PROTEIN ENGINEERING WITH COMPUTATIONAL MODELING

DISSERTATION

Submitted in Partial Fulfillment of
the Requirements for
the Degree of
DOCTOR OF PHILOSOPHY (Materials Chemistry)

at the

NEW YORK UNIVERSITY
POLYTECHNIC SCHOOL OF ENGINEERING

by

Ching-Yao Yang

May 2015

Approved:

_____

Department Head Signature

_____

Date

Copy No. ____
Univerity ID#: N14332422

Approved by the Guidance Committee:

Professor Jin Ryoun Kim
Associate Professor of Chemical Engineering
New York University

Date

Professor Jin Kim Montclare
Associate Professor of Chemistry
New York University

Date

Professor Vikas Nanda
Associate Professor of Biochemistry
Rutgers University

Date

Professor Evgeny Vulfson
Industry Professor of Biotechnology
New York University

Date

Vita

Carlo Yuvienco was born in New York, New York on March 22nd, 1983 to Dr. Francisco P. Yuvienco and Elizabeth T. Yuvienco. His academic and professional pursuits would reside in New York until the time of this work, starting with his secondary education at Stuyvesant High School. This was followed by an undergraduate education in chemical engineering at the Cooper Union for the Advancement of Science and Art, culminating in a Bachelor of Engineering (granted in 2005). Carlo then was employed at Pall Corporation (East Hills, NY) as a research scientist, working on the development of leukocyte-reduction filtration media. After two years of industry experience, Carlo pursued a Master of Science in biomedical engineering at the Polytechnic Institute of New York University (granted in 2010), immediately preceding the pursuit of his doctorate in biomedical engineering under the advisement of Prof. Jin Kim Montclare, Ph.D. at the now New York University - Polytechnic School of Engineering, in the Department of Chemical and Biomolecular Engineering. From 2008 to 2014, Carlo worked on several projects in the field of protein engineering under the guidance of Prof. Montclare, having conducted experiments in various labs within New York University, the Navy Research Laboratory, the Wright-Patterson Air Force Base Materials and Manufacturing Directorate, and the City University of New York. During the course of his doctoral research, he and his research projects have been supported by the National Science Foundation (GK-12 Program, DGE-0741714; NYU MRSEC Center, DMR-0820341; I-Corps Program, IIP-1332165; DMR-1205384), the Air Force Office of Scientific Research (FA-9550-07-1-0060 and FA-9550-08-1-0266), and the Army Research Office (W911NP-10-1-0228).

Acknowledgments

This body of work encompasses seven years of effort that would not have been possible without the professional and emotional support of my friends, family, and newfound colleagues. I declare my immense thanks to them for the following reasons.

First, to my family:

The continuous pride and understanding that my family has expressed in regards to the pursuit of my doctorate has been the driving force behind my stalwart efforts these past few years. They have continued to champion my marathon despite my long-awaited presence at the finish line. And during the times of despair and frustration that commonly plague scientific research, I have always been able to benefit from the respite that I received in the sanctuaries that were their homes. To all of them, I convey my eternal gratitude and love.

To my colleague, friend, and mentor - Jin Montclare:

From day one, you have continued to foster and promote my advancement as a scientist and intellectual. Assuredly, I am a better person for it. We have celebrated academic successes as well as suffered scientific and professional tribulations. But perhaps most importantly, regardless of the context, you have undoubtedly and consistently demonstrated your utmost trust in both my abilities and my character - a privilege for which I can only hope I have abundantly reciprocated in kind. Thank you very much for the momentous opportunity to have worked with you.

To my colleagues:

There are many individuals to whom I owe thanks for their professional support during the course of my doctoral studies. They include fellow students (high school, undergraduate, and graduate alike), accomplished scientists, aspiring professors, and university administrators and staff. A vast majority of them have provided me with an infinitely expansive sounding board for my ideas and have in turn stimulated the

advancement of my research. Others have quelled quagmires before they escalated toward affecting the completion of my degree. And others have simply promoted my advancement as a research scientist through the extension of rare professional opportunities.

To my close friends:

I submit my sincerest apologies, more so and in addition to my gratitude. A graduate career of seven years has in truth robbed me of many occasions to spend with my close personal friends. They have nevertheless continued to support my pursuit and take pride in my endeavors.

And to my Jennifer:

Without an iota of doubt, my professional accomplishments would not have been possible without the love and support that you have devoted to me. You have kept me company during late night experiments. You have consoled me during times of struggle and sadness. You have cheered my successes and announced them atop mountains. You are my rock and my cushion. And in my seven-year pursuit of knowledge of the unknown, I have come to only one surety - that I love you.

To all of you again - thank you.

ABSTRACT

PROTEIN ENGINEERING WITH COMPUTATIONAL MODELING

by

Ching-Yao Yang

Advisor: Prof. Jin Kim Montclare, Ph.D.

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy(Materials Chemistry)

May 2015

With the advancement of technologies to probe and manipulate biophysical matter, the scientific community continues to ever better engineer biological systems with the complexity and elegance in design that is necessary to address biomedical challenges. The growing maturity of the field of protein engineering is a testament to this proclamation. Herein, two fundamental ideas are explored. In Chapter I, an evaluation is presented on the effects of the incorporation of a non-canonical, fluorinated amino acid into a protein-based block copolymer. The ramifications of these results, and similar others in the field, on the promise for predictable tuning of the physicochemical behavior and properties of protein-based materials are emphasized. In Chapter II, an alternative application of an endogenous protein is examined, harnessing its inherent form and function. Hypotheses postulate the ability of a coiled-coil protein, of particularly high oligomeric order, to facilitate the delivery of small molecule therapeutics for the treatment of osteoarthritis, whilst addressing dominant hurdles pertaining to drug localization. This complete body of work rests on the themes of control and repurposed application of biophysical matter, contributing to the formalization of engineered systems within protein science.

Contents

1 Improved Stability and Half-Life of Fluorinated Phosphotriesterase Using Rosetta **1**

1.1 Introduction 1

1.1.1 Protein Engineering 1

1.1.1a Rational Design 1

1.1.1b Rosetta Design 1

1.1.1c Protein Engineering Bearing Non-natural Amino Acids 2

1.1.2 Phosphotriesterase 2

1.1.3 Incorporation of Non-natural Amino Acids 4

1.1.4 Fluorinated Amino Acids In Proteins 6

1.1.5 Scope of Work 7

1.2 Methods 7

1.2.1 General 7

1.2.2 Recombinant Gene Construction 7

1.2.3 Protein Expression 8

1.2.4 PyRosetta Design 9

1.2.5 Thermo-stability and Secondary Structure of Phosphotriesterase 10

1.2.5a Nano-DSC 10

1.2.5b Circular Dichroism 10

1.2.6 Enzyme Kinetics 10

1.2.7 MALDI-TOF Mass Spectrometry 11
1.3 Results and Discussion 12
1.3.1 Biosynthesis of Phosphotriesterase 12
1.3.2 Thermo-stability And Secondary Structure 12
1.3.3 Enzymatic Kinetics of Phosphotriesterase 15
1.3.4 Protein Design 16

