/(c2*1*nr);
ymr3 = (0.1*yd3*M)/(c3*1*nr);
ymr4 = (0.1*yd4*M)/(c4*1*nr);
ymr5 = (0.1*yd5*M)/(c5*1*nr);
ymr6 = (0.1*yd6*M)/(c6*1*nr);
ymr7 = (0.1*yd7*M)/(c7*1*nr);
ymr8 = (0.1*yd8*M)/(c8*1*nr);
ymr9 = (0.1*yd9*M)/(c9*1*nr);
ymr10 = (0.1*yd10*M)/(c10*1*nr);

plot(x0,ymr0,x1,ymr1,x2,ymr2,x3,ymr3,x4,ymr4,x5,ymr5)
xlabel 'wavelength[nm]'
ylabel 'Mean molar ellipticity per residue [mdeg cm^{-1} dmol^{-1}]'
axis ([200 250 -10 5])
title 'Circular Dichroism'
legend ('WT-AEH-run1', 'WT-AEH-run2', 'N186D/A275P/V622I',
'N186D/A275P/V584I/V622I', ...
'N186D/A275P/K426G/V584I/V622I', 'N186D/A275P/E377K/K426G/V584I/V622I')

figure
plot(x1,ymr1,x2,ymr2,x10,ymr10,x6,ymr6,x7,ymr7,x8,ymr8,x9,ymr9)
xlabel 'wavelength[nm]'
ylabel 'Mean molar ellipticity per residue [mdeg cm^{-1} dmol^{-1}]'
axis ([200 250 -10 5])
grid
title 'Circular Dichroism Scan of AEH Mutants, 50 ng/uL Enzyme, ...
10^oC, in 50 mM Phosphate'
legend ('WT-AEH-run1', 'N186D/A275P/V622I', 'WT-AEH-run3', 'AEH-cocaine esterase-TM',
'AEH-cocaine esterase-WT', 'AEH-J1Acylase-WT', 'AEH-J1Acylase-TM')
Kinetic Data Analysis

Michaelis-Menten
function yfit = michaelismenton(para, xdata)
yfit = para(1)*xdata./(xdata+para(2));

Michaelis-MentenSubstrate
function yfit = michaelismentonsosubstrate(para, xdata)
yfit = para(1)*xdata./(xdata.*(1+xdata./para(3)))+para(2));

MichaelisMentonFit
%Use this program to fit your data to the Michaelis-Menten equation
%insert location of file
S = load ('/users/janna/Documents/AEM_Figures/linweavertest.txt');

y = S(:,2);
x = S(:,1);
unitx = 'mM'
unity = 'k_{cat}Us^{V[}'

ey = bSx

%excel fit values and initial guesses
kcat = 70; % 1/sec
Km = 1; mm
ypred = (kcat)*x./(x+Km);

%R2 in the actual space using initial guess
dev=y-mean(y);
resid=y-ypred;
   SST=sum(dev.^2);
   SSE=sum(resid.^2);
   Rsq=1-SSE/SST

%R2 in the 1/y vs 1/x fit (assuming lineweaver burk initial guess)
\[1.//(beta*x)+gamma\]
oy = (Km./(x*kcat))+1/kcat;
dev = 1./y - mean(1./y);
resid = 1./y - oy;
   SST=sum(dev.^2);
   SSE=sum(resid.^2);
   Rsq2=1-SSE/SST

%nonlinear parameter estimation
[parameters, R, J] = NLINFIT(x,y,michaelismenton, [kcat Km]);

%now R3 in the actual space from nonlinear parameter estimation
ypred = (parameters(1))*x./(x+parameters(2));

dev=y-mean(y);
resid=y-ypred;
   SST=sum(dev.^2);
SSE=sum(resid.^2);
Rsqfit=1-SSE/SST;

kcatfit = parameters(1);
Kmfit = parameters(2);

covparam = inv(J'*J)*sum(R.^2)/(3)

xfit = [0:0.5:max(S(:,1))];
yfit = (parameters(1))*xfit./((xfit+parameters(2)));

