Biosynthetic Approaches to Protein Engineering Using Fluorinated Amino Acids

Thesis by

Soojin Son

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended July 27, 2006)

To My Family

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to everyone who has helped me throughout this unforgettable chapter of my life. First, I thank my advisor, David Tirrell, for his patience, guidance, and continuous support. His optimistic outlook and genuine encouragement allowed me to maintain my confidence throughout the ups and downs of my Ph.D. career. I would also like to thank my thesis committee- Anand Asthagiri, Frances Arnold, and Dennis Dougherty- for its valuable advice and input.

I have been very fortunate to work with a talented and fun group of colleagues. I want to thank Yi Tang and Pin Wang, who helped me to get started in lab. To all the Tirrell group members past and present, thank you for the fruitful discussions, creative insights, and of course, numerous happy hours. I'm grateful for the friendships that I was able to establish with many of you, especially Jin, Mandy, Kimberly, and Yvette, and I wish all of you the best in the future.

Outside of lab, I also met many wonderful people who have provided balance in my life and the chance to recharge. To my past roommates, Tomtor and Tracey, you two have made my transition to Caltech, both academically and personally, much less scary and far more enjoyable. I have also established valuable friendships with many people in the Korean community, especially Koun, SangAh, Peter, Jangwook oppa, and Hern-ee oppa. I was also lucky enough to have my college friends Grace, Tina, and Saujin nearby who were there to listen, support, and let me go crazy once in awhile. I truly appreciate all the memories I've shared with all of you and I hope our friendships can continue to grow in the future.

The most constant and powerful source of strength has always been my family. To my sister, Hyunjin, you have been my biggest cheerleader and my best friend. Thanks for always being there for me, especially through difficult times. To my parents, who have instilled in me the importance of always giving my best and never quitting, I would not be here today if it weren't for you. Thanks for your unending love and support.

Finally, the biggest accomplishment of my Caltech career has been meeting and marrying my husband, Jongwon. Oppa, I don't have enough words to thank you for everything you have given me to make this work possible. Thank you so much for your friendship, encouragement, and love. You've believed in me more than I believed in myself. From tutoring me in Statistical Mechanics to staying up with me through long nights of experiments, I share this thesis with you.

ABSTRACT

Non-canonical amino acids provide a new set of building blocks and a potential route to new chemistries that extend beyond those achieved through the 20 common amino acids. Among various methods developed for this purpose, residue-specific incorporation *in vivo* has the potential to economically engineer proteins of various sizes with altered physical and chemical behavior. The work presented in this thesis explored the incorporation of fluorinated aliphatic residues and their effect on protein structure, stability, and function.

In Chapter 2, two fluorinated amino acids, 5,5,5-trifluoroisoleucine (5TFI) and 4,4,4-trifluorovaline (4TFV), were incorporated *in vivo* into mutant GCN4 proteins. Both 5TFI and 4TFV showed replacement levels higher than 88%, as indicated by quantitative amino acid and MALDI-MS analysis. The incorporation of 5TFI into **a**-positions of the coiled-coil protein raised the thermal denaturation temperature (T_m) by 27 °C from that of its isoleucine counterpart. However, when valines were replaced by 4TFV in the same positions, T_m only increased by 4 °C. Similar trends were observed in response to chemical denaturation by guanidine hydrochloride; $\Delta\Delta G_{folding}$ upon incorporation of 5TFI and 4TFV was -2.1 and -0.3 kcal/mol, respectively. Secondary and higher order structures as well as biological activity were retained in the presence of both 5TFI and 4TFV. These results indicate that, even when introduced into the same positions within the protein, the effect of fluorination differs depending on which amino acid is fluorinated.

In Chapter 3, the stereochemical effects of 5,5,5-trifluoroleucine (5TFL) were studied using (2*S*, 4*R*)-5',5',5'-trifluoroleucine and (2*S*, 4*S*)-5',5',5'-trifluoroleucine. The results from *in vitro* activation assays correlated well with efficiency of their incorporation *in vivo*. The (2*S*, 4*S*) isomer, whose k_{cat}/K_m was100-fold lower than that of leucine, was incorporated at high levels, with 91% replacement of the encoded leucine residues in a *de novo* engineered coiled-coil protein A1. The (2*S*, 4*R*) isomer exhibited 9-fold lower k_{cat}/K_m than the (2*S*, 4*S*) isomer, resulting in a slightly lower level of incorporation, 80%. The secondary structure of A1 was undisturbed upon the incorporation of either isomer and their impact on thermostability was similar, with an increase of 11 °C in T_m as compared to that of A1 containing leucine. However, the equimolar mixture of A1 containing (2*S*, 4*S*)-TFL and A1 containing (2*S*, 4*R*)-TFL displayed a further increase in T_m of 3 °C. Although this further enhancement in thermostability was modest, it may be attributed to the ability of the coils to pack more compactly into dimers due to the stereochemical differences.

In Chapter 4, laboratory evolution was utilized to recover the catalytic activity of chloramphenicol acetyltransferase (CAT) after replacement of isoleucine by 5TFI. Upon global incorporation of 5TFI into CAT, the catalytic efficiency, k_{cat}/K_m, was reduced by more than 2-fold, from $10.2 \pm 0.8\ \mu M^{-1}$ to $3.9 \pm 0.5\ \mu M^{-1}\ min^{-1}$. Four rounds of random mutagenesis, enrichment, and screening were performed, yielding a 7-fold fluorinated mutant, tfi-G4, whose activity in fluorinated form was 2.8-fold higher than that of the fluorinated parent enzyme, tfi-WT. The total number of isoleucines decreased only by one in the 7-fold mutant, and the gap in activity between the hydrogenated (ile-G4) and fluorinated (tfi-G4) forms narrowed. Despite similar secondary structure, the

incorporation of fluorinated amino acids decreased the stability of CAT for both the wild-type and G4 pairs based on both functional and structural analysis. Fluorinated forms were more sensitive towards thermal and chemical denaturation. However, both forms of G4 enzymes had increased stabilities as compared to their wild-type counterparts. This resulted in tfi-G4 exhibiting similar thermostability as that of ile-WT. Although structural changes were noted at both high and low pH, the pK_a of the catalytically essential histidine (His-193) was not affected by the incorporation of 5TFI. Based on these results, the incorporation of 5TFI appears to be adversely affecting protein folding, which resulted in decreased activity and stability. However, laboratory evolution effectively recovered these losses and yielded a fluorinated enzyme that performed similarly to the wild-type.

Finally, in Chapter 5, the effect of fluorination on sensitivity to proteolytic degradation was explored. GCN4 proteins containing either 5TFI (tfi-INL) or 4TFV (tfv-VNL) as well as CAT proteins containing either 5TFI (tfi-G2) or 5TFL (tfl-L2A1) were treated with two proteases, trypsin and elastase. As evidenced by gel electrophoresis and densitometry analysis, the half life of tfi-INL in the presence of elastase was 4 times that of its hydrogenated counterpart INL. The increased resistance was less significant upon incorporation of 4TFV as well as in response to trypsin. The opposite trend was observed in the analysis of CAT mutants. Both of the fluorinated CAT mutants were more susceptible to elastase and trypsin as compared to their hydrogenated counterparts. Upon incorporation of fluorinated amino acids, two factors that affect proteolytic degradation are altered, the stability of the protein as well as

substrate recognition and hydrolysis by the protease. It appears that changes in both of these factors contributed to the observed rates of proteolysis.

The work explored in this thesis has expanded our understanding of fluorinated amino acids and their effects on proteins. Based on these insights, we are continuing to expand the use of fluorinated amino acids (and other non-canonical amino acids) to control protein structure, function, and stability.

TABLE OF CONTENTS

LIST OF TABLES

Number *Page*

LIST OF FIGURES

Chapter 1

Protein Engineering through *In Vivo* Incorporation of Non-Canonical Amino Acids

1.1 Biosynthesis of Polymeric Materials

Synthetic polymer chemistry is a widely used method for creating a variety of materials with interesting structural and chemical properties. The diversity and versatility of synthetic polymers arise from the broad spectrum of monomers from which they are formed. However, polymers lack structural uniqueness. More specifically, they are characterized by heterogeneity in chain length and conformation as well as stereochemistry. Nature, through billions of years of evolution, has refined the task of precise structural control over macromolecules through the method of template-directed polymerization. Proteins are the products of nature's polymer factory and they form the basis for the host organism's survival. They perform complex chemical reactions; provide organisms with structural integrity; function in transport of mass, energy, and information; and frequently exhibit performance properties superior to man-made materials.[1] Fortunately, recombinant DNA technology has allowed us to tap into this robust set of machinery to generate improved materials with control both on microscopic and macroscopic levels.

Through this approach of protein engineering, our laboratory has been successful in developing protein-based materials with monodispersity and stereoregularity. In general, our method of material synthesis involves the insertion of a chemically synthesized piece of DNA that encodes the desired pattern of amino acids into a cloning plasmid. Once the necessary ligations of oligonucleotides are completed, the sequenced artificial gene is inserted into the expression plasmid and transformed into an *E.coli* expression host. After the accumulation of biomass, expression of the artificial protein is turned on through an inducible promoter. This method has yielded, among others,

smectic liquid-crystalline structures with exact layer spacing,[2] hydrogels that undergo pH or temperature-controlled reversible gelation,[3] and artificial extracellular matrix proteins that exhibit biological signaling capabilities.[4, 5]

1.2 Incorporation of Non-Canonical Amino Acids

Despite a number of advantages that recombinant DNA technology has over the traditional chemical synthesis of polymeric materials, one main drawback lies in the limited number of naturally existing amino acids, the monomeric building blocks of proteins. In order to expand the scope of protein engineering through the introduction of novel functional groups, our laboratory has been exploring a variety of non-canonical amino acids for incorporation into proteins using nature's machinery. Technology has been available to synthesize peptides through solid-phase multistep synthesis outside the realm of biological systems. However, the peptides are limited to a length of around 50 amino acids, and therefore most of the interesting protein materials cannot be synthesized through this method. Another available technology for the incorporation of artificial amino acids is through an *in vitro* expression system utilizing chemical aminoacylation methods.[6-10] One of these methods uses chemically acylated suppressor tRNAs to insert the artificial amino in response to a stop codon in place of the original codon encoding the residue of interest.[9, 10] Another approach introduces four and even five-base codons that are read by chemically acylated tRNAs.[11-14] Since these methods circumvent the use of an aminoacyl-tRNA synthetase in charging of the analog to the tRNA, they support the incorporation of a wide variety of non-canonical amino acids that are structurally divergent from the natural counterpart. On the other hand, since the tRNAs cannot be

recharged through the cell's machinery, a stoichiometric amount of the chemically acylated tRNAs is required for translation. Worsened by this constraint, these *in vitro* approaches suffer in protein yield, which is usually within the microgram range, motivating us to explore other alternatives.

Our laboratory, along with several others, has developed an *in vivo* method for the residue-specific incorporation of non-canonical amino acids.[15-17] A key component of this method is the *E. coli* auxotroph, *i.e.,* bacterial hosts that have lost the ability to synthesize the amino acid(s) of interest. Since these bacterial strains rely on the growth media for obtaining the deficient amino acid (s), one can control the set of amino acids that is available for the translation of proteins. While the cells are growing up to a sufficient cell density, the medium contains all 20 amino acids to allow a healthy rate of growth. However, through a medium shift procedure prior to induction, we provide the cells with only 19 amino acids and an analog for the deficient amino acid. If the cell's machinery, most importantly the corresponding aminoacyl-tRNA synthetase that normally charges the tRNA with the amino acid, remains functional towards the analog, incorporation of the analog into the target protein is possible. While some amino acids are incorporated efficiently without any other modifications to the system, other analogs require multiple copies of the aminoacyl-tRNA synthetase and some require mutations in/near the active site. As shown in Figure 1.1, research in our laboratory has demonstrated the replacement of methionine, leucine, isoleucine, proline, and phenylalanine with their various analogs.[18-25]

1.2.1 Incorporation of Fluorinated Analogs

Fluorocarbons are used widely in industrial, medical, and home applications in the forms of plastics, coatings, refrigerants, pharmaceuticals, and agrochemicals, to name a few. However, fluorine is extremely rare in biological molecules. Since the first fluorinated natural product was identified more than 50 years ago, only about a dozen fluorinated compounds have been found in nature.[26] Correspondingly, there are no canonical amino acids that contain fluorine groups. For protein engineers, several unique features of fluorocarbons, rendering them attractive for the above uses, have inspired efforts to incorporate fluorine into proteins.

First, fluorine is the most electronegative element, and therefore the electronic properties of the C-F bond are quite different from those of a C-H bond. While the C-H bond has a relatively small dipole, the dipole of the C-F bond is much larger and opposite in direction. Similarly, fluorine exerts a strong inductive effect that is sufficient to perturb the reactivity of atoms several bonds away. Since polarity strongly influences acidity or basicity of nearby functional groups, the pK_a can be shifted up to several log units.[27] In addition, organofluorine compounds possess the ability to act as hydrogen-bond acceptors.[28] Since hydrogen bonds are indispensable features in protein structure, and thus their biological function, the ability to incorporation of fluorinated amino acids has the potential to be a powerful protein engineering tool.