2 Effects of Phenylalanines Outside Dimer Interface of Phosphotriesterase 22

2.1 Introduction 22
2.1.1 Phosphotriesterase 22
2.1.2 Side-chain effects 22
2.2 Methods 22
2.2.1 General 22
2.2.2 Protein preparation methods 23
2.2.2a Biosynthesis 23
2.2.2b Protein purification 23
2.2.2c Endotoxin removal 23
2.2.2d Protein concentration 23
2.2.3 In vitro methods 23
2.2.3a Cell culture methods 23
2.2.3b Reverse transcription-polymerase chain reaction (PCR) and real-time PCR analysis 23
2.2.4 Biophysical methods 24
2.3 Results 24
2.3.1 Scaling biosynthetic methods 24
2.3.2 Effects of COMPcc on BMS493 therapy 24
2.3.3 Endotoxin levels in COMPcc protein 24

2.4 Discussion 24
2.4.1 Contamination of recombinant COMPcc with endotoxins . . . 24
2.4.2 Meso-scale features of COMPcc promote intra-articular delivery applications 25
2.4.3 Future work 25

List of Figures

1-1 The scheme of scoring function of Rosetta. 1: Lennard-Jones Potential; 2: implicit solvent model; 3: hydrogen bonding; 4: electrostatics; 5: PDB drived torsion potential. 2

1-2 Structure of and active site of phosphotriesterase: (A) Crystal structure of PTE (PDB 1HZY). Wild-type PTE consists of two monomers. Shown in light blue is one of them, and dark blue is the other. Yellow dots represent zinc atoms; (B) Substrates that PTE hydrolyzed: (1) paraoxon, (2) 2-naphthyl acetate, and (3) chloropyfios. (C) Small pocket residues are labeled in red: G60,I106, L303, S308; large pocket residues are labeled in blue: H254, H257, M317, while leaving group residues are labeled in grey: W131, F132, F306, Y309. 3

1-3 Illustration of NAA incorporation via (a) RSI using auxotrophic strain, (b) engineering additional copies of endogenous AARS, (c) expanding the AARS binding pocket and (d) shrinking the AARS editing pocket. 5

1-4 SDS-PAGE of purified proteins: (A) purified PTE and 104A, (B) purified *p*FF-PTE and *p*FF–104A. 12

1-5 MALDI-TOF mass spectra of tryptic peptide fragments. 13

1-6 CD wavelength scans of PTE (circles), 104A (triangles), *p*FF-PTE (diamonds) and *p*FF-104A (squares). 14

1-7 Differential scanning calorimetry thermograms of (A) F104A and (B) *p*FF-F104A. 14

List of Tables

1.1 (A) physical properties of the C-F bond. (B) comparison of C-H and C-F bonds, van der Waals radius, and total size 7

Chapter 1

Improved Stability and Half-Life of Fluorinated Phosphotriesterase Using Rosetta

1.1 Introduction

1.1.1 Protein Engineering

1.1.1a Rational Design

1.1.1b Rosetta Design

Computational tools are widely used for protein engineering[1–8]. Rosetta suite was first developed in University of Washington. Baker *et al.* adapted this suite for prediction of three-dimensional structure of proteins[5]. This suite provides a handful of protocols for analyzing and mutating protein structures. The simulation replies heavily on knowledge-based potentials. It is a suite of libraries and tools for macromolecular ligand docking, to thermo-stabilize proteins, to design a hydrogen-bond network, to design novel protein folds, to create novel protein interfaces, and to de-

sign enzymes, including some containing unnatural amino acids in the active sites (Figure 1-1).

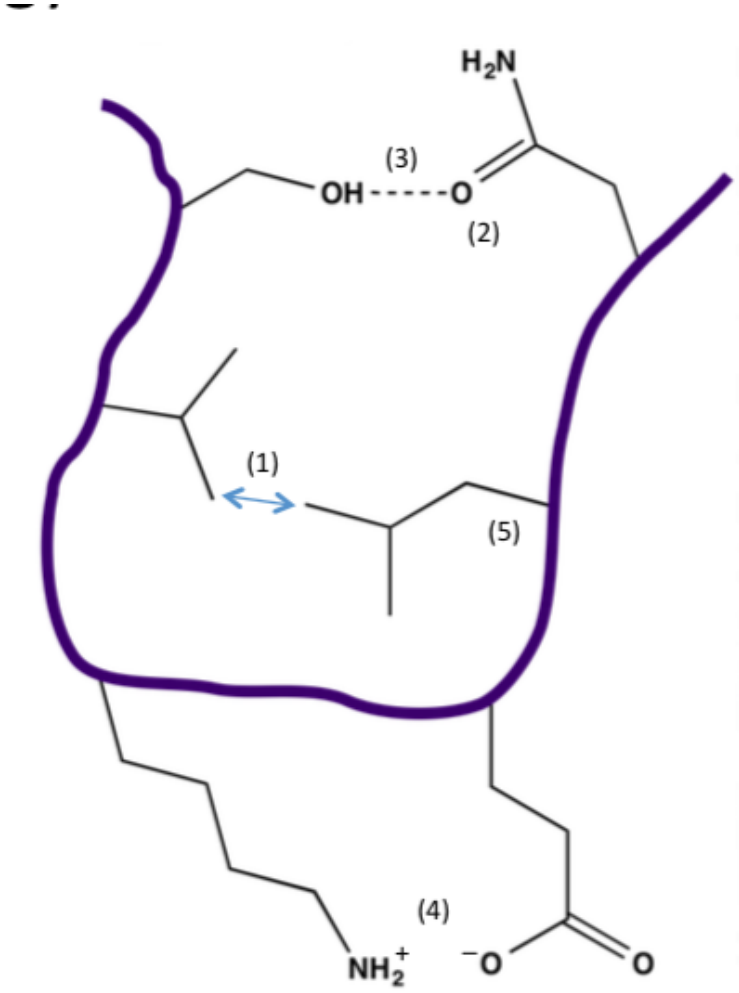

Figure 1-1: The scheme of scoring function of Rosetta. 1: Lennard-Jones Potential; 2: implicit solvent model; 3: hydrogen bonding; 4: electrostatics; 5: PDB drived torsion potential.

1.1.1c Protein Engineering Bearing Non-natural Amino Acids

1.1.2 Phosphotriesterase

PTE is a homodimeric protein composed of two monomers, each of which contains a metallo-active site.Phosphotriesterase (PTE) are enzymes, which hydrolyze

organophosphates (OPs) as well as synthetic esters (Figure 1-2).[9] OPs are a synthetic class of small molecule that irreversibly inactivate acetylcholinesterase (AChE), disrupting neural transmission. AChE is an enzyme that degrades the neurotransmitter, acetylcholine, at the neuromuscular junction in the cholinergic nervous system. After the acetylcholine is hydrolyzed, the synaptic transmission would be terminated. Inhibition of AChE lead to hyper-stimulation from toxic accumulation of acetylcholine.[10] Army also adapted this protein for chemical weapons neutralization. [11]