%Lineweaver-Burk
lb = polyfit(1./x,1./y,1);
ylb = lb(:,1).*1./x+lb(:,2);
kcatlb = 1./lb(:,2);
Kmlb =kcatlb .* lb(:,1);
dev = 1./y - mean(1./y);
resid = 1./y - ylb;
    SST=sum(dev.^2);
    SSE=sum(resid.^2);
Rsqlb=1-SSE/SST;

%Eadie-Hoffstee
eh = polyfit(y./x, y,1);
yeh = eh(:,1)*y./x +eh(:,2);
kcateh = eh(:,2);
Kmeh = -eh(:,1);
dev = y- mean(y);
resid = y - yeh;
    SST=sum(dev.^2);
    SSE=sum(resid.^2);
Rsqeh=1-SSE/SST;

%Hanes plot
h = polyfit(x, x./y,1);
yh = h(:,1)*x +h(:,2);
kcath = 1/h(:,1);
Kmh = kcath*h(:,2);
dev = x./y - mean(x./y);
resid = x./y - yh;
    SST=sum(dev.^2);
    SSE=sum(resid.^2);
Rsqh=1-SSE/SST;

figure(1)
subplot (2,2,1)
plot (x,y,'ok',xfit,yfit,'-k')
xlabel (unitx)
ylabel (unity)
title 'Michaelis-Menten'

subplot (2,2,2)
plot (1./x,1./y,'ok',1./x,ylb,'-k')
xlabel '1/mM'
ylabel 'sec'
title 'Lineweaver-Burk'

subplot (2,2,3)
plot (y./x, y, 'ok', y./x, yeh, '-k')
xlabel '1/mM-sec'
ylabel 'sec'
title 'Eadie Hoffstee'

subplot (2,2,4)
plot (x, x./y, 'ok', x, yh, '-k')
xlabel 'mM'
ylabel 'mM-sec'
title 'Hanes plot'

figure(2)
plot (x,y,'ok',xfit,yfit,'-k')
xlabel ('unitx')
ylabel ('unity')
title ('Michaelis-Menten Plot')

format short g

help = {'kcatfit' 'Kmfit' 'Rsqfit';'kcatlineweaverburk' 'Kmlineweaverburk'...
'Rsqlineweaverburk';'kcateadie' 'Kmeadie' 'Rsqeadie';...
'kcathanes' 'Kmhanes' 'Rsqhanes'}

Results = [kcatfit Kmfit Rsqfit;kcatlb Kmlb Rsqlb;kcateh Kmehe Rsqueh;kcatlh Kmh Rsqh]

Michaelis-Menten Substrate Inhibition
%Use this program to fit your data to the Michaelis-Menten equation
%insert location of file
S = load ('/users/janna/Documents/AEM_Figures/lineweavertest.txt');

y = S(:,2);
x = S(:,1);
unitx = 'mM'
unity = 'k_{cat},s^{-1}'

%excel fit values and initial guesses
kcat = 70; % 1/sec
Km = 1; % mM

ypred = (kcat)*x./(x*Km);

%R2 in the actual space using initial guess
dev=y-mean(y);
resid=y-ypred;
    SST=sum(dev.^2);
    SSE=sum(resid.^2);
    Rsq=1-SSE/SST

%R2 in the 1/y vs 1/x fit (assuming lineweaver burk initial guess)
\[1./(beta*x)+gamma\]
\[oy = (Km.\,(x*kcat))+1/kcat;\]
\[dev = 1./y - mean(1./y);\]
\[resid = 1./y - oy;\]
\[SST=sum(dev.^2);\]
\[SSE=sum(resid.^2);\]
\[Rsq2=1-SSE/SST\]