Another important characteristic of fluorocarbons is their increased hydrophobic character. Many fluorinated organic compounds display lower solubility in water as compared to their hydrogenated counterparts. The increased hydrophobicity can also be evidenced by partitioning measurements. The partition constant for the trifluoromethyl

group is 1.07, which is more than twice that of a methyl group.[29] At the same time, perfluorinated molecules also exhibit poor solubility in hydrocarbon solvents. When water, hexane, and perfluorohexane are mixed, the result is three separate layers, each compound being mutually immiscible with the other two.[30] Therefore, based on this fluorous effect, perfluorocarbons are characterized as fluorophilic. The well known polymer Teflon takes advantage of the fluorous effect, preventing the interaction with both hydrophilic and lipophilic compounds. The increased hydrophobicity as well as fluorophilicity could have significant effects on protein folding and protein-protein interaction.

Fluorine also has the potential to serve as a probe in spectroscopic analysis of protein structure and dynamics. The aforementioned lack of fluorine in biological molecules allows the use of ^{19}F NMR spectroscopy as a powerful and sensitive tool in protein analysis. With no naturally occurring background and a large range of chemical shift values, even minor differences in the environment of the ^{19}F label can be detected.[31] For example, domain-domain interactions in the *E. coli* chaperone protein PapD[32] and conformational changes in the chromophore of a GFP variant[33] were analyzed through ^{19}F NMR using 4-, 5-, and 6-fluorotryptophan as probes.

Another crucial feature of fluorine is its size; fluorine atoms are extremely small, with a van der Waals radius (1.35 Å) closely resembling that of a hydrogen atom (1.20 Å). Although the C-F bond is 0.4 Å longer than a C-H bond, the substitution of a fluorine for a hydrogen is very often sufficiently conservative, as evidenced by the recognition of various fluorinated amino acid analogs by their cognate amino acyl-tRNA synthetases.[22, 25, 31, 34-36] This feature, in a practical sense, is one of the most important

characteristics of fluorinated amino acids that make them attractive for *in vivo* incorporation into proteins.

Successful *in vivo* incorporation of fluorinated analogs using endogenous synthetases has been shown to be possible for aromatic as well as aliphatic residues. The first reported examples involved the incorporation of monofluorinated phenylalanine and tryrosine analogs into *E. coli*.[37, 38] A few years later, Rennert and coworkers demonstrated nearly quantitative substitution of leucine by 5,5,5-trifluoroleucine (5TFL).[39] More recently, Duewel *et al.* showed that both low (31%) and high (70%) levels of trifluoromethionine incorporation at methionine sites are possible in lysozyme through the use of an auxotrophic *E. coli* strain and the media shift procedure.[35] This protein was further analyzed using ^{19}F NMR. In our laboratory, Yoshikawa and coworkers showed 95-100% replacement of phenylalanine with p-fluorophenylalanine (pfF) in repetitive protein -$[(Ala\text{-}Gly)_3.pfF.Gly]_{13}$.[40] Shortly thereafter, Tang and coworkers demonstrated *in vivo* incorporation of 5TFL with up to 92% efficiency into the coiled-coil peptide domain.[36] Tang followed this work with the incorporation of hexafluoroleucine into a similar peptide using *E. coli* hosts that over-expressed LeuRS and therefore exhibited increased LeuRS activity.[18] Most recently, Wang and others have shown the incorporation of 5,5,5-trifluoroisoleucine (5TFI, via endogenous IleRS)[25] and (2S,3R)-4,4,4-trifluorovaline (via over-expression of IleRS or ValRS)[41] into murine DHFR in *E. coli*.

Our interest in incorporating fluorinated amino acids into proteins stemmed from the hypothesis that we could access some of the previously described properties of fluorine-containing compounds and polymers. Initial confirmation of this hypothesis

came when genetically engineered protein films containing fluorinated leucines were shown to support a hexadecane contact angle similar to that of Teflon (70°), an increase of 53° from that of the hydrogenated counterpart (unpublished work). Further work demonstrated that the increased hydrophobicity of fluorinated amino acids could enhance the strength of hydrophobic interactions that serve as the main driving force for protein folding and protein-protein interaction. In this work, Tang and coworkers chemically synthesized GCN4-p1,[42] a leucine zipper peptide from a yeast transcription factor, using 5TFL in place of leucine. The result was an elevation in thermal denaturation temperature of 13 °C. A similar trend was observed when 5TFL was incorporated *in vivo* into *de novo* engineered A1,[3] while the incorporation of hexafluoroleucine through over-expression of LeuRS gave rise to a 22°C increase in melting temperature.[18, 36]

The fluorophilic nature of fluorinated molecules was also captured in peptides through the use of hexafluoroleucine.[43, 44] A leucine zipper peptide system with hydrophobic cores composed entirely of leucine or 5,5,5,5',5',5'-α-*S*-hexafluoroleucine was chemically synthesized. At equilibrium, the homodimeric assembly was greatly preferred over the heterodimer form as indicated by a disulfide-exchange assay. The authors suggested that the instability of the heterodimer and the hyperstability of the fluorinated dimer provided the driving force for preferential homodimer formation. This study clearly demonstrates the ability to engineer specific protein-protein interaction through the incorporation of fluorinated amino acids.

The effects of fluorination on the biological activity of proteins are also of great interest. Upon incorporation, changes in activity could arise because fluorine produces a global change in protein structure or conformational dynamics. It may also be that a

particular aspect of catalysis is altered by the presence of fluorine near the active site. As expected, the change in activity varies widely, from no effect to both an increase and a decrease in activity, depending on the fluorinated amino acid, target protein, and the position(s) of incorporation.[31, 34] For example, when 5TFL was incorporated in GCN4 peptides, the DNA binding activity as well as specificity were not affected.[45] Similarly, when 5TFI was incorporated into murine interleukin-2, the same maximal response as the wild-type was observed.[25] On the contrary, the global replacements of tyrosine by 2- and 3-fluorotyrosine in mutants of *Aequorea victoria* green fluorescent proteins resulted in altered biophysical properties.[46] As evidenced by these studies, the impact of fluorinated amino acids on protein function is difficult to predict due to complexities still not well understood.

1.3 Directed Evolution as a Complementary Tool

The effects of non-canonical amino acids are complex and in some cases can lead to the perturbation of one or more properties. Therefore, a general technique, used in conjunction with non-canonical amino acids, that recovers the lost property could enhance the versatility of non-canonical amino acids. Among the numerous protein engineering methods, directed evolution is one of the most effective and generalizable.[47, 48] Mirroring natural evolution, directed evolution aims to generate proteins with new or improved properties by screening or selecting for the desired function from a large pool of protein variants.[49-51] By exploring the sequence space composed of 20 natural amino acids, directed evolution has been used successfully to engineer proteins with enhanced activity, thermostability, altered substrate specificity, and improved protein function in

non-natural applications.[52-55] Extending laboratory evolution to non-canonical amino acids could yield functional proteins outfitted with novel chemistries.

There are a few key reasons why directed evolution has the potential to be an excellent complement to the *in vivo* incorporation of non-canonical amino acids. First, both techniques utilize recombinant DNA technology for the *in vivo* expression of target proteins. Therefore, the two techniques can readily be performed concurrently. Diversity is introduced at the DNA level through various genetic manipulations, which can then be translated into protein variants containing non-canonical amino acids using auxotrophic *E. coli* hosts and the medium shift procedure. Another attractive feature is that structural information is not necessary to yield successful outcomes. Unlike other "rational" design techniques that try to predict the amino acids responsible for key properties based on structural information, directed evolution assumes no prior knowledge. This is important since rarely is the structure known for proteins containing non-canonical amino acids. Finally, directed evolution is very effective when used to obtain proteins that perform "unnatural" tasks or function in "unnatural" environments. Clearly, nature did not evolve proteins to include amino acids with fluorines, azides, ketones, etc. Therefore, when we ask proteins to fold and function with "unnatural" amino acids, directed evolution has the potential to provide significant improvements.

A preliminary study in our laboratory has shown promise in this regard. When 5TFL was incorporated into 11 leucine positions in chloramphenicol acetyltransferase, the half life ($t_{1/2}$) of thermal inactivation of the enzyme at 60 °C was reduced by 20-fold.[56] Through two rounds of random mutagenesis and screening, a more thermostable, fluorinated CAT was found, exhibiting 27-fold improvement in $t_{1/2}$.

1.4 Proteins of Interest

In this thesis, two types of model proteins are explored for the incorporation of fluorinated amino acids. The first set of proteins belongs to the leucine zipper family while the other is based on the enzyme chloramphenicol acetyltransferase. The following sections briefly describe these two types of proteins.

1.4.1 Leucine Zipper Proteins

Leucine zippers constitute a subcategory of coiled-coils,[57] widespread structural motifs found both as the dominant structures in fibrous proteins and as oligomerization domains in a variety of proteins.[58] The existence of the leucine zipper motif was first hypothesized by Landschultz, McKnight, and coworkers based on the study of C/EBP, an enhancer binding protein found in rat liver.[59, 60] They noted that a segment within the DNA binding domain of C/EBP shared notable sequence similarities with regions of Myc, Fos, and Jun transforming proteins as well as that of the yeast transcription factor GCN4. When the sequences from these regions were displayed on an idealized α-helix, Landschultz *et al.* discovered a periodic repetition of leucine residues at every seventh position, forming an array of at least four leucine repeats in eight helical turns, hence the name leucine zipper.

Frequently, the seven residue module is characterized by positions **a** through **g**, where **a** and **d** positions are occupied by hydrophobic residues and **e** and **g** positions are occupied by charged residues (Figure 1.2). The hydrophobic interactions of **a** and **d** positions drive the intermolecular association of amphipathic α-helices.[61] Therefore, leucine zipper proteins are ideal models for the study of fluorinated amino acids and their

effect on protein structure and stability. The exaggerated hydrophobic and fluorophilic nature of fluorocarbons can be examined in a protein context through the incorporation of fluorinated amino acids into **a** and/or **d** positions. Since the roles of **a** and **d** positions have been extensively studied and shown to primarily participate in hydrophobic interactions, the analysis of the fluorination effects can be simplified. Furthermore, several studies have already demonstrated the utility of leucine zippers in elucidating the effects of incorporating fluorinated amino acids.[36, 43, 45, 62] Therefore, in this thesis, leucine zipper proteins again served as the context in which to further our understanding of fluorination of proteins.

1.4.2 Chloramphenicol Acetyltransferase

The effects of fluorination in the hydrophobic core of coiled-coils have been explored by several laboratories.[36, 43, 45, 62-64] However, the impact on global incorporation of fluorinated amino acids into larger, more complex proteins has yet to be studied in depth. Therefore, in order to extend the versatility of fluorinated amino acid incorporation, analysis of proteins with different sizes, structures, and functions is necessary. For this purpose, chloramphenicol acetyltransferase (CAT) was chosen.

CAT confers chloramphenicol resistance in bacteria by catalyzing acetyl-group transfer from acetyl coenzyme A to the hydroxyl groups of the antibiotic.[65] Unlike simple coiled-coil structures, CAT is comprised of a mixed six-stranded parallel and antiparallel β-pleated sheet supported by six α-helices.[66] As shown in Figure 1.3, CAT, with a monomeric molecular weight of 27 kDa, assembles into a trimer where each subunit extends its β-strand across the subunit interface to interact with its neighbor.

Chloramphenicol binds in a deep pocket that is located at the interface between the adjacent subunits of the trimer. CAT is readily expressed in *E. coli* and can be easily assayed for activity and stability.[65, 67, 68] Furthermore, CAT has been previously evolved successfully for various purposes.[56, 69, 70] Based on these factors, CAT is a suitable target protein to explore fluorinated amino acid incorporation as well as directed evolution.

1.5 Thesis Objectives

Although numerous studies have begun to explore the effects of fluorination on proteins, in order to fully capture the potential of fluorinated amino acids in protein engineering, a more thorough understanding in a wider context is necessary. Also, the ability to complement this technology with other protein engineering techniques would further enhance its versatility. In addition, the insights gained through the exploration of fluorinated amino acids could provide general understanding of non-canonical amino acids and their impact on protein behavior. The work described in this thesis makes important progress towards these goals.

Figure 1.1 Non-canonical amino acids incorporated *in vivo* by the Tirrell laboratory. Amino acids in black are accepted by the wild-type *E. coli* biosynthetic machinery. Those in blue require the overexpression of wild-type *E. coli* aminoacyl-tRNA synthetase. Amino acids in green require mutations in the editing site whereas those in red require mutations in the active site of the aminoacyl-tRNA synthetase.