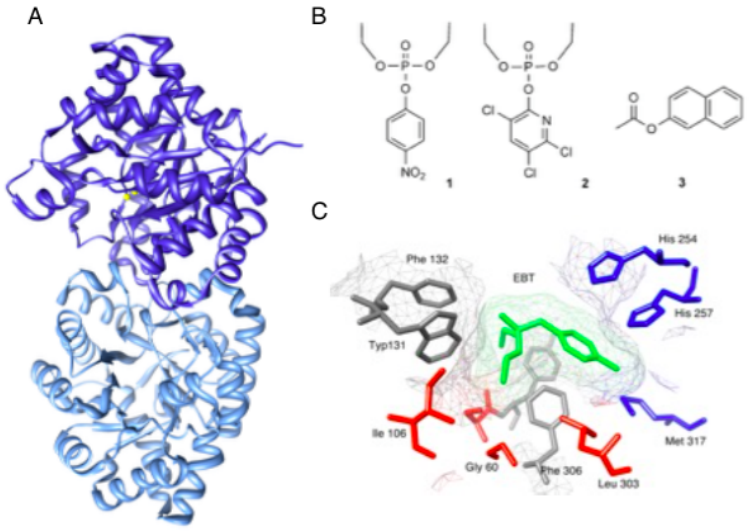

Figure 1-2: Structure of and active site of phosphotriesterase: (A) Crystal structure of PTE (PDB 1HZY). Wild-type PTE consists of two monomers. Shown in light blue is one of them, and dark blue is the other. Yellow dots represent zinc atoms; (B) Substrates that PTE hydrolyzed: (1) paraoxon, (2) 2-naphthyl acetate, and (3) chloropyfios. (C) Small pocket residues are labeled in red: G60,I106, L303, S308; large pocket residues are labeled in blue: H254, H257, M317, while leaving group residues are labeled in grey: W131, F132, F306, Y309.

1.1.3 Incorporation of Non-natural Amino Acids

Several methods have been developed for the incorporation of unnatural amino acids into proteins: solid-phase synthesis (SPPS)[12], in vivo and in vitro site-specific incorporation,[13, 14] and residue-specific incorporation (Figure 1-3).[15] In SPPS, activated amino acids are immobilized on a solid support and synthesized step-by-step in the reactant solution. This method is convenient for the introduction of functional groups into peptides, but it is still restricted to the yield and the expense of peptides. For example, if each coupling step has 99% yield, a 26-amino acid peptide would be synthesized in 77% final yield. To synthesis longer chain peptides and proteins bearing UAAs, biosynthetic methods have been developed. There exists two contemporary methods to biosynthetically incorporate non-natural amino acids into proteins: site-specific incorporation and residue-specific incorporation. Bain *et al.* have developed a general approach for the *in vitro* synthesis of proteins.[16] The approach relies on the suppression of an amber termination codon (UAG) in the mRNA by an amber suppressor tRNA charged with the amino acid analog. This method has been well studied and developed in research of protein structures and functions.[17, 18]

Methods to incorporate amino acid analogues site-specifically into proteins in vivo greatly expand research of unnatural amino acids. We are not only able to synthesize large amounts of protein, but capable of overcoming potential problems including post translational modifications. An *in vivo* site-specific method to incorporation UAAs was developed by Schultz and coworkers.[19, 20] A stop codon at the position of interest is encoded in the mRNA. For *in vivo* site-specific UAA incorporation, an orthogonal aminoacyl-tRNA synthetase charges an orthogonal tRNA with particular UAA, and the suppressor tRNA would help the incorporation of UAA with recognition of a stop codon. As cells contain 20 aminoacyl-tRNA synthetase/suppressor tRNA pairs, a new one is required for the incorporation. An orthogonal aminoacyl-

tRNA synthetase/suppressor tRNA pair based on a TyrRS/tRNATyr pair in the *Methanococcus jannaschii* has been engineered for use in *E. coli* for the incorporation of tyrosine analogs.[19]

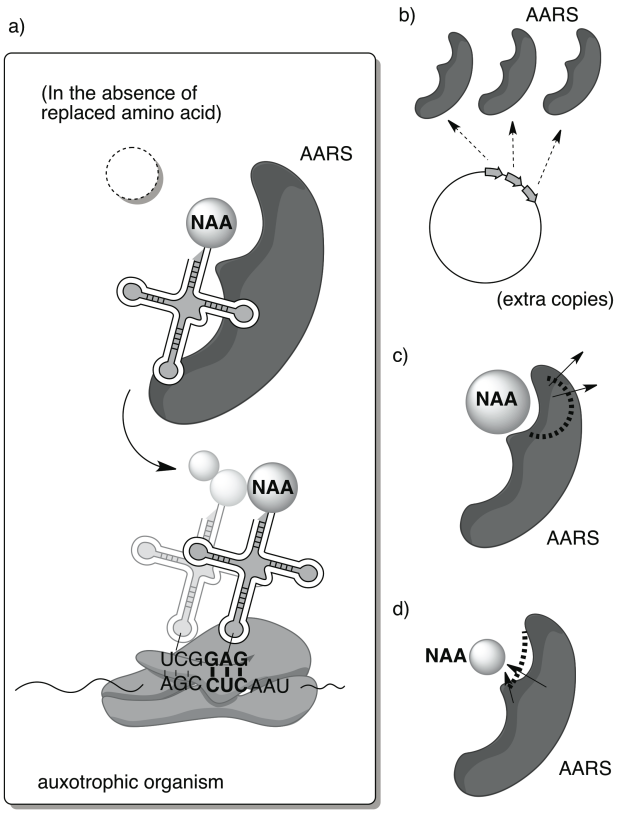

Figure 1-3: Illustration of NAA incorporation via (a) RSI using auxotrophic strain, (b) engineering additional copies of endogenous AARS, (c) expanding the AARS binding pocket and (d) shrinking the AARS editing pocket.

As an alternative to site-specific incorporation, residue-specific incorporation has been developed in which a natural amino acid is replaced with an UAA. Auxotrophic strains or organisms that cannot biosynthesize a particular natural amino acid, has been used to introduce multiple UAAs throughout the protein sequence. UAAs that are isosteric to natural amino acids are capable of being recognized by the natural aminoacyl-tRNA synthetase (aaRS), charging the appropriate tRNA enabling the introduction of UAA into the protein sequence without alteration of the biosynthetic

machinery. However, to introduce UAAs with gross differences from the natural amino acids, further engineering of the aaRS is required. To incorporate refractory methionine analogs, Tirrell and coworkers engineered additional copied of the methionyl-tRNA synthetase (MetRS) by adding the MetRS gene under constitutive promotor.[21] Alternatively, Schimmel and coworkers mutated editing pocket of valyl-tRNA synthetase (ValRS) to facilitate the incorporation of analogs that normally would not be accepted by endogenous aaRS.[22] Finally, Kast and coworkers generated a mutated phenylalanyl-tRNA synthetase (PheRS), ePheRS* under a constitutive promoter, with a large binding pocket (T251G) and showed relaxed specificity.[23]