%nonlinear parameter estimation
[parameters, R, J] = NLINFIT(x,y,@michealismenton, [kcat Km]);

%now R3 in the actual space from nonlinear parameter estimation
ypred = (parameters(1))*x./(x+parameters(2));

\[dev=y-mean(y);\]
\[resid=y-ypred;\]
\[SST=sum(dev.^2);\]
\[SSE=sum(resid.^2);\]
\[Rsqfit=1-SSE/SST;\]

kcatfit = parameters(1);
Kmfit = parameters(2);

covparam = inv(J'*J)*sum(R.^2)/3

xfit = [0:0.5:max(S(:,1))];
yfit = (parameters(1))*xfit./(xfit+parameters(2));

%Lineweaver-Burk
lb = polyfit(1./x,1./y,1);
ylb = lb(:,1).*1./x+lb(:,2);
kcatlb = 1./lb(:,2);
Kmlb = kcatlb .* lb(:,1);
\[dev = 1./y - mean(1./y);\]
\[resid = 1./y - ylb;\]
\[SST=sum(dev.^2);\]
\[SSE=sum(resid.^2);\]
\[Rsqlb=1-SSE/SST;\]

%Eadie-Hoffstee
\[eh = polyfit(y./x, y,1);\]
\[yeh = eh(:,1)*y./x +eh(:,2);\]
\[kcateh = eh(:,2);\]
\[Kmeh = -eh(:,1);\]
\[dev = y- mean(y);\]
\[resid = y - yeh;\]
\[SST=sum(dev.^2);\]
\[SSE=sum(resid.^2);\]
\[Rsqeh=1-SSE/SST;\]

%hanes plot
\[h = polyfit(x, x./y,1);\]
\[yh = h(:,1)*x +h(:,2);\]
kcath = 1/h(:,1);
Km = kcat*h(:,2);
dev = x./y - mean(x./y);
resid = x./y - yh;
    SST=sum(dev.^2);
    SSE=sum(resid.^2);
    Rsqh=1-SSE/SST;

figure(1)
subplot (2,2,1)
plot (x,y,'ok',xfit,yfit,'-k')
xlabel (unitx)
ylabel (unity)
title 'Michaelis-Menten'

subplot (2,2,2)
plot (1./x,1./y,'ok',1./x,ylb,'-k')
xlabel '1/mM'
ylabel 'sec'
title 'Lineweaver-Burk'

subplot (2,2,3)
plot (y./x, y./x, 'ok', y./x, yeh,'-k')
xlabel '1/mM-sec'
ylabel 'sec'
title 'Eadie Hoffstee'

subplot (2,2,4)
plot (x, x./y, 'ok', x, yh,'-k')
xlabel 'mM'
ylabel 'mM-sec'
title 'Hanes plot'

figure(2)
plot (x,y,'ok',xfit,yfit,'-k')
xlabel (unitx)
ylabel (unity)
title 'Michaelis-Menten Plot'

format short g

help = {'kcatfit' 'Kmfit' 'Rsqfit';'kcatlineweaverburk' 'Kmlineweaverburk'...
    'Rsqlineweaverburk';'kcateadie' 'Kmeadie' 'Rsqueadie';...
    'kcathanes' 'Kmhanes' 'Rsqhanes'}

Results = [kcatfit Kmfit Rsqfit;kcatlb Kmlb Rsqlb;kcateh...
    Kmeh Rsqeh;kcat Kmh Rsqh]

Beta and gamma Synthesis Analysis

function yfit = jannamodel(para, xdata)
yfit = (1/para(1))*xdata./((xdata+1)/(para(1)*para(2)));
end

y = [0.68;1.1;1.34;1.6];
x = [20;33;45;60];

% excel fit values
beta = 38.75*10^(-3);
gamma = 0.1714;

ypred = (1/gamma)*x./(x+1/(beta*gamma));