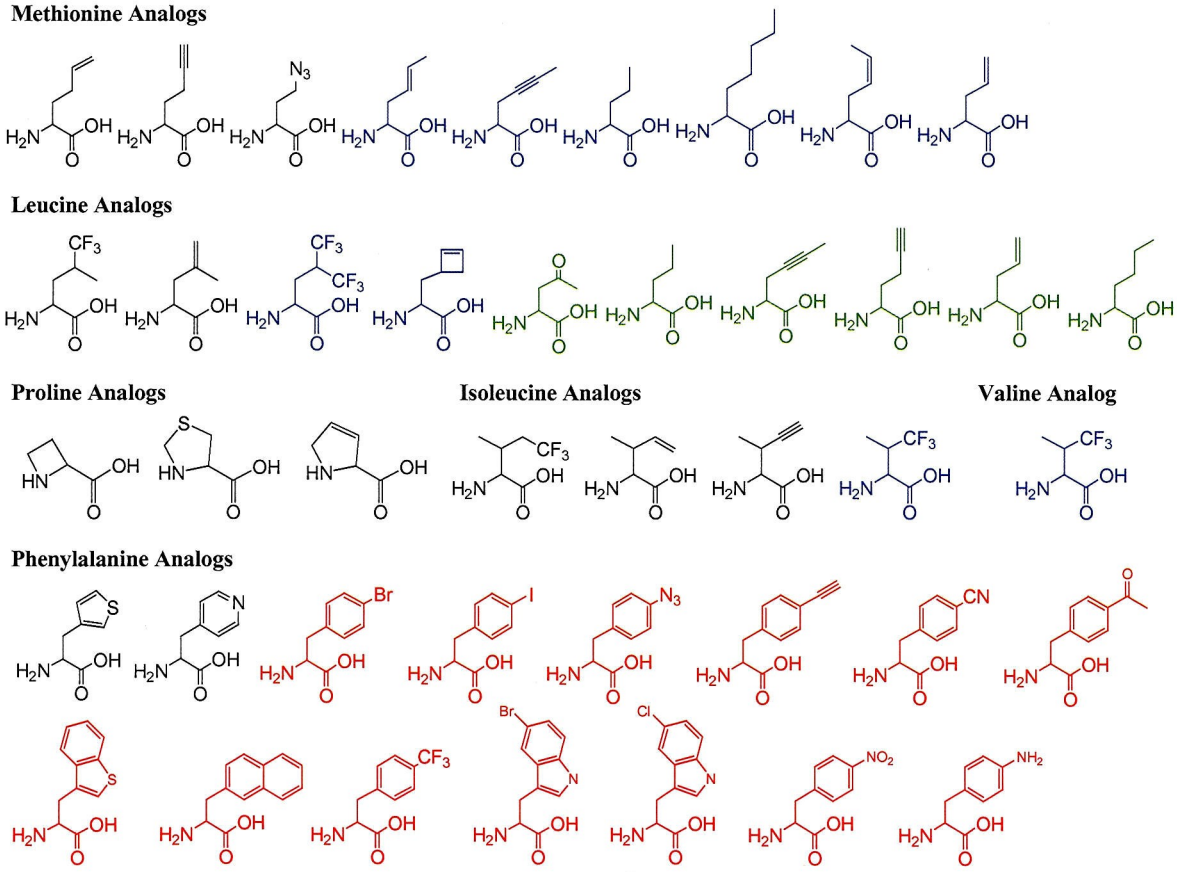
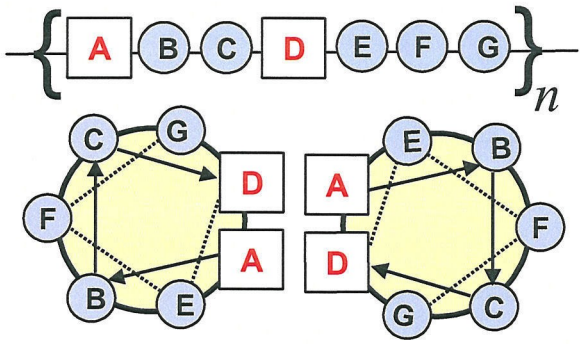


Figure 1.2 Description of the leucine zipper motif. **A**: The linear sequences of leucine zippers are characterized by repeats of heptads. The **a** and **d** positions are usually occupied by hydrophobic residues such as leucine, valine, or isoleucine. The view down the helix axis of a dimer leucine zipper indicates that the burial of **a** and **d** positions drives the coiled-coil interaction. **B**: The crystal structure of GCN4-bZIP protein bound to DNA containing the AFT/CREB recognition sequence.[71]

A



B

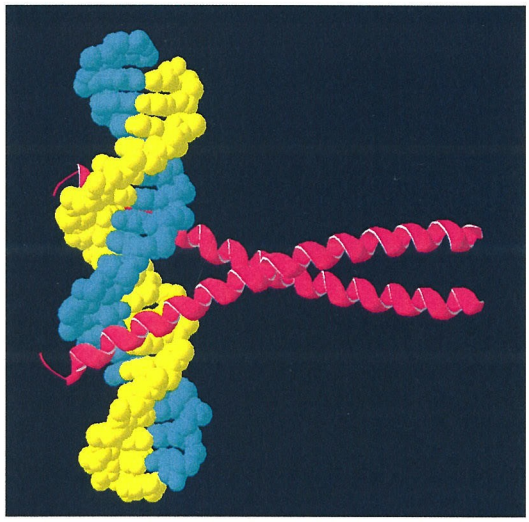
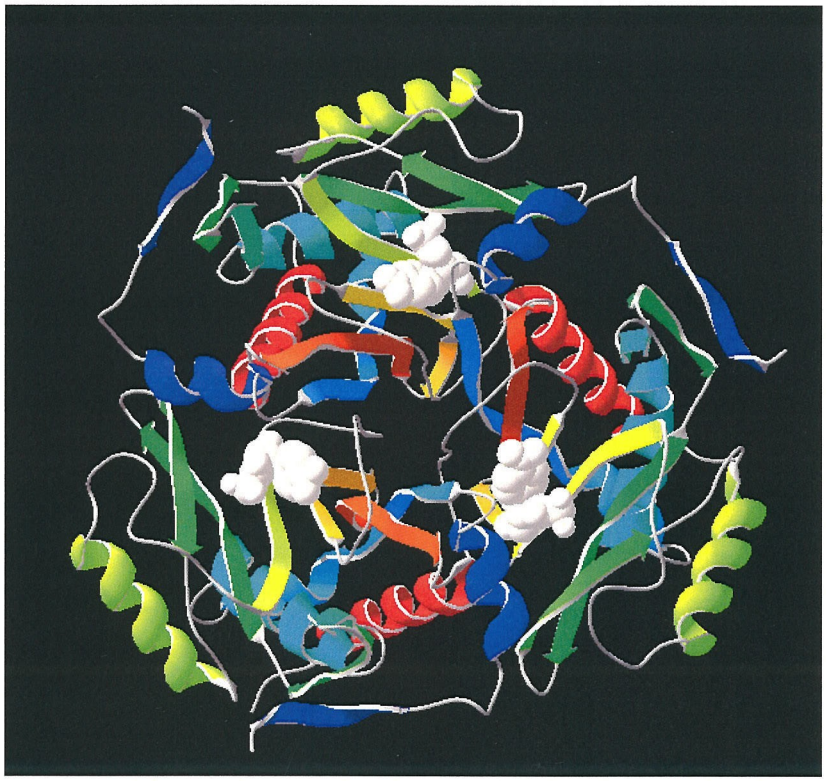


Figure 1.3 The crystal structure of type III chloramphenicol acetyltransferase from *E. coli*.[72] The ribbon representation is colored according to secondary structure succession. The white balls correspond to chloramphenicols, which are located at the boundary between adjacent subunits of the trimer.



1.6 References

1. Creighton, T. E., *Proteins*. W. H. Freeman and Company: New York, **1993**.

2. Yu, S. J. M.; Conticello, V. P.; Zhang, G. H.; Kayser, C.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., *Nature* **1997**, 389, (6647), 167-170.

3. Petka, W. A.; Harden, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A., *Science* **1998**, 281, (5375), 389-392.

4. Welsh, E. R.; Tirrell, D. A., *Biomacromolecules* **2000**, 1, (1), 23-30.

5. Panitch, A.; Yamaoka, T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., *Macromolecules* **1999**, 32, (5), 1701-1703.

6. Hecht, S. M.; Alford, B. L.; Kuroda, Y.; Kitano, S., *J. Biol. Chem.* **1978**, 253, (13), 4517-20.

7. Gallivan, J. P.; Lester, H. A.; Dougherty, D. A., *Chem. Biol.* **1997**, 4, (10), 739-749.

8. Hohsaka, T.; Sato, K.; Sisido, M.; Takai, K.; Yokoyama, S., *FEBS Lett.* **1994**, 344, (2-3), 171-174.

9. Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R.; Diala, E. S., *J. Am. Chem. Soc.* **1989**, 111, (20), 8013-8014.

10. Noren, C. J.; Anthonycahill, S. J.; Griffith, M. C.; Schultz, P. G., *Science* **1989**, 244, (4901), 182-188.

11. Sisido, M.; Hohsaka, T., *Appl. Microbiol. Biotechnol.* **2001**, 57, (3), 274-81.

12. Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M., *Biochemistry* **2001**, 40, (37), 11060-4.

13. Taki, M.; Hohsaka, T.; Murakami, H.; Taira, K.; Sisido, M., *FEBS Lett.* **2001**, 507, (1), 35-38.

14. Murakami, H.; Hohsaka, T.; Ashizuka, Y.; Hashimoto, K.; Sisido, M., *Biomacromolecules* **2000**, 1, (1), 118-25.

15. Budisa, N., *Angew. Chem.-Int. Edit.* **2004**, 43, (47), 6426-6463.

16. Wang, L.; Schultz, P. G., *Angew. Chem.-Int. Edit.* **2005**, 44, (1), 34-66.

17. Link, A. J.; Mock, M. L.; Tirrell, D. A., *Curr. Opin. Biotechnol.* **2003**, 14, (6), 603-609.

18. Tang, Y.; Tirrell, D. A., *J. Am. Chem. Soc.* **2001**, 123, (44), 11089-90.

19. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R., *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, (1), 19-24.

20. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A., *FEBS Lett.* **2001**, 502, (1-2), 25-30.

21. Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A., *FEBS Lett.* **2000**, 467, (1), 37-40.

22. Furter, R., *Protein Sci.* **1998**, 7, (2), 419-426.

23. van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A., *J. Am. Chem. Soc.* **2000**, 122, (7), 1282-1288.

24. Kirshenbaum, K., Carrico, I., Tirrell, D., *ChemBioChem* **2002**, 3, 235-237.

25. Wang, P.; Tang, Y.; Tirrell, D. A., *J. Am. Chem. Soc.* **2003**, 125, (23), 6900-6906.

26. O'Hagan, D.; Harper, D. B., *J. Fluor. Chem.* **1999**, 100, (1-2), 127-133.

27. Bohm, H. J.; Banner, D.; Bendels, S.; Kansy, M.; Kuhn, B.; Muller, K.; Obst-Sander, U.; Stahl, M., *ChemBioChem* **2004**, 5, (5), 637-643.

28. Desiraju, G. R., *Accounts Chem. Res.* **2002**, 35, (7), 565-573.

29. Resnati, G., *Tetrahedron* **1993**, 49, (42), 9385-9445.

30. Hudlicky, M., *Chemistry of Organic Fluorine Compounds*. Ellis-Horwood: New York, **1992**.

31. Gerig, J. T., *Prog. NMR Spectrosc.* **1994**, 26, 293-370.

32. Bann, J. G., Pinkner, J., Hultgren, S. J., Frieden, C., *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 709-714.

33. Seifert, M. H.; Kziazek, D.; Azim, M. K.; Smialowski, P.; Budisa, N.; Holak, T. A., *J. Am. Chem. Soc.* **2002**, 124, 7932-7942.

34. Sykes, B. D., Weiner, J. H., *Magnet. Reson. Biol.* **1980**, 1, 171-196.

35. Duewel, H.; Daub, E.; Robinson, V.; Honek, J. F., *Biochemistry* **1997**, 36, (11), 3404-16.

36. Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A., *Angew. Chem.-Int. Edit.* **2001**, 40, (8), 1494-1496.

37. Munier, R.; Cohen, G. N., *Biochim. Biophys. Acta* **1956**, 21, (3), 592-593.

38. Munier, R.; Cohen, G. N., *Biochim. Biophys. Acta* **1959**, 31, (2), 378-391.

39. Rennert, O. M.; Anker, H. S., *Biochemistry* **1963**, 2, (3), 471-476.

40. Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., *Macromolecules* **1994**, 27, (19), 5471-5475.

41. Wang, P.; Fichera, A.; Kumar, K.; Tirrell, D. A., *Angew. Chem.-Int. Edit.* **2004**, 43, (28), 3664-3666.

42. O'Shea, E. K.; Rutkowski, R.; Kim, P. S., *Science* **1989**, 243, (4890), 538-42.

43. Bilgicer, B.; Xing, X.; Kumar, K., *J. Am. Chem. Soc.* **2001**, 123, (47), 11815-11816.

44. Bilgicer, B.; Kumar, K., *Tetrahedron* **2002**, 58, (20), 4105-4112.

45. Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A., *Biochemistry* **2001**, 40, (9), 2790-2796.

46. Pal, P. P.; Bae, J. H.; Azim, M. K.; Hess, P.; Friedrich, R.; Huber, R.; Moroder, L.; Budisa, N., *Biochemistry* **2005**, 44, (10), 3663-3672.

47. Arnold, F. H., *Accounts Chem. Res.* **1998**, 31, (3), 125-131.

48. Tao, H. Y.; Cornish, V. W., *Curr. Opin. Chem. Biol.* **2002**, 6, (6), 858-864.

49. Stemmer, W. P. C., *Nature* **1994**, 370, (6488), 389-391.

50. Moore, J. C.; Arnold, F. H., *Nat. Biotechnol.* **1996**, 14, (4), 458-467.

51. Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. C., *Nat. Biotechnol.* **1996**, 14, (3), 315-319.

