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

List of Figures

List of Tables

Chapter 1

Improved Stability and Half-Life of Fluorinated Phosphotriesterase Using Rosetta

1.1 Introduction

1.1.1 Rosetta and Protein Engineering

Computational tools are widely used for protein engineering [ref]. Rosetta suite was first developed in University of Washington [ref]. Baker et al. adapted this suite for prediction of three: dimensional structure of proteins. 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 design enzymes, including some containing unnatural amino acids in the active sites. (see Figure 1-1)

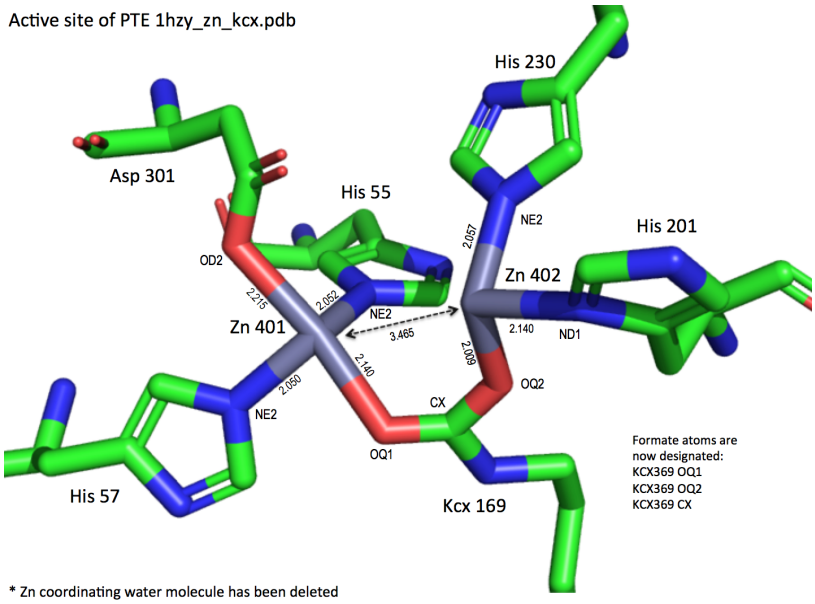

Figure 1-1: Rosetta Test

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.[1] 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.[2] Army also adapted this protein for chemical weapons neutralization. [3]

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)[4], in vivo and in vitro site-specific incorporation, 16 and residue-specific incorporation (Fig. 1)1d, 17. 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. Schultz and their coworkers18 have developed a general approach for the in vitro synthesis of proteins. 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 functions19.

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 coworkers21. 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 analogs21a.

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 promotor22. 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 aaRS23. 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.24

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 hydrophobzgicity 26 as compared to their hydrogenated counterparts. As we compare the structure of fluorocarbon groups, t(Bh)e C-F bond is highly dipolar while the hydrocarbon is less. The C-F bond is roughly 0.24 Å longer than C-H bond26 (Table 1). While in some cases the global replacement of hydrophobic amino acids with fluorinated analogs has led to the stabilization of protein structure27, it has all been shown that in some cases they can reduce the thermodynamic stability28.

1.1.5 Scope of Work

The primary goals of this work were to done 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 thermostability 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.[5] 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 g/mL ampicillin, 34 g/mL chloramphenicol. A Single colony was picked and grown in medium (M9 medium supplemented with 0.2 wt % glucose, 35 mg/L thiamine, 1mM $MgSO_4$, 0.1 mM $CaCl_2$, 200 g/mL ampicillin, and 34 g/mL chloramphenicol) with 20 mg/L 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. 1mM 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 µmol · L^{-1} $CoCl_2$. 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 Thermo-stability and Secondary Structure of Phosphotriesterase

1.2.4a Nano-DSC

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

1.2.4b 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 M 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: mrw = MRW(obs) / (10 * c * l) 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 M. Spectra was recorded from 190 nm to 250 nm with a scan speed of 1 nm/min.

1.2.5 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 M^{-1} cm^{-1}. Reactions were monitored spectrophotometrically (Synergy H1, Bio-Tek, Winooski VT) at 405 nm for paraoxon (coefficient = 17,000 M^{-1} cm^{-1}). Reactions for paraoxon (13 104 M) was done in 0.4% methanol. KM and kcat values were determined by a Lineweaver-Burk plot (1/v vs 1/[S]).[5] The equation used is shown below: 1/v = (KM/ VMax) ? (1/[S]) + 1/VMax S2 where [S] represents substrate concentration; KM represents the substrate

concentration at which the reaction rate is half of Vmax. The data reported is the average of three trials and the error represents the standard deviation of those trials.

1.2.6 MALDI-TOF Mass Spectrometry

To determine level of *p*FF incorporation, 20 µL of purified PTE pFF-PTE, F104A, or *p*FF-104A was incubated with 12.5 ng/L of trypsin solution (in 50 mM of ammonium bicarbonate) at 37 °C overnight. 2 L 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 mg/mL 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 g/mol), neurotensin (MW = 1671.92g/mol), ACTH (1-17) (MW = 2092.09 g/mol), ACTH (18-39) (MW = 2464.20 g/mol), and ACTH (7-38) (MW =3656.93 g/mol). 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.

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 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 S3). Surprisingly, comparison of the non-fluorinated counterparts revealed that F104A was less structured than PTE (Figure S3). To assess the stability, differential scanning calorimetry (DSC) was performed (Figure 2). 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; Table S1); this is consistent with our previous studies.[5] 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.[17] 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); Figure 2B, Table S1). 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 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 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 tran-

sition. 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.*).[18] That is, *p*FF-F104A unfolds cooperatively in a single step, concurrent with its dissociation into monomeric species. Although it was expected that the pFF-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 transformationthrough 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.[19]

1.3.3 Enzymatic Kinetics of PTE

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 (kcat/KM = 327,000 s-1 M-1; Table 1); *p*FF-F104A was slightly lower (kcat/KM=223,000 s-1 M-1); non-fluorinated PTE exhibited kcat/KM of 200,000 s-1 M-1, similar to those both fluorinated proteins; however, F104A was dramatically less active (kcat/KM = 23,000 s-1 M-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 pFF-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.

1.3.4 Protein Design

Although methods enabling the biosynthesis of artificial proteins bearing NCAAs are abundant,[1] 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.[3c, 5, 20] 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. Ghanem, E. & Raushel, F. M. *Detoxification of organophosphate nerve agents by bacterial phosphotriesterase* in *Toxicology and Applied Pharmacology* **207** (2005).

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

3. 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).

4. 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).

5. Baker, P. J. & Montclare, J. K. Enhanced refoldability and thermoactivity of fluorinated phosphotriesterase. *Chembiochem : a European journal of chemical biology* **12,** 1845–8 (Aug. 2011).