% R2 in the actual space
dev=y-mean(y);
resid=y-ypred;
  SST=sum(dev.^2);
  SSE=sum(resid.^2);
  Rsq=1-SSE/SST

% R2 in the 1/y vs 1/x fit
oy = 1./(beta*x)+gamma;
dev = 1./y - mean(1./y);
resid = 1./y - oy;
SST=sum(dev.^2);
  SSE=sum(resid.^2);
  Rsq2=1-SSE/SST

% nonlinear parameter estimation
[parameters, R, J] = NLINFIT(x,y,@jannamodel, [gamma beta])

% now R2 in the actual space
ypred = (1/parameters(1))*x./(x+1/(parameters(1)*parameters(2)));

dev=y-mean(y);
resid=y-ypred;
  SST=sum(dev.^2);
  SSE=sum(resid.^2);
  Rsq3=1-SSE/SST

covparam = inv(J'*J)*sum(R.^2)/(3)

xfit = [0:0.5:60]';
yfit = (1/parameters(1))*xfit./(xfit+1/(parameters(1)*parameters(2)));

plot (x,y,'ok',xfit,yfit,'-k');
xlabel ' [6-APA]';
ylabel ' [S:H]_0^-^-1';

AEH Synthesis Analysis

function dydx = svedassynthesis( x,y,beta,gamma,alpha,ado,no)
%Svedassynthesis defines the equation to determine the maximum product yield
% from the enzymatic synthesis of antibiotics from ado=[DPGME]o, no=[6APA]o
% and parameters beta, gamma, and alpha.
\[
dydx = (((\beta \cdot (n-\alpha) \cdot (a-\alpha) - \alpha \cdot x^2) \cdot (1 + \beta \cdot \gamma \cdot (n-\alpha))) / ((1 + \beta \cdot \gamma \cdot (n-\alpha)) \cdot (a-\alpha + \alpha \cdot y)))
\]

end

Svedas Plot

%set parameters and run ode for svedasynthesis equation
load /users/janna/Documents/AEM_Figures/Xc/seedasplotraw.txt;
y0 = 0;
alpha = 0.25; (% kcat/Km(s))/(kcat/Km(h))
%beta = 42.8; % 1/M
beta = 42.8
%gamma = 0.1714;
gamma = 0.23
no = 0.020; % Initial concentration of nucleophile (M)
ado = 0.020; % Initial concentration of electrophile (M)
xspan = [0, 0.020];
ode = @(x,y) svedasynthesis( x, y, beta, gamma, alpha, ado, no);
[x, y] = ode45 (ode, xspan, y0);
x1 = 1000*x;
y1 = 1000*y;

no = 0.033; % Initial concentration of nucleophile (M)
ado = 0.020; % Initial concentration of electrophile (M)
xspan = [0, 0.020];
ode = @(x,y) svedasynthesis( x, y, beta, gamma, alpha, ado, no);
[x, y] = ode45 (ode, xspan, y0);
x2 = 1000*x;
y2 = 1000*y;

no = 0.045; % Initial concentration of nucleophile (M)
ado = 0.020; % Initial concentration of electrophile (M)
xspan = [0, 0.020];
ode = @(x,y) svedasynthesis( x, y, beta, gamma, alpha, ado, no);
[x, y] = ode45 (ode, xspan, y0);
x3 = 1000*x;
y3 = 1000*y;

no = 0.060; % Initial concentration of nucleophile (M)
ado = 0.020; % Initial concentration of electrophile (M)
xspan = [0, 0.020];
node = @(x,y) svedasynthesis( x, y, beta, gamma, alpha, ado, no);
[x, y] = ode45 (ode, xspan, y0);
x4 = 1000*x;
y4 = 1000*y;

no = 0.020; % Initial concentration of nucleophile (M)
ado = 0.033; % Initial concentration of electrophile (M)
xspan = [0, 0.033];
node = @(x,y) svedasynthesis( x, y, beta, gamma, alpha, ado, no);
[x,y] = ode45 (ode,xspan,y0);
x5 = 1000*x;
y5 = 1000*y;