52. Miyazaki, K.; Arnold, F. H., *J. Mol. Evol.* **1999**, 49, (6), 716-720.

53. Arnold, F. H.; Volkov, A. A., *Curr. Opin. Chem. Biol.* **1999**, 3, (1), 54-59.

54. Joo, H.; Lin, Z. L.; Arnold, F. H., *Nature* **1999**, 399, (6737), 670-673.

55. Merz, A.; Yee, M. C.; Szadkowski, H.; Pappenberger, G.; Crameri, A.; Stemmer, W. P. C.; Yanofsky, C.; Kirschner, K., *Biochemistry* **2000**, 39, (5), 880-889.

56. Montclare, J. K., Tirrell, D. A., *Angew. Chem.-Int. Edit.* **2006**, 45, (27), 4518-4521.

57. Crick, F. H. C., *Acta Crystallogr.* **1953**, 6, 689-697.

58. Lupas, A., *Trends Biochem. Sci.* **1996**, 21, 375-382.

59. Landschulz, W. H.; Johnson, P. F.; McKnight, S. L., *Science* **1988**, 240, (4860), 1759-64.

60. McKnight, S. L.; Landschulz, W. H.; Johnson, P. F., *Curr. Top. Microbiol. Immunol.* **1988**, 141, 186-8.

61. O'Shea, E. K., Klemm, J. D., Kim, P. S., Alber T., *Science* **1991**, 254, 539-544.

62. Bilgicer, B.; Fichera, A.; Kumar, K., *J. Am. Chem. Soc.* **2001**, 123, (19), 4393-4399.

63. Lee, K. H.; Lee, H. Y.; Slutsky, M. M.; Anderson, J. T.; Marsh, E. N. G., *Biochemistry* **2004**, 43, (51), 16277-16284.

64. Lee, H. Y.; Lee, K. H.; Al-Hashimi, H. M.; Marsh, E. N. G., *J. Am. Chem. Soc.* **2006**, 128, (1), 337-343.

65. Shaw, W. V.; Leslie, A. G. W., *Annu. Rev. Biophys. Biophys. Chem.* **1991**, 20, 363-386.

66. Leslie, A. G. W., Liddell, J. M., Shaw, W. V., *J. Mol. Biol.* **1986**, 188, (283-285).

67. Shaw, W. V., *Methods Enzymol.* **1975**, 43, 737-755.

68. Turner, S. L., Ford, G. C., Mountain, A., Moir, A., *Protein Eng.* **1992**, 5, (6), 535-541.

69. Robben, J.; Vanderschueren, J.; Verhasselt, P.; Aert, R.; Volckaert, G., *Protein Eng.* **1995**, 8, (2), 159-165.

70. Van der Schueren, J.; Robben, J.; Volckaert, G., *Protein Eng.* **1998**, 11, (12), 1211-1217.

71. Keller, W.; Konig, P.; Richmond, T. J., *J. Mol. Biol.* **1995**, 254, (4), 657-667.

72. Leslie, A. G. W., *J. Mol. Biol.* **1990**, 213, (1), 167-186.

Chapter 2

Stabilization of *bzip* Peptides through Incorporation of Fluorinated Aliphatic Residues

This chapter appeared as a full paper in Son, S., Tanrikulu, I. C., and Tirrell, D. A. *ChemBioChem* **2006**, 7, (8), 1251-1257.

2.0 Abstract

Two fluorinated amino acids, 5,5,5-trifluoroisoleucine (5TFI) and 2S,3R-4,4,4-trifluorovaline (4TFV), which have been shown to serve as isoleucine surrogates in protein synthesis in *Escherichia coli*, have been incorporated *in vivo* into basic leucine zipper (*bzip*) peptides derived from GCN4. The extents of residue-specific incorporation of 5TFI and 4TFV were 90% and 88%, respectively, of the encoded isoleucine residues as evidenced by MALDI mass spectrometry and amino acid analysis. Both circular dichroism and equilibrium sedimentation studies of the fluorinated *bzip* peptides indicated preservation of secondary and higher-order protein structure. Thermal denaturation experiments showed an increase of 27 °C in melting temperature when isoleucine was replaced by 5TFI. However, as compared to the peptide containing valine, that containing 4TFV exhibited an increase in T_m of only 4 °C. Similar trends were observed in chemical denaturation studies where $\Delta\Delta G_{unfold}$ in water was determined to be 2.1 and 0.3 kcal/mol upon incorporation of 5TFI and 4TFV, respectively. When the fluorinated peptides were tested for DNA binding, both affinity and specificity were similar to those of the respective hydrogenated peptides. These results suggest that fluorinated amino acids, even when introduced into the same positions, can have markedly different effects on the physical properties of proteins, while having little impact on secondary and higher-order structure.

2.1 Introduction and Background

The incorporation of non-natural amino acids into recombinant proteins has enhanced the power and versatility of protein engineering.[1-4] In particular, introduction of non-natural amino acids in multi-site fashion allows systematic engineering of the overall physical behavior of recombinant proteins. Multi-site incorporation is easy to implement and yields protein products in amounts sufficient for detailed physical characterization.[5-12] Fluorinated amino acids have been incorporated successfully at high levels into several different proteins *in vivo*.[13-15] Fluorine atoms have slightly larger van der Waals radii than hydrogen atoms (r(F) = 1.35 Å, r(H) = 1.20 Å), and the C-F bond is 0.4 Å longer than the C-H bond.[16, 17] Despite these differences, fluorinated amino acids serve as excellent substrates for the natural biosynthetic machinery and generally cause minimal structural perturbation.[13, 18, 19] Fluorine atoms exhibit low polarizability, and fluorocarbons yield higher contact angles with water than their hydrocarbon counterparts, suggesting enhanced hydrophobic character.[20] In light of the fact that hydrophobic forces play key roles in protein folding and protein-protein interactions, incorporation of fluorinated amino acids might be expected to alter protein stability and binding behavior. Indeed, Marsh and coworkers found that the $\Delta\Delta G_{unfold}$ increase resulting from the incorporation of L-5,5,5,5',5',5'-hexafluoroleucine (hFLeu) into a four-α-helix bundle, 0.3 kcal/mol per hFLeu, could be attributed to the increase in hydrophobicity of the side chain, determined experimentally to be 0.4 kcal/mol per hFLeu from partitioning measurements of hFLeu and leucine.[21] Similar trends were reported in recent work by Kumar and coworkers and by this laboratory, wherein fluorinated aliphatic residues were

inserted into the interhelical interfaces of coiled-coil peptides, resulting in enhanced stability with respect to thermal and chemical denaturation.[14, 18, 19, 22] In addition, we have found that protein function can be retained following introduction of fluorinated amino acids, as shown by the DNA binding affinity and specificity of fluorinated peptides derived from GCN4, as well as by preservation of the activity of murine interleukin-2 activity following replacement of isoleucine by 5,5,5-trifluoroisoleucine.[13, 18]

Fluorination has also been shown to serve as a tool for engineering of protein-protein interactions.[23, 24] Kumar and coworkers found that mixtures of hydrogenated and fluorinated coiled-coil peptides exhibited striking self-sorting behavior with a strong preference for homodimeric assembly.

In order to explore more fully the prospects for using fluorinated amino acids to control protein folding and protein-protein interaction, we have examined the consequences of introducing 2S,3R-4,4,4-trifluorovaline (4TFV) and 5,5,5-trifluoroisoleucine (5TFI) into coiled-coil peptides derived from GCN4. We find striking differences in the effects of these two amino acids on the stability of the coiled-coil assembly.

2.2 Materials and Methods

2.2.1 Fluorinated Amino Acids

4,4,4-DL-trifluorovaline was purchased from Matrix Scientific (Columbia, SC) and used without further purification. We have shown previously that only the 2S,3R-

isomer of 4,4,4-trifluorovaline is incorporated into recombinant proteins expressed in *E. coli*.[25] 5,5,5-DL-trifluoroisoleucine was synthesized according to published procedures with slight modifications.[13, 26-28]

2.2.2 Cloning of the Mutant bzip DNA Construct

The DNA sequence used to encode the mutant *bzip* peptides was derived from the yeast transcription factor, GCN4. The peptides include the DNA binding domain as well as the dimerization (zipper) domain, which together account for 56 residues in the C-terminal region of GCN4.[29] Apart from the valine/isoleucine substitutions in the **a**-positions of the heptad repeats, the DNA sequences closely resemble the *S. cerevisiae* sequence encoding the C-terminal region of GCN4 with a few silent mutations to reflect preferred codon usage of *E. coli*. To facilitate purification, an N-terminal His-tag was fused to the *bzip* peptides, resulting in sequence **1a**.

Oligonucleotides (prepared in the Caltech DNA Synthesis Facility) were annealed to yield DNA duplexes that encode both the basic and zipper regions. Following phosphorylation and purification, the duplex was ligated into *Sac*I/*Hind*III digested pQE30 (Qiagen) and used to transform *E. coli* cloning strain XL1-Blue. The expression plasmid pQE30 contains a T5 promoter and two *lac* operator sequences that allow tight regulation of expression. The plasmids isolated from the transformants, designated pQE-INL, were verified by sequencing.

Cloning of the DNA construct VNL was facilitated by the original design of the pQE-INL, which allowed simple replacement of the zipper domain through a restriction site. Oligonucleotides encoding the new zipper domain were synthesized and annealed.

Following purification and phosphorylation, the duplex was ligated into *Pst*I/*Hind*III digested pQE-INL. This digestion allowed release of only the zipper domain of INL, and insertion of a new zipper with Val replacing Ile at each of the **a**-positions of the heptad repeats. The resulting plasmid, designated pQE-VNL, was transformed into XL1-Blue and verified by sequencing.

To enhance incorporation of 4TFV into the mutant peptide tfv-VNL, an isoleucyl-tRNA synthetase (IleRS) gene, cloned directly from *E. coli* genomic DNA, was inserted into *Nhe*I digested pQE-INL.[25] The resulting plasmid was designated pQE-INL-IRS.

2.2.3 Construction of the Protein Expression System

An isoleucine auxotrophic derivative of *E. coli* strain BL21(DE3), designated AI (*E. coli* B F$^-$ *ompT hsdS*(r_B^- m_B^-) *gal dcm* λ(DE3) *ilv*D691), constructed in our laboratory, was used as the host strain for protein expression.[13] The repressor plasmid pLysS-IQ was constructed by Sharma in our laboratory, and carried the *lacI*^{q} gene for *laq* repressor.[7] The AI strain with the repressor plasmid pLysS-IQ was designated AI-IQ. The final expression system AI-IQ[pQE-INL] for the expression of INL and tfi-INL was obtained by transformation of pQE-INL into AI-IQ. The expression strains for VNL and tfv-VNL were constructed similarly and designated AI-IQ[pQE-VNL] and AI-IQ[pQE-INL-IRS], respectively.

2.2.4 Protein Expression and Incorporation of 5TFI and 4TFV into Mutant bzip Peptides

M9 minimal medium (1 L) supplemented with 1 mM $MgSO_4$, 1 mM $CaCl_2$, 0.4 wt% glucose, 1 mg/L thiamine, and the antibiotics ampicillin (100 mg/L) and

chloramphenicol (35 mg/L) was inoculated with 10 mL of fresh overnight culture (M9) of the expression strain. After the culture had grown to OD_{600} of 1.0, a medium shift procedure was performed. Cells were sedimented by centrifugation for 8 minutes at 6000 g at 4 °C. The supernatant was removed and the cell pellets were washed twice with 0.9% NaCl. The cell pellet was then resuspended in 1 L M9 medium supplemented with 5TFI or 4TFV (200 mg/L or 300 mg/L, respectively). IPTG (1 mM) was added after 10 minutes to induce protein expression. Cells were collected after 4 hours by centrifugation (6000 g, 10 min, 4 °C).

2.2.5 Protein Purification

Harvested cells were resuspended in Buffer A, pH 8.0 (30 mL; 6 M GuHCl, 0.1 M NaH_2PO_4, 0.01 M TrisCl) and stored at -80 °C overnight. The cells were thawed on ice and shaken slowly at room temperature for 30 minutes to effect complete lysis. Cell debris was sedimented (10,000 g, 30 min, r.t.), and the supernatant was applied to a Ni-NTA column. Following a series of washes with Buffer A ranging from pH 7.0 to 5.9, target protein was eluted at pH 4.5. The purified protein was subjected to matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) and amino acid analysis. Amino acid analysis was performed at W. M. Keck Facility at Yale University using a Beckman Model 7300 ion-exchange instrument following a 16 hour hydrolysis at 115 °C in 100 µl of 6 N HCl, 0.2% phenol that also contains 2 nmol norleucine. Because both 5TFI and 4TFV coeluted with either a natural amino acid or standards used in the analysis, the percent incorporation was calculated by the absence of natural Ile.

2.2.6 Analytical Ultracentrifugation

Sedimentation equilibrium analysis was performed on a Beckman XLI Analytical Ultracentrifuge, recording radial absorbance at 236 nm. Initial peptide concentrations ranged from 10 to 50 μM and data were collected at 20 °C. The peptides were prepared in and dialyzed against PBS buffer (137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4 \cdot 7H_2O$, pH 7.4). Data were collected at three speeds (32,000, 38,000, and 45,000 rpm for VNL and tfv-VNL; 34,000, 40,000, and 47,000 rpm for INL and tfi-INL) until equilibrium was reached as verified by the program WinMatch.[30] Equilibrium was reached within 24 hours in all experiments. The program WinReedit was used to select the ranges of data for further analysis and the equilibrium distributions were then fit globally using the software Winnl106.[31] The program Sednterp was used to calculate the solvent density and the protein partial specific volume from the amino acid composition of each protein, substituting leucine and valine for 5TFI and 4TFV, respectively, in the fluorinated peptides for lack of available data.[32]

2.2.7 Circular Dichroism Spectroscopy

Circular dichroism spectra were recorded on an Aviv 62DS spectropolarimeter. All experiments were performed in a rectangular cell with path length of 1 mm with a thermostatically controlled cuvette. Protein concentrations were determined using the Bradford assay (BioRad, Hercules, CA). Guanidine hydrochloride (GuHCl) concentrations were based on dry weight of 99.8% UltraPure guanidine hydrochloride (Sigma, St. Louis, MO). In preparing GuHCl solutions of various concentrations, an 8 M

GuHCl solution was prepared and diluted with appropriate amounts of PBS buffer, pH 7.0.