1.1.4 Fluorinated Amino Acids In Proteins

Fluorinated amino acids (FAAs), represent a unique class of UAAs. They have different bond energies, electron distributions, and hydrophobicity[24] as compared to their hydrogenated counterparts. As we compare the structure of fluorocarbon groups, the C-F bond is highly dipolar while the hydrocarbon is less. The C-F bond is roughly 0.24 Å longer than C-H bond (Table 1.1).[25] While in some cases the global replacement of hydrophobic amino acids with fluorinated analogs has led to the stabilization of protein structure[24], it has all been shown that in some cases they can reduce the thermodynamic stability.[26] The expansion of the genetic code has led to the biosynthetic incorporation of a wide range of noncanonical amino acids (NCAAs) into proteins.[27] In particular, fluorinated amino acids (FAAs) have been integrated into small coiled-coil proteins,[25, 28]a range of enzymes,[26, 29–32] and biomaterials.[33] Although incorporation of FAAs into a target protein can lead to enhanced function or stability, in some cases loss of activity or stability occurs, and further improvements to the artificial protein have been made by rational mutagenesis[30] and directed evolution strategies.[34]

Bond	Length	Van der Waals radius	Total size
C-H	1.09	1.2	2.29
C-F	1.35	1.7	2.82

Table 1.1: (A) physical properties of the C-F bond. (B) comparison of C-H and C-F bonds, van der Waals radius, and total size

1.1.5 Scope of Work

The primary goals of this work were to adapt Rosetta for phosphotriesterase. Overall, with incorporation of *p*FF into protein, we will be able evaluate the performance of scoring function. In advance, we would evaluate the shelf life and thermo-stability of phosphotriesterase.

1.2 Methods

1.2.1 General

All chemicals, reagents, and substrate were purchased from Sigma. T4 DNA ligase was purchased from Roche. DNA sequence was confirmed by Eurofins MWG Operon. 96-well plates were purchased from Thermo Fisher Scientific (Waltham, MA).

1.2.2 Recombinant Gene Construction

pQE30-S5 was used as described before.[35] The pQE30-104A plasmid was prepared with forward primers (5-GATGTGTCGACTGCCGATATCGGTCG-3, Fisher Scientific), reverse primers (5-CGACCGATATCGGCAGTCGACACA-3, Fisher Scientific). The PCR parameters were set as follow for 18 cycles: initial denaturation in 95 °C for 30 seconds, sequential denaturation in 95 °C for 30 seconds, annealing in

55 °C for 1 minute, and extension in 68 °C for 4 minutes. The mixture was then incubated 37 °C overnight with DpnI to digest methylated parent DNA strands, which lack the desired mutation. DNA sequence was further confirmed by Eurofins MWG Operon.

1.2.3 Protein Expression

Mutant and wild type plasmids were transformed into *E. coli* phenylalanine auxotrophic strains (AF-IQ cells).[5] Electroporation was done at 25 µF, 100 Ω, 2.5 kV (Biorad Gene Pulser II). Cells were plated on agar plates containing 200 $\mu g \cdot mL^{-1}$ ampicillin, 34 $\mu g \cdot mL^{-1}$ chloramphenicol. A Single colony was picked and grown in medium (M9 medium supplemented with 0.2 wt % glucose, 35 $mg \cdot L^{-1}$ thiamine, 1 mM $MgSO_4$, 0.1 mM$CaCl_2$, 200 $\mu g \cdot mL^{-1}$ ampicillin, and 34 $\mu g \cdot mL^{-1}$ chloramphenicol) with 20 $mg \cdot L^{-1}$ of 20 amino acids at 37 °C, 300 r.p.m. Afterwards, 250 mL of M9 medium for large-scale expression was innoculated 1:50 with an overnight culture. After optical density reached 1.0 at 600 nm, media shift was carried out by washing the cells three times with 0.9% 4 °C NaCl. Cells were then transferred to M9 minimal medium containing either 20 amino acids or 19 amino acids (-Phe). *p*FF-PTE and *p*FF-104A expression media were supplemented with and 3 mM of *p*FF and 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) to induce protein expression. 1 mM of $CoCl_2$ was added in each post-induction medium. After three hours incubation at 37 °C, 300 r.p.m., the cells were harvested and then resuspended with 20 mM Tris-HCl, 500 mM NaCl, 5 mM imidazole, 10% glycerol (pH 8.0) and 1 µM $CoCl_2$. Cell lysate was sonicated on ice for 1.5 minutes and then a clarification spin was performed (20, 000 g, 4 °C, 30 min). Clarified supernatants were loaded into a His Trap column (G.E Healthcare, Piscataway, NJ) using KTA FPLC purifier (G.E. Healthcare, Piscataway, NJ). Protein elution was generated using elution buffer B (20 mM Tris-HCl, 500 mM sodium chloride, 500 mM imidazole (pH 8.0)). The purified samples

were then transferred for buffer exchange using 12 L 20 mM phosphate buffer (pH 8.0). Dialyzed protein was subjected to kinetic assays immediately.

1.2.4 PyRosetta Design

Rosetta[2, 5] was used to generate a symmetric, *p*FF-incorporated PTE structure used by all simulations. The Holden and coworkers structure (PDB code: 1HZY) of wild type PTE was used as the input. In addition to the phenylalanine positions being mutated to *p*FF, three positions in the wild-type PTE sequence were mutated (K185R, D208G, and R319S) to generate *p*FF-PTE.[36] Mutations were made using the Rosetta *fixbb* application and were followed by side chain repacking and minimization. The amino acids directly interacting with the Co^{2+} ions are crucial in binding the necessary divalent cation for PTE activity,[3] so they were fixed in their native rotamers during repacking and minimization. PyRosetta, a python interface to the Rosetta libraries,[4] was used to make and characterize point mutations. Every *p*FF position was individually mutated into any natural amino acid minus phenylalanine. To simulate a mutation, a single *p*FF position would be mutated and neighboring amino acid within 10Å(as measured by Cα-Cα atom distance) was allowed to repack and minimize to accommodate the point mutation to fill in potential high-cost-energy voids or to supplement the hydrophobicity, polarity, or charge in the vicinity. For each *p*FF position, 500 decoys were generated. After the mutations were made, representative structures of each mutation were chosen based on the overall stability of the enzyme, reflected by the total score. The binding energy is the total energy minus the energy of both monomers separated by 1000Å. Point mutations were chosen based on the difference between relative total and predicted binding energies of the mutant and *p*FF-PTE sequence. As above, amino acids directly interacting with the Co^{2+} ions were fixed in their native rotamers during repacking and minimization. All Rosetta and PyRosetta calculations were done using the *score12* score function, and included

extra rotamer sampling, including the native rotamers.

1.2.5 Thermo-stability and Secondary Structure of Phosphotriesterase

1.2.5a Nano-DSC

DSC (Nano-DSC, TA instrument, USA) was preformed by using 600 µL ($0.1\,\mathrm{mg}\cdot\mathrm{mL}^{-1}$) of protein right after dialysis. Measurements were conducted at a scan rate of $1\,^{\circ}\mathrm{C}\cdot\mathrm{min}^{-1}$. Signals was blanked with buffer under the same condition. The observed diagram was then analyzed by using NanoAnalyze software.