no = 0.020; % Initial concentration of nucleophile (M)
ado = 0.045; % Initial concentration of electrophile (M)
xspan = [0,0.045];
ode = @(x,y) svedassynthesis( x,y,beta,gamma,alpha,ado,no);
[x,y] = ode45 (ode,xspan,y0);
x6 = 1000*x;
y6 = 1000*y;

no = 0.020; % Initial concentration of nucleophile (M)
ado = 0.060; % Initial concentration of electrophile (M)
xspan = [0,0.060];
ode = @(x,y) svedassynthesis( x,y,beta,gamma,alpha,ado,no);
[x,y] = ode45 (ode,xspan,y0);
x7 = 1000*x;
y7 = 1000*y;

no = 0.040; % Initial concentration of nucleophile (M)
ado = 0.040; % Initial concentration of electrophile (M)
xspan = [0,0.060];
node = @(x,y) svedassynthesis( x,y,beta,gamma,alpha,ado,no);
[x,y] = ode45 (ode,xspan,y0);
x8 = 1000*x;
y8 = 1000*y;

x2060 = Xccsvedasplotraw(:,1)-Xccsvedasplotraw(1,1);  
y2060 = Xccsvedasplotraw(:,2);  
x2045 = Xccsvedasplotraw(:,3)-Xccsvedasplotraw(1,3);  
y2045 = Xccsvedasplotraw(:,4);  
x2033 = Xccsvedasplotraw(:,5)-Xccsvedasplotraw(1,5);  
y2033 = Xccsvedasplotraw(:,6);  
x2020 = Xccsvedasplotraw(:,7)-Xccsvedasplotraw(1,7);  
y2020 = Xccsvedasplotraw(:,8);  
x3320 = Xccsvedasplotraw(:,9)-Xccsvedasplotraw(1,9);  
y3320 = Xccsvedasplotraw(:,10);  
x4520 = Xccsvedasplotraw(:,11)-Xccsvedasplotraw(1,11);  
y4520 = Xccsvedasplotraw(:,12);  
x6020 = Xccsvedasplotraw(:,13);  
y6020 = Xccsvedasplotraw(:,14);  
t = Xccsvedasplotraw(:,15);

xlabel('[\text{(R)PG}],[\text{mM}]');
ylabel('[\text{AMP}],[\text{mM}]');
legend('20:60','60:20','20:45','45:20','20:33','33:20','20:20');
Figure2 = plot (t,y2060,t,y2045,t,y2033,t,y2020,t,y3320,t,y4520,t,y6020);
xlabel ('time (min)');
ylabel ('[AMP] mM');
REFERENCES


VITA

Janna Billy-Blum was born in Pensacola, Florida in 1978. She graduated from Pine Forest High School in 1996 and went on to earn a B.S. in Chemical Engineering, Highest Honors, from the Georgia Institute of Technology in May 2001. During her undergraduate tenure at Georgia Tech, she worked as a co-op at BP Amoco Chemical in Wando, SC. Following graduation, she worked at Merck and Company in Elkton, VA, and was involved in the large-scale production of SINGULAR®, ZOCOR®, and CRIXIVAN®. In 2003, she transferred to Merck Research Labs in West Point, PA, where she worked on the production of clinical vaccines. In 2005, she joined the Bommarius Research lab at Georgia Tech to pursue her Ph.D. in Chemical and Biomolecular Engineering, studying the development of novel enzyme catalysts. Janna has been very involved in the Graduate Student Government serving as department senator since 2007 and holding several offices turning her tenure. She helped to organize the 2009 and 2010 Georgia Tech Research and Innovation Conferences (GTRIC), serving as Event Chair in 2009 and Technical Chair in 2010. Janna is married to Richard Blum and enjoys traveling, reading, and hiking in her spare time.