Wavelength scans were performed at 4 °C from 255 to 195 nm. Data points were collected every 1 nm with an averaging time of 5 seconds at each wavelength. For each sample, the scan was performed three times and averaged. The CD deconvolution software, CDNN, was used to calculate the helical content from the mean residue ellipticity vs. wavelength data.[33]

Temperature scans were performed at 222 nm from 0 °C to 100 °C with points taken every 1 °C. Two scans were performed on a single sample and averaged. Each data point was collected after 1 minute of thermal equilibration at the desired temperature with an averaging time of 20 seconds at each temperature.

Thermodynamic quantities were estimated from the thermal melts by fitting the denaturation curves to a monomer-dimer equilibrium according to procedures published by Schneider and coworkers.[34] Mean residue ellipticity at 222 nm was fit as a function of temperature and total peptide concentrations using Origin7 software (Northampton, MA) using a 1 M peptide standard state. The data were obtained from global fits of thermal denaturation data at peptide concentrations of 20 and 50 μM. All thermodynamic quantities were calculated on a per mole of monomer basis.

Chemical denaturation curves were obtained from guanidine hydrochloride titration experiments. Two stock solutions of 20 μM peptide concentration were prepared for each peptide, one in 0 M GuHCl (PBS, pH 7) and the other at 8 M GuHCl (pH 7). For each desired GuHCl concentration, appropriate amounts of each stock solution were mixed and allowed to reach equilibrium at the set temperature of 25 °C.

Thermodynamic quantities were estimated from chemical denaturation experiments as described above, except the fitting was done at a particular concentration (rather than as a global fit). The apparent free energy difference between folded dimer and unfolded monomer states is assumed to be linearly dependent on the GuHCl concentration.[19] Therefore, ΔG^o can be written as

$$\Delta G^o = \Delta G^o_{H2O} - m[GuHCl]$$

All thermodynamic quantities reported are per mole of monomer.

2.2.8 Gel Retardation Assay

Oligonucleotides containing the CREB binding site (CREB, 5'-TGGAGATGACGTCATCTCCT-3') and the non-specific sequence (NON, 5'-GATCCCAACACGTGTTGGGATC-3') were synthesized at the Caltech DNA Synthesis Facility. The oligonucleotides were labeled with γ-[^{32}P]ATP using polynucleotide kinase, annealed, and purified. Serially diluted solutions of peptides (0 to 100 nM, in PBS) were made from a single stock and a constant concentration of labeled DNA ($\leq$ 480 pM) was added. The reactions were incubated at 4 °C for 30 minutes, and 8 μL of each reaction was loaded onto a running 5% polyacrylamide gel and run at 2 watts, 5 °C for 1.5 hours. Free and protein-bound DNA was visualized by autoradiography. The relative intensities of the bands were measured by densitometry.

2.3 Results and Discussion

2.3.1 Design Principles

The current study explores the effects of fluorination on two mutant *bzip* peptides derived from the yeast transcription factor GCN4. These two peptides are identical to the 56 residues in the C-terminal region of GCN4, except for the amino acids in the **a**-positions of the heptad repeats. As shown in Figure 2.1, four of the five amino acids in the **a**-positions have been mutated to either isoleucine or valine, to yield peptides designated INL and VNL, respectively. Three of the five **a**-position amino acids are valines in wildtype GCN4. The conserved asparagine in the **a**-position of the third heptad repeat was retained in the design of all target proteins because of its important role in determining oligomerization state, orientation specificity, and the overall geometry of the coiled-coil.[35-38]

2.3.2 Expression of Mutant bzip Peptides and Their Fluorinated Counterparts

The four peptides of interest, INL, VNL, tfi-INL, and tfv-VNL, were expressed in *E. coli* according to established procedures.[13, 25] All four proteins expressed in good yield and were purified cleanly by nickel-affinity chromatography (Table 2.1). The incorporation of 5TFI and 4TFV into tfi-INL and tfv-VNL, respectively, was confirmed by MALDI-TOF mass spectrometry and quantitative amino acid analysis. The results indicate that the incorporation level was 88% or higher in all samples (Table 2.1, Figure 2.2), with 5TFI being incorporated at slightly higher levels than 4TFV. These results are

consistent with previous reports of incorporation of 5TFI and 4TFV into recombinant proteins.[13, 25]

2.3.3 Equilibrium Sedimentation of Mutant bzip Peptides

Equilibrium sedimentation experiments showed all four mutant *bzip* peptides to be predominantly dimeric in the concentration range of interest in this study. The equilibrium profiles were fit to a model assuming a single species to yield an average molecular weight that corresponds to the dimeric form of each peptide (VNL and tfv-VNL: 17,400 $\pm$ 300, INL and tfi-INL: 17,300 $\pm$ 300). In all fits, the square root of the variance ranged between $5x10^{-3}$ and $8x10^{-3}$. These results indicate that the incorporation of 5TFI and 4TFV into mutant *bzip* peptides preserves the oligomerization state and dimerization affinity characteristic of the coiled-coil region of GCN4.

2.3.4 Spectroscopic Characterization

All four proteins are highly helical at 4 °C as evidenced by strong minima at 222 and 208 nm in the CD spectrum (Figure 2.3). At 20 μM peptide concentration, the molar ellipticity at 222 nm indicates 45 to 48% helical content.[33] The spectra of the hydrogenated and fluorinated peptides are essentially coincident, suggesting nearly identical secondary structures. These results are consistent with our expectation that incorporation of 5TFI and 4TFV would not disrupt the folding of the coiled-coil structure.[18, 19]

Thermal stability was determined by monitoring the molar ellipticity at 222 nm as a function of temperature. The melting experiments were performed in PBS for VNL and

tfv-VNL, while INL and tfi-INL experiments required 3 M guanidine hydrochloride (GuHCl) in order to observe complete melting of the peptides. Incorporation of either 4TFV or 5TFI into mutant *bzip* peptides gave rise to increased dimer melting temperatures (Figure 2.4); however, the difference in thermal denaturation temperature between INL and tfi-INL (27 °C) is much greater than that between VNL and tfv-VNL (4 °C) at 20 μM peptide concentration. Based on global analysis of the thermal unfolding curves at two different concentrations (20 μM and 50 μM), the free energy of unfolding at 25° C of tfi-INL is estimated to be 2.3 kcal/mol greater than that of INL. The corresponding change in unfolding free energy is only 0.4 kcal/mol when valine is replaced by 4TFV in tfv-VNL.

The stability of each peptide toward chaotropic reagents was determined through GuHCl titration experiments. The results of the chemical denaturation study were consistent with those of the thermal unfolding experiments. At 25 °C, the concentration of GuHCl needed to unfold 50% of the sample was higher for both of the fluorinated peptides as compared to their hydrogenated counterparts (Figure 2.5). The apparent free energy of folding in the absence of GuHCl (ΔG^{o}) calculated at 25 °C was 16.6 $\pm$ 0.5 kcal/mol for tfi-INL vs. 14.5 $\pm$ 0.5 kcal/mol for INL. The corresponding values for tfv-VNL and VNL were determined to be 8.7 $\pm$ 0.3 kcal/mol and 8.4 $\pm$ 0.4 kcal/mol, respectively. Literature values for ΔG^{o} of GCN4-p1, the 33-residue leucine zipper region of GCN4, range from 8.0 to 10.5 kcal/mol at temperatures of 4 to 25 °C.[18, 19, 39, 40]

The enhancement in stability caused by replacement of valine by 4TFV was much smaller than that caused by replacement of isoleucine by 5TFI. The smaller stabilizing effect of 4TFV can be inferred from previous results as well. Tang and coworkers found

an increase of 13 °C in the melting temperature of GCN4-p1 when the four **d**-position leucine residues were replaced by trifluoroleucine.[18] When Kumar and coworkers made the further substitution of three valine residues by trifluorovaline, a nearly identical (15 °C) increase in melting temperature was observed.[19] Although quantitative comparison may be frustrated by differences in experimental conditions, these results suggest that most of the stability enhancement observed by Kumar and coworkers arose from fluorination of the **d**-position leucine residues.

The 27-28 °C increase in thermal melting temperature associated with fluorination of the four isoleucine residues in the **a**-positions of INL is remarkable. Although we do not know why 5TFI yields a larger increase in T_m than 4TFV, we suggest that the difference may arise from both entropic and packing effects. Upon helix formation, valine loses substantial side chain entropy through steric clashes between its γ-methyl groups and the helix backbone.[41] Given the increased steric bulk of the trifluoromethyl group, it is reasonable to propose that 4TFV suffers further loss of conformational entropy, and that the entropy loss largely negates the stabilizing effect of the increased hydrophobicity of 4TFV. An alternative, and perhaps complementary, explanation may be found in the unique stabilizing effect of the homotypic interaction between isoleucine residues in the **a**- and **a'**-positions of the dimeric coiled coil.[42] Vinson and coworkers attribute this effect to special packing properties of the isoleucine side chain – packing properties that may allow more complete burial of the trifluoromethyl group in 5TFI as compared to that in 4TFV.

2.3.5 DNA Binding Studies

One of the goals of protein engineering is create proteins that resist denaturation but exhibit normal or enhanced activity.[43, 44] In that context, it was of interest to us to determine whether tfi-INL and tfv-VNL can bind DNA with affinities and specificities similar to those of their hydrogenated counterparts. Qualitative determination of DNA binding activity was performed by CD analysis. Because the DNA-binding domain of GCN4 is conformationally disordered in the absence of DNA and becomes helical only when bound to target DNA, monitoring of the helical content of the peptide serves as an indicator of DNA binding.[45-49] As mentioned earlier, in the absence of DNA-binding partners, the peptides of interest were 45 to 48% helical. When oligonucleotides containing the CREB target binding sequence were added, all four peptides showed more intense CD minima at 208 and 222 nm, corresponding to increases in helical content to 56 to 58% (Figure 2.6). The similarity of the spectroscopic changes observed for the fluorinated peptides and their natural counterparts upon addition of target DNA implies similar DNA binding behavior.

Gel-retardation assays were used to explore the affinity and specificity of DNA binding. The mobility shift assay indicated that the binding affinity and specificity do not change upon incorporation of fluorinated amino acids (Figure 2.7). Based on densitometric analysis, the values of K_d were found to be indistinguishable for tfi-INL and INL (12.0 $\pm$ 0.6 nM and 12.8 $\pm$ 0.6 nM, respectively) and for tfv-VNL and VNL (12.1 $\pm$ 0.5 nM and 12.3 $\pm$ 0.6 nM, respectively). These values are similar to those of a control peptide with a wild-type GCN4 sequence (11.3 $\pm$ 0.9 nM), and correlate well with previous measurements on GCN4.[45, 47, 50-53] At concentrations up to 100 nM, none of the

peptides showed detectable binding to a non-specific oligonucleotide sequence. These results indicate that the DNA binding affinity of *bzip* peptides can be maintained upon introduction of fluorinated amino acids, consistent with previous studies of fluorinated variants of GCN4-p1.[18]

2.4 Conclusions

The fluorinated aliphatic amino acids 5TFI and 4TFV are both incorporated efficiently into *bzip* peptides *in vivo* and both yield coiled-coil dimers of enhanced stability. Surprisingly, when introduced into the same positions within the protein, 5TFI gives rise to an increment in the free energy of unfolding of ca. 2.3 kcal/mol, while 4TFV incorporation results in a much smaller increase (ca. 0.3 kcal/mol) in stability. Incorporation of 5TFI or 4TFV into mutant *bzip* peptides leaves DNA binding affinity and specificity unchanged. We are continuing to explore the use of fluorinated amino acids (and other non-canonical amino acids) to control protein structure, function, and stability.

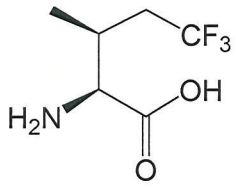
Table 2.1 Protein yield and extent of amino acid replacement in mutant *bzip* peptides. Molecular weight was measured by MALDI-TOF analysis. The theoretical molecular weights of tfi-INL and tfv-INL were calculated by assuming 100% replacement of all four isoleucines by fluorinated amino acids. The extent of replacement was determined both by amino acid analysis and by MALDI-TOF analysis of tryptic fragments.