1.2.5b Circular Dichroism

CD spectra were recorded on a JASCO J-815 Spectropolarimeter (Easton, MD) using Spectra Manager software. Temperature was controlled using a Fisher Isotemp Model 3016S water bath. Proteins concentrations were 10 µM in 20 mM phosphate buffer (pH 8.0). 20 mM phosphate buffer was used for blanking signals. To calculate ellipticities, the following formula was used(Eq. 1.1):

$$mrw = MRW(obs)/(10 * c * l) \tag{1.1}$$

where MRW is the mean residue weight of the specific phosphotriesterase, obs is the observed ellipticities (mdeg), l is the path length (cm), c is the concentration in µM. Spectra was recorded from 190 to 250 nm with a scan speed of $1\,\mathrm{nm}\cdot\mathrm{min}^{-1}$.

1.2.6 Enzyme Kinetics

The protein was diluted to a final concentration of 30 nM in 20 mM sodium phosphate (pH 8.0) by using the extinction coefficient $29\,280\,\mathrm{M}^{-1}\cdot\mathrm{cm}^{-1}$. Reactions were

monitored spectrophotometrically (Synergy H1, BioTek, Winooski VT) at 405 nm for paraoxon (coefficient = $17\,000\,\text{M}^{-1} \cdot \text{cm}^{-1}$). Reactions for paraoxon (13 to 104 µM) was done in 0.4% methanol. K_M and k_{cat} values were determined by a Lineweaver-Burk plot.[35] The equation used is shown below (Eq. 1.2):

$$\frac{1}{v} = \frac{K_M}{V_{max}} \times \frac{1}{S} + \frac{1}{V_{max}} \tag{1.2}$$

where S represents substrate concentration; K_M represents the substrate concentration at which the reaction rate is half of V_{max}. The data reported is the average of three trials and the error represents the standard deviation of those trials.

1.2.7 MALDI-TOF Mass Spectrometry

To determine level of *p*FF incorporation, 20 µL of purified PTE *p*FF-PTE, F104A, or *p*FF-104A was incubated with $12.5\,\text{ng} \cdot \mu\text{L}^{-1}$ of trypsin solution (in 50 mM of ammonium bicarbonate) at 37 °C overnight. 2 µL of 10% trifluoroacetic acid (TFA) was used to quench each reaction. Reaction was then purified with a C_{18} packed zip-tip (Millipore, Billerica, MA). Tips were wetted in 50% acetonitirile (ACN), equilibrated in 0.1% TFA, and eluted with 0.1% TFA in 75% ACN. Matrix was dissolved in $10\,\text{mg} \cdot \text{mL}^{-1}$ α-cyano-4-hydrocinnamic acid (CCA) in 50% ACN, 0.05% TFA. Theoretical trypsin digest were calculated from Peptide Mass (www.expasy.org/tools/peptide-mass.html). Samples were added to the matrix at a 1:1 ratio and spotted on MALDI plate. Five standards were spotted separately for calibration: angiotensin I (MW = $1295.69\,\text{g} \cdot \text{mol}^{-1}$), neurotensin (MW = $1671.92\,\text{g} \cdot \text{mol}^{-1}$), ACTH (1-17) (MW = $2092.09\,\text{g} \cdot \text{mol}^{-1}$), ACTH (18-39) (MW = $2464.20\,\text{g} \cdot \text{mol}^{-1}$), and ACTH (7-38) (MW = $3656.93\,\text{g} \cdot \text{mol}^{-1}$). Compass 1.4 for flex software was then used to analyze the MALDI spectra (www.bruker.com/).

1.3 Results and Discussion

1.3.1 Biosynthesis of Phosphotriesterase

The *p*FF-F104A variant and the *p*FF-PTE parent were biosynthesized by residue-specific incorporation with the phenylalanine auxotrophic *Escherichia coli* strain AFIQ.[6] As controls, the non-fluorinated counterparts, PTE and F104A, were expressed under conventional conditions. As expected, all four proteins exhibited good expression in the presence of phenylalanine or *p*FF. *p*FF-F104A and *p*FF-PTE exhibited 80% and 92% incorporation, respectively, as determined by MALDI-TOF mass spectrometry. Notably, purified yields of *p*FF-F104A were twofold higher than for *p*FF-PTE, thus indicating more soluble protein yield. (Figure 1-4)

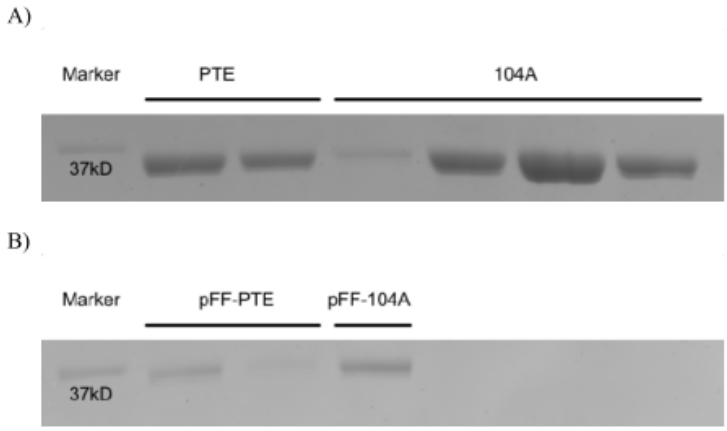

Figure 1-4: SDS-PAGE of purified proteins: (A) purified PTE and 104A, (B) purified *p*FF-PTE and *p*FF–104A.

1.3.2 Thermo-stability And Secondary Structure

Circular dichroism (CD) was performed to determine whether the mutation had an impact on the overall secondary structure and stability. Far-UV wavelength scans of *p*FF-F104A and *p*FF-PTE showed a double minimum at 208 nm and 222 nm (25 °C),

as expected for a $(\beta/\alpha)_8$-barrel protein, thus suggesting that the mutation did not affect the overall structure (Figure 1-6). Surprisingly, comparison of the non-fluorinated

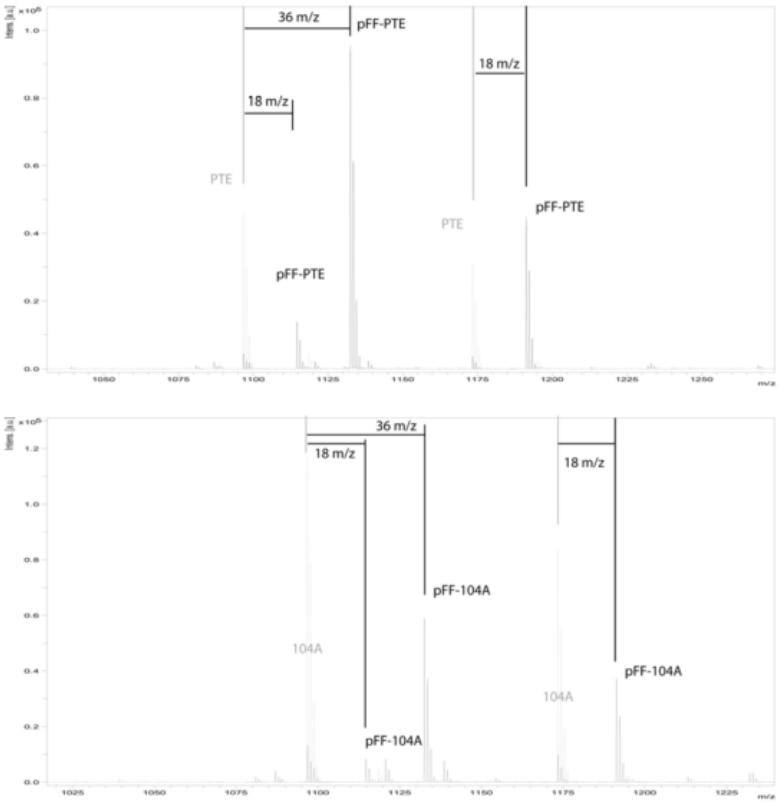

Figure 1-5: MALDI-TOF mass spectra of tryptic peptide fragments.