Protein	**Yield (mg/L)**	**Molecular Weight (Measured, Theoretical)**	**% Replacement (AAA, MALDI)**
INL	41	8586.3, 8584.9	N.A.
tfi-INL	30	8803.7, 8800.9	90.0, 90.1
VNL	38	8530.5, 8528.8	N.A.
tfv-VNL	27	8744.0, 8744.8	87.8, 88.0

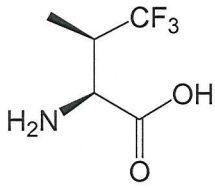
Figure 2.1 Description of the model system: **1a** amino acid sequences of *bzip* peptides INL and VNL, **1b** chemical structure of (2*S*, 3*R*)-5,5,5-trifluoroisoleucine (5TFI), **1c** chemical structure of (2*S*, 3*R*)-4,4,4-trifluorovaline (4TFV), **1d** schematic of the coiled-coil region of the zipper peptides with four **a**-positions highlighted based on the x-ray crystal structure of GCN4.[54]

1a MRGSHHHHHHGSACELMDPAALKRARNTEAARRSRARKLQR *I/V* KQLEDK *I/V* EELLSKNYHLENE *I/V* ARLKKL *I/V* GER

1b



1c



1d

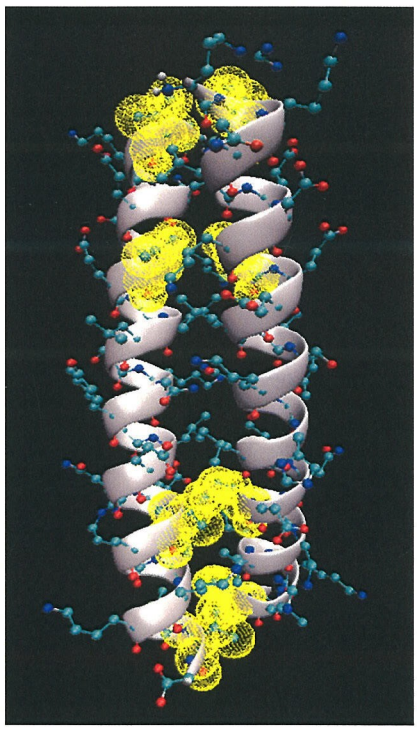


Figure 2.2 MALDI analysis of mutant *bzip* peptides following trypsin digestion. A fragment of sequence NYHLENEIAR (residues 56-65), yields the spectra shown. (A) fragments digested from tfv-VNL, (B) fragments digested from tfi-INL.

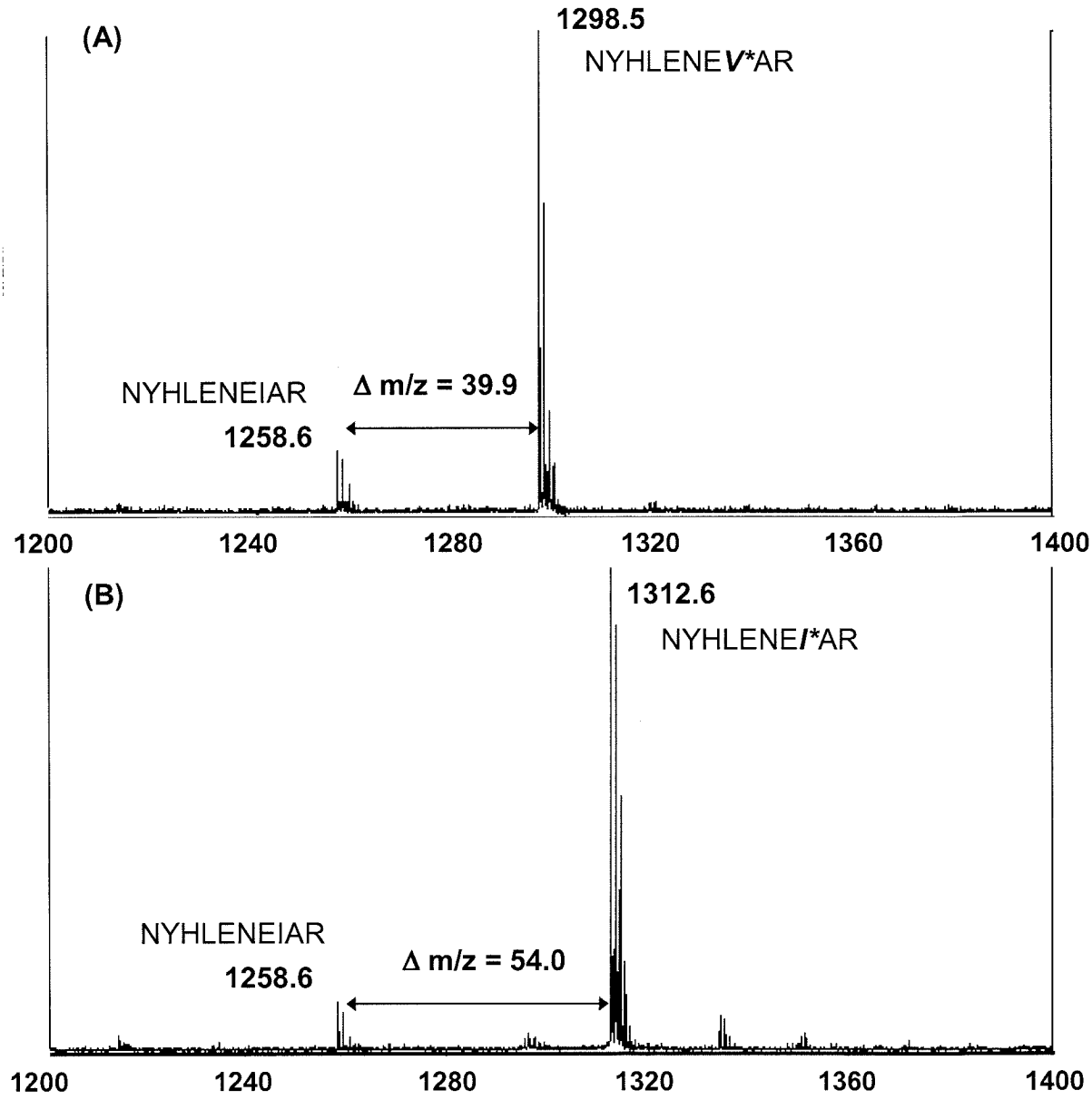


Figure 2.3 Circular dichroism spectra of 20 μM INL (closed circles), tfi-INL (open circles), VNL (closed triangles), and tfv-VNL (open triangles) in PBS (10 mM phosphate, 137 mM NaCl, 2.7 mM KCl, pH 7.4) at 4 °C. The minima at 222 and 208 nm indicate that all peptides are α-helical.

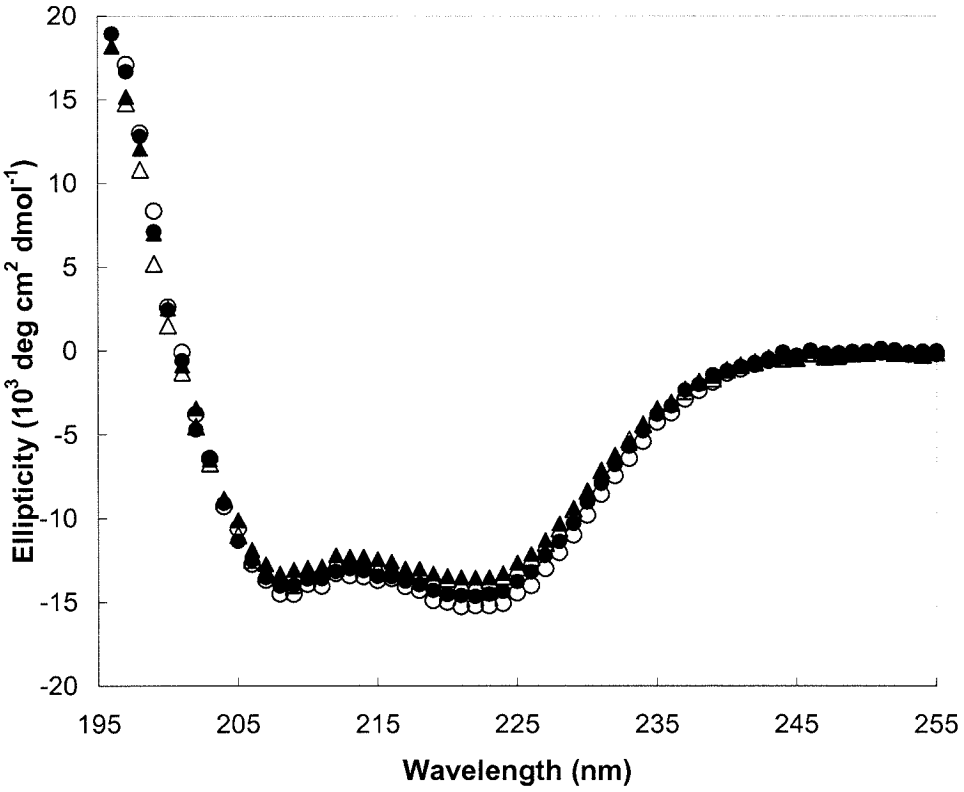


Figure 2.4 Thermal denaturation of mutant *bzip* peptides monitored by CD ([θ_{222}]). (A) Thermal unfolding profiles for VNL and tfv-VNL at 20 μM (open symbols) and 50 μM (closed symbols) in PBS (10 mM phosphate, 137 mM NaCl, 2.7 mM KCl, pH 7.4). At 20 μM, the values of T_m for VNL and tfv-VNL are 57 and 61 °C, respectively. (B) Thermal unfolding profiles for INL and tfi-INL at 20 μM (open symbols) and 50 μM (closed symbols) in 4 M GuHCl, 10 mM phosphate, pH 7.0 solution. At 20 μM, the values of T_m for INL and tfi-INL are 33.5 and 60 °C, respectively.

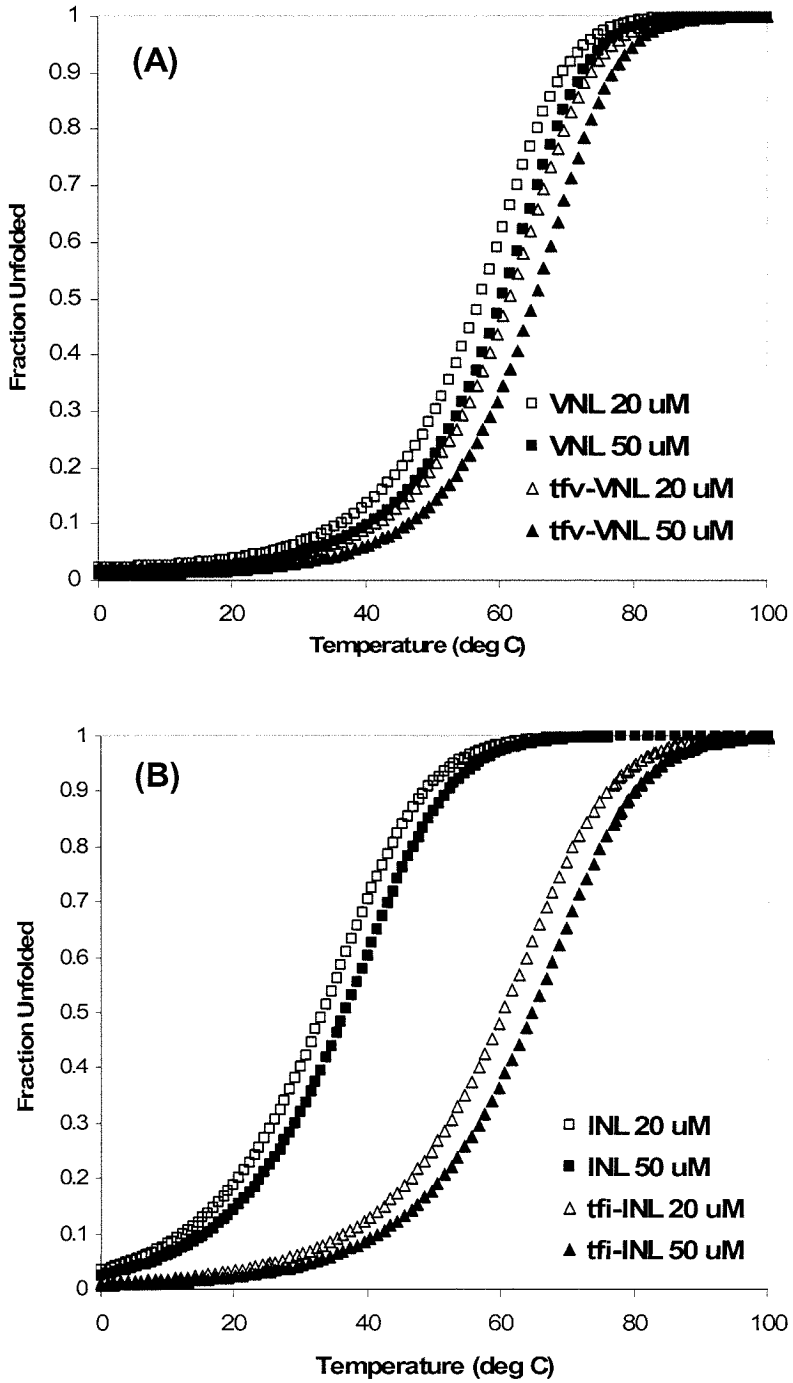


Figure 2.5 Guanidinium hydrochloride (GuHCl) titration of (A) VNL (closed circles) and tfv-VNL (open circles), and (B) INL (closed squares) and tfi-INL (open squares). Data were recorded at 25 °C. The ellipticity was monitored at 222 nm at a peptide concentration of 20 μM.

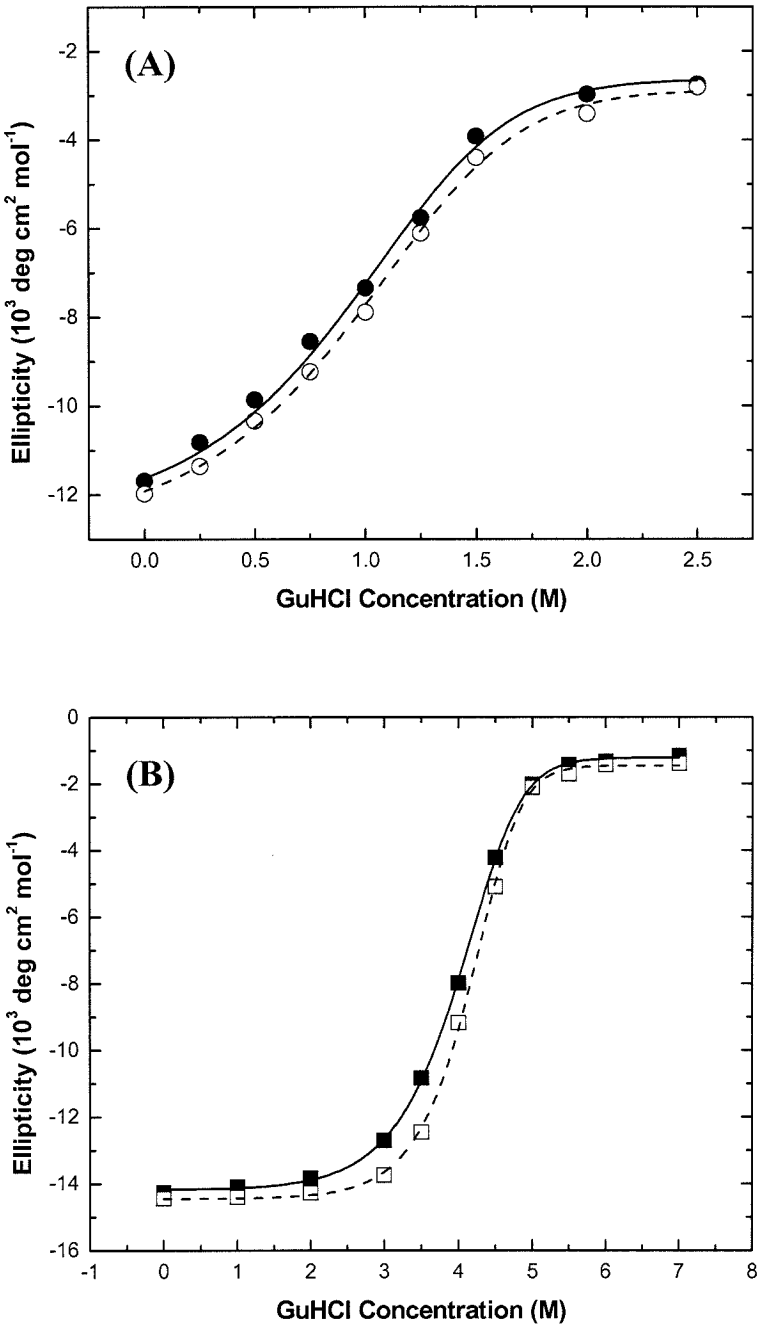


Figure 2.6 CD spectra for INL (squares), tfi-INL (triangles), VNL (circles), tfv-VNL (diamonds) with (open symbols) and without (closed symbols) CREB at 4 °C. Based on the molar ellipticity at 222 nm, the peptide helicity increased from ~45-48% to 56-58% upon DNA binding. Both the DNA and protein concentrations were 20 μM in PBS.

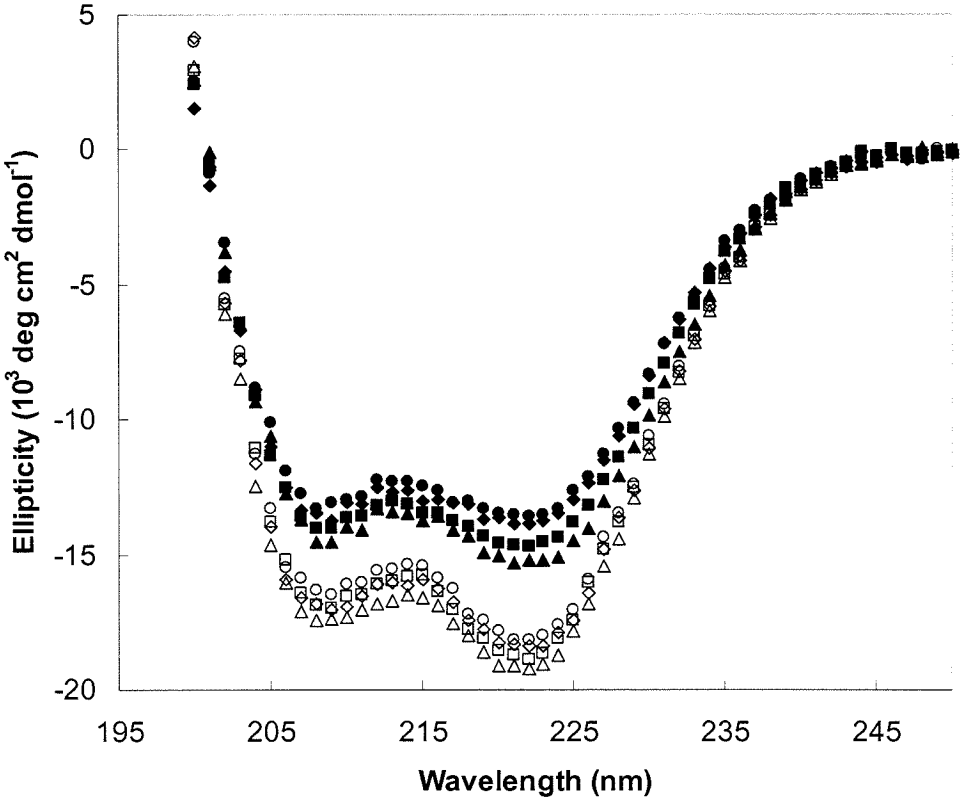
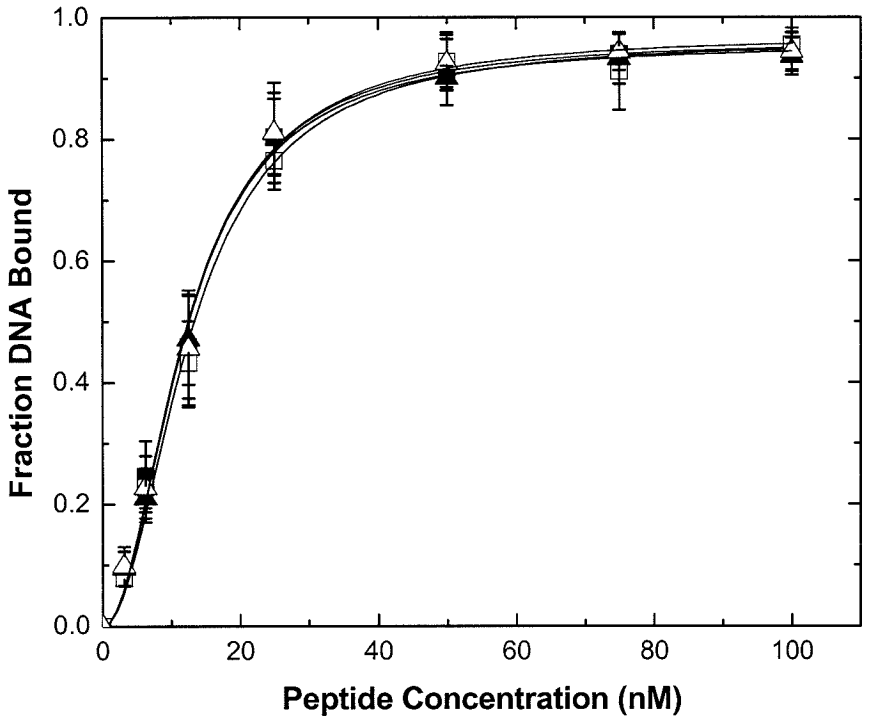


Figure 2.7 Mobility shift assay of VNL (closed triangles), tfv-VNL (open triangles), INL (closed squares), and tfi-INL (open squares) binding to oligonucleotides containing the CREB binding site (5'-TFGAGATGACGTCATCTCCT-3'). Similar binding affinities were observed for all four peptides. No binding to the non-specific sequence (5'-GATCCCAACACGTGTTGGGATC-3') could be detected in this concentration range (data not shown).



2.5 References

1. Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P., *Annu. Rev. Biochem.* **2004,** 73, 147-176.

2. Budisa, N., *Angew. Chem.-Int. Edit.* **2004,** 43, (47), 6426-6463.

3. Wang, L.; Schultz, P. G., *Angew. Chem.-Int. Edit.* **2005,** 44, (1), 34-66.

4. Link, A. J.; Mock, M. L.; Tirrell, D. A., *Curr. Opin. Biotechnol.* **2003,** 14, (6), 603-609.

5. van Hest, J. C. M.; Tirrell, D. A., *FEBS Lett.* **1998,** 428, (1-2), 68-70.

6. Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A., *Angew. Chem.-Int. Edit.* **2000,** 39, (12), 2148-2152.

7. Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A., *FEBS Lett.* **2000,** 467, (1), 37-40.

8. Kirshenbaum, K., Carrico, I., Tirrell, D., *ChemBioChem* **2002,** 3, 235-237.

9. Budisa, N.; Minks, C.; Medrano, F. J.; Lutz, J.; Huber, R.; Moroder, L., *Proc. Natl. Acad. Sci. U.S.A.* **1998,** 95, (2), 455-459.

10. Ibba, M.; Hennecke, H., *FEBS Lett.* **1995,** 364, (3), 272-275.

11. Kim, W. Y.; George, A.; Evans, M.; Conticello, V. P., *ChemBioChem* **2004,** 5, (7), 928-936.

12. Doring, V.; Mootz, H. D.; Nangle, L. A.; Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P.; Marliere, P., *Science* **2001,** 292, (5516), 501-504.

13. Wang, P.; Tang, Y.; Tirrell, D. A., *J. Am. Chem. Soc.* **2003,** 125, (23), 6900-6906.

14. Tang, Y.; Tirrell, D. A., *J. Am. Chem. Soc.* **2001,** 123, (44), 11089-90.

15. Duewel, H.; Daub, E.; Robinson, V.; Honek, J. F., *Biochemistry* **1997,** 36, (11), 3404-16.

16. Resnati, G., *Tetrahedron* **1993,** 49, (42), 9385-9445.

17. Kukhar, V. P., Soloshonok, V. A., *Fluorine Containing Amino Acids: Synthesis and Properties*. John Wiley & Sons: New York, **1995**.

18. Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A., *Biochemistry* **2001,** 40, (9), 2790-2796.

19. Bilgicer, B.; Fichera, A.; Kumar, K., *J. Am. Chem. Soc.* **2001,** 123, (19), 4393-4399.

20. Sheppard, W. A., Sheetz, C. M., *Organic Fluorine Chemistry*. W. A. Benjamin, Inc.: New York, **1969**.

21. Lee, K. H.; Lee, H. Y.; Slutsky, M. M.; Anderson, J. T.; Marsh, E. N. G., *Biochemistry* **2004,** 43, (51), 16277-16284.

22. Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A., *Angew. Chem.-Int. Edit.* **2001,** 40, (8), 1494-1496.

23. Bilgicer, B.; Kumar, K., *Tetrahedron* **2002,** 58, (20), 4105-4112.

24. Bilgicer, B.; Xing, X.; Kumar, K., *J. Am. Chem. Soc.* **2001,** 123, (47), 11815-11816.

25. Wang, P.; Fichera, A.; Kumar, K.; Tirrell, D. A., *Angew. Chem.-Int. Edit.* **2004,** 43, (28), 3664-3666.

26. Muller, N., *J. Fluor. Chem.* **1987,** 36, (2), 163-170.

27. Charvillon, F. B.; Amouroux, R., *Synth. Commun.* **1997,** 27, (3), 395-403.

28. Koziara, A.; Zwierzak, A., *Synthesis* **1992**, (11), 1063-1065.

29. Hope, I. A.; Struhl, K., *Cell* **1986,** 46, (6), 885-894.

30. WinReedit and Win Match were written by J. W. Lary and D. A. Yphantis and are available on the National Analytical Ultracentrifugation Facility (NAUF) FTP site.

31. Winnl106 was written by M. L. Johnson, J. W. Lary, and D. A. Yphantis and is available on the National Analytical Ultracentrifugation Facility (NAUF) FTP site.

32. Sednterp was written by D. T. Hayes, T. M. Laue, and J. Philo and is available on the Reversible Associations in Structural and Molecular Biology anonymous FTP site.

33. CD Spectra Deconvolution was written by G. Bohm and is available on http://bioinformatik.biochemtech.uni-halle.de/ internet site.

34. Schneider, J. P.; Lear, J. D.; DeGrado, W. F., *J. Am. Chem. Soc.* **1997,** 119, (24), 5742-5743.

35. Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T., *Science* **1993,** 262, (5138), 1401-1407.