counterparts revealed that F104A was less structured than PTE (Figure 1-6). To assess the stability, differential scanning calorimetry (DSC) was performed (Figure 1-7). Upon heating the sample from 0 to 70 °C, *p*FF-PTE exhibited two transitions (T_m1: 42.0 ± 0.1 °C; T_m2 : 48.6 ± 0.2 °C); this is consistent with our previous studies.[35] This biphasic unfolding was also observed by Grimsley *et al.* in a study of organophosphorus hydrolase, and was attributed to the presence of a dimeric unfolded intermediate.[37] In contrast, *p*FF-F104A exhibited a single transition at 49.7 ± 0.2 °C, which was higher than both *p*FF-PTE values (by 7.7 and 1.1 °C). Remarkably, after heating, *p*FF-F104A retained the single T_m of 49.2 ± 0.1 °C, thus demonstrating regaining of structure after undergoing thermal unfolding. In the absence of *p*FF, F104A

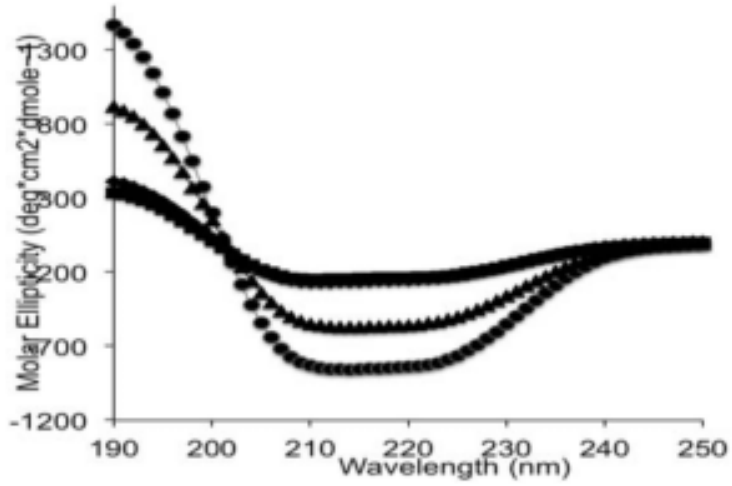

Figure 1-6: CD wavelength scans of PTE (circles), 104A (triangles), *p*FF-PTE (diamonds) and *p*FF-104A (squares).

demonstrated two transitions similar to *p*FF-PTE (Figure 2A), thus suggesting that fluorination was critical for stability. These data demonstrate the overall thermodynamic stability of *p*FF-F104A. Moreover, the data further suggest that the unfolding

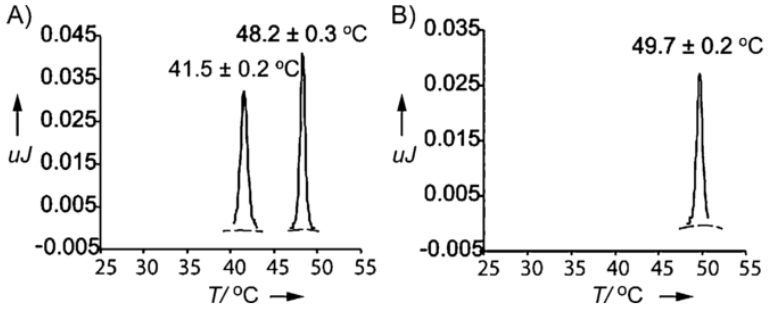

Figure 1-7: Differential scanning calorimetry thermograms of A) F104A and B) pFF-F104A.

model has been altered from a 3-state to 2-state transition, and that the energy requirement to attaining the unfolded intermediate was increased, thereby resulting in a more cooperative transition. Mutations resulting in this particular transformation have been observed for other proteins; for example, Fan *et al.* showed that removal

of a C-terminal domain of the oligomeric *E. coli* trigger factor protein resulted in the transformation of an otherwise n-state unfolding process to a distinct two-state unfolding process, indicative of pronounced stabilization of the native structure by interdomain interactions.[18] We propose that the *p*FF-F104A mutation might also be stabilizing the native structure of the overall protein (in effect the reverse of the mutation observed by Fan *et al.*).[38] That is, *p*FF-F104A unfolds cooperatively in a single step, concurrent with its dissociation into monomeric species. Although it was expected that the *p*FF-F104A mutation would have an effect on interdomain stability (such that neighboring residues would be allowed to repack and energy would be minimized between monomers at the dimer interface), the apparent stabilization of the native structure concluded from the 3-state to 2-state transition transformation through these new interdomain interactions were indeed unanticipated. Prior examples of this transformation exist in cases involving subdomains of similar stabilities or where strong coupling exists between subdomains.[39]

1.3.3 Enzymatic Kinetics of Phosphotriesterase

To assess function, we determined the MichaelisMenten kinetics of *p*FF-F104A, *p*FF-PTE, F104A, and PTE with paraoxon. At 25 °C, *p*FF-PTE exhibited the highest activity ($k_{cat}/K_M = 327\,000\ M^{-1} \cdot s^{-1}$; Table 1); *p*FF-F104A was slightly lower ($k_{cat}/K_M = 223\,000\ M^{-1} \cdot s^{-1}$); non-fluorinated PTE exhibited k_{cat}/K_M of $200\,000\ M^{-1} \cdot s^{-1}$,similar to those both fluorinated proteins; however, F104A was dramatically less active ($k_{cat}/K_M = 23\,000\ M^{-1} \cdot s^{-1}$). Thus, the fluorinated amino acids appear to be necessary for *p*FF-F104A activity. Proteins were then incubated at 35 °C, 45 °C and 55 °C for one hour, and then cooled to room temperature to determine residual activity. A decline in residual activity was observed for all proteins as a function of elevated temperature. *p*FF-F104A, which was designed to stabilize the fluorinated protein, exhibited 50% retention of activity at 55 °C (Figure 3A, Table 1). In contrast, at 55 °C,

*p*FF-PTE and PTE exhibited 24% and 23% initial activity, respectively; F104A exhibited a significant loss in activity at or above 45 °C (Figure 3A, Table 1). As *p*FF-F104A retained substantial activity after elevated temperatures, we then investigated the half-life of the activity. The parent *p*FF-PTE exhibited more than 50% loss of activity after three days, whereas the non-fluorinated PTE showed more than 50% activity reduction after seven days (Figure 3B, Table 2). Remarkably, *p*FF-F104A retained 66% activity after seven days. The non-fluorinated counterpart F104A failed to exhibit activity after one day. Together these data confirm that *p*FF-F104A is able to delay heat inactivation while maintaining function after one week.

protein		25 °C	35 °C	45 °C	55 °C
PTE	k_{cat}/K_M	2.00± 0.13	0.76 ± 0.11	0.72 ± 0.12	0.46 ± 0.18
	k_{cat}	2.1 ± 0.4	1.3 ± 0.1	1.4 ± 0.1	0.9 ± 0.1
F104A	k_{cat}/K_M	3.27 ± 0.11	2.42 ± 0.10	1.84 ± 0.21	0.80 ± 0.09
	k_{cat}	6.0 ± 1.1	5.6 ± 0.1	4.0 ± 1.1	2.0 ± 0.9
*p*FF-PTE	k_{cat}/K_M	0.23 ± 0.04	0.21 ± 0.03	n.a	n.a
	k_{cat}	0.01 ± 0.0	0.1 ± 0.0	n.a	n.a
*p*FF-F104A	k_{cat}/K_M	2.23 ± 0.15	1.94 ± 0.18	1.49 ± 0.20	1.11 ± 0.09
	k_{cat}	3.3 ± 0.3	3.3 ± 0.6	2.6 ± 1.0	2.0 ± 0.7

n.a = not available; k_{cat}/K_M: $\times 10^5 M^{-1} \cdot s^{-1}$; k_{cat}: s^{-1}.

1.3.4 Protein Design

Although methods enabling the biosynthesis of artificial proteins bearing NCAAs are abundant,[29] tools to help further improve the overall activity and stability are needed. Mutagenesis and evolutionary approaches have been employed successfully to identify variants with enhanced function; however, these rely heavily on testing or screening several to millions of constructs.[30, 34, 40] We demonstrate the use of

computational methods to identify a fluorinated protein variant that exhibits superior heat stability and half-life. Notably, the *p*FF-F104A variant is only functional in the fluorinated form, thus validating Rosetta-based design with *p*FF. This provides another useful tool for protein design and could be employed in conjunction with the above mentioned approaches.

References

1. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453,** 190–5 (May 2008).

2. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & André, I. Modeling symmetric macromolecular structures in Rosetta3. *PloS one* **6,** e20450 (Jan. 2011).

3. Korkegian, A., Black, M. E., Baker, D. & Stoddard, B. L. Computational thermostabilization of an enzyme. *Science (New York, N.Y.)* **308,** 857–860 (2005).

4. Leaver-Fay, A. *et al.* Scientific benchmarks for guiding macromolecular energy function improvement. *Methods in enzymology* **523,** 109–43 (Jan. 2013).

5. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology* **487,** 545–74 (Jan. 2011).

6. Drew, K. *et al.* Adding Diverse Noncanonical Backbones to Rosetta: Enabling Peptidomimetic Design. English. *Plos One* **8** (July 2013).

7. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. & Meiler, J. Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry* **49,** 2987–98 (Apr. 2010).

8. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein structure prediction using Rosetta. *Methods in enzymology* **383,** 66–93 (Jan. 2004).

9. Ghanem, E. & Raushel, F. M. *Detoxification of organophosphate nerve agents by bacterial phosphotriesterase* in *Toxicology and Applied Pharmacology* **207** (2005).

10. Soreq, H. & Seidman, S. *Acetylcholinesterase–new roles for an old actor.* 2001.

11. Yang, C.-Y., Renfrew, P. D., Olsen, A. J., Zhang, M., Yuvienco, C., Bonneau, R. & Montclare, J. K. Improved stability and half-life of fluorinated phosphotriesterase using Rosetta. eng. *Chembiochem: A European Journal of Chemical Biology* **15,** 1761–1764 (Aug. 2014).

12. Mahto, S. K., Howard, C. J., Shimko, J. C. & Ottesen, J. J. A reversible protection strategy to improve Fmoc-SPPS of peptide thioesters by the N-acylurea approach. *ChemBioChem* **12,** 2488–2494 (2011).

13. Cellitti, S. E. *et al.* In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *Journal of the American Chemical Society* **130,** 9268–9281 (2008).

14. Hassan, A. Q. *Site-specific incorporation of chemical probes into proteins for NMR* 2008.

15. Johnson, J. A., Lu, Y. Y., Van Deventer, J. A. & Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Current opinion in chemical biology* **14,** 774–80 (Dec. 2010).

16. Bain, J. D., Diala, E. S., Glabe, C. G., Wacker, D. A., Lyttle, M. H., Dix, T. A. & Chamberlin, A. R. Site-specific incorporation of nonnatural residues during in vitro protein biosynthesis with semisynthetic aminoacyl-tRNAs. *Biochemistry* **30,** 5411–5421 (1991).

17. Martoglio, B., Hofmann, M. W., Brunner, J. & Dobberstein, B. The protein-conducting channel in the membrane of the endoplasmic reticulum is open laterally toward the lipid bilayer. English. *Cell* **81,** 207–214 (Apr. 1995).

18. Eichler, J., Brunner, J. & Wickner, W. The protease-protected 30 kDa domain of SecA is largely inaccessible to the membrane lipid phase. *EMBO Journal* **16,** 2188–2196 (1997).

19. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science (New York, N.Y.)* **292,** 498–500 (2001).

20. Wang, L., Brock, A. & Schultz, P. G. Adding L-3-(2-naphthyl)alanine to the genetic code of E. coli. *Journal of the American Chemical Society* **124,** 1836–1837 (2002).

21. Kiick, K. L. & Tirrell, D. A. *Protein Engineering by In Vivo Incorporation of Non-Natural Amino Acids: Control of Incorporation of Methionine Analogues by Methionyl-tRNA Synthetase* 2000.

22. Döring, V., Mootz, H. D., Nangle, L. A., Hendrickson, T. L., de Crécy-Lagard, V., Schimmel, P. & Marlière, P. Enlarging the amino acid set of Escherichia coli by infiltration of the valine coding pathway. *Science (New York, N.Y.)* **292,** 501–504 (2001).

23. Kast, P. & Hennecke, H. Amino acid substrate specificity of Escherichia coli Phenylalanyl-tRNA synthetase altered by distinct mutations. *Journal of Molecular Biology* **222,** 99–124 (1991).

24. Biffinger, J. C., Kim, H. W. & DiMagno, S. G. The polar hydrophobicity of fluorinated compounds. *Chembiochem : a European journal of chemical biology* **5,** 622–7 (May 2004).

25. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F. & Tirrell, D. A. Fluorinated Coiled-Coil Proteins Prepared In Vivo Display Enhanced Thermal and Chemical Stability This work was supported by a grant from the U.S. Army Research Office. Y. Tang is supported by a Whitaker Graduate Research Fellowship. We thank Dr. Gary Hat. *Angewandte Chemie (International ed. in English)* **40,** 1494–1496 (Apr. 2001).

26. Panchenko, T., Zhu, W. W. & Montclare, J. K. Influence of global fluorination on chloramphenicol acetyltransferase activity and stability. *Biotechnol Bioeng* **94,** 921–930 (2006).

27. Voloshchuk, N. & Montclare, J. K. Incorporation of unnatural amino acids for synthetic biology. *Mol Biosyst* **6,** 65–80 (Jan. 2010).

28. Montclare, J. K., Son, S., Clark, G. A., Kumar, K. & Tirrell, D. A. Biosynthesis and stability of coiled-coil peptides containing (2S,4R)-5,5,5-trifluoroleucine and (2S,4S)-5,5,5-trifluoroleucine. *Chembiochem* **10,** 84–86 (Jan. 2009).

29. Voloshchuk, N., Zhu, A. Y., Snydacker, D. & Montclare, J. K. Positional effects of monofluorinated phenylalanines on histone acetyltransferase stability and activity. *Bioorg Med Chem Lett* **19,** 5449–5451 (Sept. 2009).

30. Voloshchuk, N., Lee, M. X., Zhu, W. W., Tanrikulu, I. C. & Montclare, J. K. Fluorinated chloramphenicol acetyltransferase thermostability and activity profile: improved thermostability by a single-isoleucine mutant. *Bioorg Med Chem Lett* **17,** 5907–5911 (Nov. 2007).

31. Mehta, K. R., Yang, C. Y., Montclare, J. K., KR Mehta, C. Y. & Montclare., J. Modulating substrate specificity of histone acetyltransferase with unnatural amino acids. en. *Mol Biosyst* **7,** 3050–3055 (Nov. 2011).

32. Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C. & Mehl, R. A. Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nature protocols* **2,** 2601–7 (Jan. 2007).

33. Yuvienco, C., More, H. T., Haghpanah, J. S., Tu, R. S. & Montclare, J. K. Modulating supramolecular assemblies and mechanical properties of engineered protein materials by fluorinated amino acids. *Biomacromolecules* **13,** 2273–2278 (Aug. 2012).

34. Montclare, J. K. & Tirrell, D. A. Evolving proteins of novel composition. *Angew Chem Int Ed Engl* **45,** 4518–4521 (2006).

35. Baker, P. J. & Montclare, J. K. Enhanced refoldability and thermoactivity of fluorinated phosphotriesterase. *Chembiochem* **12,** 1845–1848 (Aug. 2011).

36. Roodveldt, C. & Tawfik, D. S. Directed evolution of phosphotriesterase from Pseudomonas diminuta for heterologous expression in Escherichia coli results in stabilization of the metal-free state. *Protein Engineering, Design and Selection* **18,** 51–58 (Jan. 2005).

37. Grimsley, J. K., Scholtz, J. M., Pace, C. N. & Wild, J. R. Organophosphorus hydrolase is a remarkably stable enzyme that unfolds through a homodimeric intermediate. English. *Biochemistry* **36,** 14366–14374 (Nov. 1997).

38. Fan, D.-J., Ding, Y.-W., Pan, X.-M. & Zhou, J.-M. Thermal unfolding of Escherichia coli trigger factor studied by ultra-sensitive differential scanning calorimetry. *Biochimica et biophysica acta* **1784,** 1728–34 (Nov. 2008).

39. Tsytlonok, M. & Itzhaki, L. S. The how's and why's of protein folding intermediates. *Archives of biochemistry and biophysics* **531,** 14–23 (Mar. 2013).

40. Yoo, T. H., Link, A. J. & Tirrell, D. A. Evolution of a fluorinated green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 13887–90 (Aug. 2007).

Chapter 2

Effects of Phenylalanines Outside Dimer Interface of Phosphotriesterase

2.1 Introduction

2.1.1 Phosphotriesterase

2.1.2 Side-chain effects

2.2 Methods

2.2.1 General

All chemicals, reagents, and substrate were purchased from Sigma. T4 DNA ligase was purchased from Roche. DNA sequence was confirmed by Eurofins MWG Operon. 96-well plates were purchased from Thermo Fisher Scientific (Waltham, MA)[1].

2.2.2 Protein preparation methods

2.2.2a Biosynthesis

In anticipation for the need of large quantities of protein mass for the ... delete comtent of the supernatant and storage at $-20\,^{\circ}\mathrm{C}$.

2.2.2b Protein purification

All solutions used in the extraction and purification of recombinant proteins ... delete content 5 CVs of buffer prior to each injection.

2.2.2c Endotoxin removal

Protein elutions were separated from pyrogenic lipopolysaccharides using a ... delete content Chromogenic Endotoxin Quantitation kit (Thermo Scientific).

2.2.2d Protein concentration

Purified protein product was assayed for concentration by way of a Thermo

2.2.3 In vitro methods

2.2.3a Cell culture methods

Chondrocytes were treated with formulations based on COMPcc and BMS493 to ... delete content of $1\,\mu\mathrm{mol}\cdot\mathrm{L}^{-1}$ COMPcc and BMS493.

2.2.3b Reverse transcription-polymerase chain reaction (PCR) and real-time PCR analysis

RNA was isolated from cell cultures using an RNeasy Mini ... delete content levels were calculated according to the equation $x = s^{-\Delta Cq}$, where $\Delta Cq = Cq_{exp} - Cq_{18S}$.

2.2.4 Biophysical methods

2.3 Results

2.3.1 Scaling biosynthetic methods

Expression of COMPcc was carried out in auto-inducing media, encompassing a . . . delete content typical in inducible expression systems.

Auto-induction growth can be sustained in baffled shaker flasks, according to . . . delete content to purification.

2.3.2 Effects of COMPcc on BMS493 therapy

. . . lots content delete here. . . potential for COMPcc to behave as an inhibitor of hypertrophic differentiation.

2.3.3 Endotoxin levels in COMPcc protein

Endotoxin levels of the protein were measured using a limulus amebocyte lysate catabolic events reported in Section 2.3.2.

2.4 Discussion

2.4.1 Contamination of recombinant COMPcc with endotoxins

The importance of endotoxin removal from recombinant protein preparations has . . . lots content deleted . . . COMPcc may provide additional benefit to the delivery

of RAR/RXR antagonists and inverse agonists as demonstrated by early evidence of down-regulating hypertrophic gene expression and terminal differentiation.

2.4.2 Meso-scale features of COMPcc promote intra-articular delivery applications

small molecules and macromolecules, including Paclitaxel, Mefoxin,

2.4.3 Future work

These preliminary results suggest that the COMPcc does not significantly improve the therapeutic effects of BMS493. The signal from these experiments suffers, however, from potential noise originating from the presence of endotoxins in the sample, possibly eliciting similar catabolic gene expression to what is observed. While dosage experiments were performed in this study of COMPcc and COMPcc that had undergone endotoxin removal procedures, additional control samples can improve the confidence of these results. While LPS have been shown to induce catabolic gene expression, known standards of LPS may be administered to chondrocytes in future studies to gauge actual response from a positive control. In addition, purified lysate from non-transformed BL21 (DE3) *E.coli* that undergoes endotoxin removal procedures may be used and considered as a negative control sample group toward the testing of COMPcc alone.

these conditions. A balance must be stricken between the solvation of BMS493 (or any other hydrophobic/non-polar small molecule compound) and the (molecular and meso-scale) structural integrity of the protein-based carrier.

References

1. Yang, C.-Y., Renfrew, P. D., Olsen, A. J., Zhang, M., Yuvienco, C., Bonneau, R. & Montclare, J. K. Improved stability and half-life of fluorinated phosphotriesterase using Rosetta. eng. *Chembiochem: A European Journal of Chemical Biology* **15,** 1761–1764 (Aug. 2014).