36. Potekhin, S. A.; Medvedkin, V. N.; Kashparov, I. A.; Venyaminov, S. Y., *Protein Eng.* **1994,** 7, (9), 1097-1101.

37. Oakley, M. G.; Kim, P. S., *Biochemistry* **1998,** 37, (36), 12603-12610.

38. Akey, D. L.; Malashkevich, V. N.; Kim, P. S., *Biochemistry* **2001,** 40, (21), 6352-6360.

39. Zitzewitz, J. A.; Bilsel, O.; Luo, J. B.; Jones, B. E.; Matthews, C. R., *Biochemistry* **1995,** 34, (39), 12812-12819.

40. Sosnick, T. R.; Jackson, S.; Wilk, R. R.; Englander, S. W.; DeGrado, W. F., *Proteins: Struct., Funct., Genet.* **1996,** 24, (4), 427-432.

41. Creamer, T. P.; Rose, G. D., *Proc. Natl. Acad. Sci. U.S.A.* **1992,** 89, (13), 5937-5941.

42. Acharya, A.; Ruvinov, S. B.; Gal, J.; Moll, J. R.; Vinson, C., *Biochemistry* **2002,** 41, (48), 14122-14131.

43. Burton, S. G.; Cowan, D. A.; Woodley, J. M., *Nat. Biotechnol.* **2002,** 20, (1), 37-45.

44. Dordick, J. S., *Biotechnol. Progr.* **1992,** 8, (4), 259-267.

45. Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K., *Nature* **1990,** 347, (6293), 575-578.

46. Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C., *Cell* **1992,** 71, (7), 1223-1237.

47. Konig, P.; Richmond, T. J., *J. Mol. Biol.* **1993,** 233, (1), 139-154.

48. Keller, W.; Konig, P.; Richmond, T. J., *J. Mol. Biol.* **1995,** 254, (4), 657-667.

49. Oneil, K. T.; Shuman, J. D.; Ampe, C.; Degrado, W. F., *Biochemistry* **1991,** 30, (37), 9030-9034.

50. Since the N-terminal leader sequence, including the 6xHis tag, was fused to the peptides, we wanted to determine the effect of this sequence, if any, on the

DNA binding activity and specificity of the proteins. Therefore, a control peptide was expressed and purified which consists of the wild-type sequence of 56 residues encoding the DNA binding and leucine zipper regions from GCN4, fused to the same N-terminal leader sequence.

51. Sellers, J. W.; Vincent, A. C.; Struhl, K., *Mol. Cell. Biol.* **1990,** 10, (10), 5077-5086.

52. Hollenbeck, J. J.; Oakley, M. G., *Biochemistry* **2000,** 39, (21), 6380-6389.

53. Bird, G. H.; Lajmi, A. R.; Shin, J. A., *Biopolymers* **2002,** 65, (1), 10-20.

54. O'Shea, E. K., Klemm, J. D., Kim, P. S., Alber, T., *Science* **1991**, 254, (5031), 539-544.

Chapter 3

Biosynthesis of Stable Coiled-Coils Bearing (2*S*, 4*R*)-5',5',5'-Trifluoroleucine and (2*S*, 4*S*)-5',5',5'-Trifluoroleucine

The synthesis of (2*S*, 4*R*)-5',5',5'-trifluoroleucine and (2*S*, 4*S*)-5',5',5'-trifluoroleucine was performed by Ginevra Giorgi and Krishna Kumar. The remaining work was completed in collaboration with Jin Montclare.

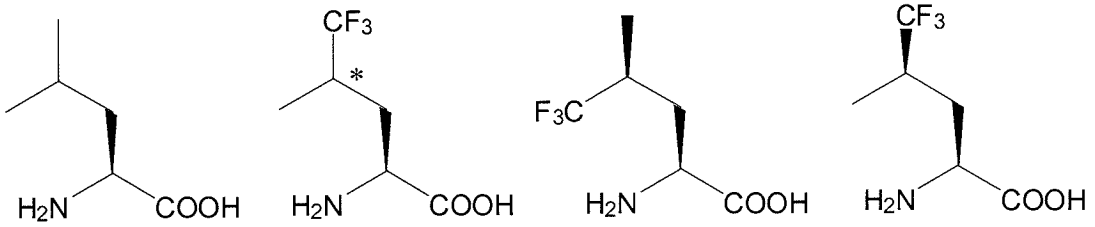
3.0 Abstract

Stereochemical effects on two trifluoroleucine isomers were explored with respect to their ability to support protein biosynthesis as well as their effects on protein structure and function. In a leucine-auxotrophic *E. coli* expression host, *in vivo* incorporation assays showed efficient translation of A1, a *de novo* engineered coiled-coil protein, in the presence of either stereoisomer. *In vitro* activation assays using the wild-type *E. coli* leucyl-tRNA synthetase revealed 100- and 900-fold slower activation rates for SS- and SR-TFL, respectively, as compared to leucine. *In vivo* incorporation levels were consistent with the *in vitro* kinetic results, with higher levels of replacement (91%) for SS-TFL than for SR-TFL (80%), as determined by quantitative amino acid analysis. Upon incorporation of either stereoisomer, secondary structure was unaffected while an 11 °C increase in T_m was observed. However, an equimolar mixture of SS- and SR-A1 exhibited a further increase of 3 °C in T_m as compared to that of the homodimers, suggesting improved packing between stereochemically different trifluoroleucines.

3.1 Introduction and Background

Non-canonical amino acid incorporation into proteins provides a powerful technique for the creation of macromolecules with novel properties.[1-4] In particular, 5',5',5'-trifluoroleucine (TFL, **2**), has been utilized as a hyperhydrophobic leucine surrogate in various contexts (Scheme 3.1).[5, 6] When incorporated into the hydrophobic cores of certain coiled-coil proteins, TFL triggers an increase in stability, rendering proteins more resistant toward both thermal and chemical denaturation.[7, 8] Furthermore, due to its nearly isosteric nature as compared to leucine, native protein folds and biological activity are retained in many circumstances.[9] The canonical amino acid, leucine (**1**), contains no other stereogenic centers besides C_α. However, replacement of a methyl group by a trifluoromethyl group at C_γ introduces an additional stereogenic center, resulting in two possible stereoisomers, (2*S*, 4*S*)-5',5',5'-trifluoroleucine (**3**) and (2*S*, 4*R*)-5',5',5'-trifluoroleucine (**4**).

Scheme 3.1 Structures of leucine and its fluorinated analogs. * designates both S and R forms.



1 **2** **3** **4**

The fidelity of translational incorporation is governed in part by the activation of the amino acid by the aminoacyl-tRNA synthetase (AARS). Although some AARSs

have been known to be tolerant to non-canonical substrates, they also can demonstrate stereoselectivity for particular amino acids and analogs. Isoleucine is one of two canonical amino acids that contain two chiral centers: one at the α carbon and another at the β carbon. The (*2S, 3S*) stereoisomer of 2-amino-3-methyl-heptanoic acid is incorporated into proteins naturally. The (*2S, 3R*) isomer is not incorporated into proteins, although there is evidence that it is bound and activated by isoleucyl-tRNA synthetase (IleRS).[10-13] The importance of stereochemistry also carries over to non-canonical amino acids in determining their possible fate as a substrate for protein synthesis in *E. coli*. Recent results demonstrate that the isoleucine analog 2-amino-3-methyl-4-pentenoic acid is accepted only in its (2*S*,3*S*) form, while 4,4,4-trifluorovaline (4TFV) is accepted only in its (2*S*,3*R*) form,[14, 15] indicating that stereochemistry can influence incorporation levels. Since stereochemistry can be critical for protein translation, we sought to explore whether the leucyl-tRNA synthetase (LeuRS) exhibited a stereochemical preference for either of the trifluoroleucines.

Here, we investigate the effects of TFL stereochemistry on protein synthesis and stability. We demonstrate that both **3** and **4** are activated and incorporated into recombinant proteins prepared in *Escherichia coli.* Coiled-coil homo- and heterodimers bearing **3** or **4** exhibit improved stability, with the heterodimer showing enhanced thermal stability relative to the homodimers.

3.2 Materials and Methods

3.2.1 Fluorinated Amino Acids

The amino acids **3** and **4** were obtained from Prof. Krishna Kumar, while the mixture of D,L-trifluoroleucine was purchased from Oakwood Products (West Columbia, SC).

3.2.2 Protein Biosynthesis and Purification

Leucine auxotrophic *E. coli* strain LAM1000 transformed with pREP4 (Qiagen) was used as the host to express A1, which was encoded within pQEA1 under the control of a T5 promoter. Protein expression and purification were performed as described previously.[7]

3.2.3 Protein Characterization

CD data were collected on an Aviv 62DS spectropolarimeter (Lakewood, NJ) using a 1 mm path length cell. Protein solutions, all at 10 μM concentration, were prepared in phosphate buffered saline, pH 7.4. Wavelength scans were taken from 195 to 250 nm with points taken every 1 nm. Temperature scans were performed from 0 to 95 °C in 1.5 °C steps. Each plot represents an average of three scans.

3.2.4 Activation Kinetics

An N-terminal hexahistidine-tagged LeuRS fusion was expressed and purified as previously reported.[16] Measurement of the rates of activation of leucine and analogs was performed by an ATP-PP_i exchange assay. The assay buffer conditions were 30 mM HEPES, pH 7.4, 10 mM $MgCl_2$, 1 mM DTT, 2 mM ATP, and 2 mM [^{32}P]-PPi (0.5 TBq/mol). A fixed concentration of 75 nM of His_6-LeuRS was used. Depending upon

the activity of the enzyme toward the substrate, the following concentration ranges were chosen for kinetics (**1**: 0.6-312.5 μM; **2**, **3**, **4**: 6.1-6250 μM). At various time intervals, the reaction mixture was quenched by addition of 200 mM PP_i, 7% w/v $HClO_4$ and 3% activated charcoal. The charcoal was washed twice with 10 mM PP_i and 0.5% $HClO_4$, and the absorbed ATP was counted via liquid scintillation counting. Kinetic parameters (k_{cat} and K_m) were determined by nonlinear curve fitting of the data to a Michaelis-Menten model.

3.3 Results and Discussion

*3.3.1 Biosynthesis of A1 in the Presence of **3** and **4***

The stereoisomers **3** and **4** were evaluated for incorporation into the *de novo* engineered leucine zipper protein, A1[17], using an *E. coli* auxotrophic strain. As a control, A1 was also expressed in the presence of **1**. SDS PAGE results show that both **3** and **4** efficiently supported protein synthesis as indicated by the strong bands in lanes d and e, respectively (Figure 3.1). The enhanced electrophoretic mobility of the fluorinated proteins relative to the control lane (b) was consistent with earlier observations of the behavior of A1 prepared in cultures supplemented with **2**.[7] Following purification on a Ni-NTA affinity column, protein yields were determined to be 18 ± 4 mg/L and 9 ± 3 mg/L for proteins bearing **3** (SS-A1) and **4** (SR-A1), respectively, compared to 45 ± 6 mg/L for A1 prepared with **1** (Leu-A1).

3.3.2 Determination of Replacement Levels of 3 and 4

MALDI mass spectrometry analysis of tryptic A1 fragments was performed to detect the extent of substitution by **3** or **4**. The fragment, **L**KNEIED**L**KAEIGD**L**NNTSGIR, corresponding to residues 46-67 of A1 with expected mass of 2442.7 Da, contains three leucine sites. Fragments that correspond to 0, 1, 2, and 3 sites of replacement of leucine by either **3** or **4** were observed (Figure 3.2). The expected mass difference of 57 Da was observed for each leucine substitution. MALDI and quantitative amino acid analysis indicated extents of incorporation of 81 $\pm$ 4 % for SR-A1 and 89 $\pm$ 4 % for SS-A1.

3.3.3 In Vitro Kinetic Assays

The rates of activation of **1** and fluorinated analogs **2-4** by LeuRS were characterized by an *in vitro* ATP-PP_i exchange assay. The kinetic parameters are shown in Table 3.1. The k_{cat}/K_m values demonstrated that **3** was a better substrate for LeuRS than **4**, by approximately nine fold, consistent with the slightly higher yields and incorporation levels mentioned above. As expected, k_{cat}/K_m for **2** fell between the values of **3** and **4**. The activation rates for both **3** and **4** were within the range of rates that have been shown to support protein synthesis in conventional hosts for other non-canonical amino acids.[18] The fact that both **3** and **4** are incorporated into proteins in *E. coli* is also consistent with a previous study that demonstrated the incorporation of hexafluoroleucine into proteins when *E. coli* LeuRS was overexpressed.[16] Fluorination at either of the C_δ positions is tolerated by LeuRS with a slight preference for the SS stereoisomer.

3.3.4 Characterization of SS-A1 and SR-A1

The secondary structures of the fluorinated and leucine forms of A1 were examined by circular dichroism (CD) spectroscopy. Because the proteins can form dimers, Leu-A1, SR-A1, and SS-A1 represent homodimers, whereas the equimolar mixture of SR-A1 and SS-A1 might form the heterodimer. All four spectra overlapped with one another, exhibiting greater than 90% helical content as determined from molar ellipticity at 222 nm (Figure 3.3A). Incorporation of TFL does not affect the secondary structure of the proteins.

Previous studies illustrated that the incorporation of fluorinated amino acids into the hydrophobic core enhanced the thermostability of coiled coils.[8, 9, 16, 19] However the extent of stabilization varied depending on what type of fluorinated amino acid was used.[16, 19] To establish if stereochemistry influenced the stability of coiled coils, thermal denaturation experiments were conducted and monitored via CD spectroscopy (Figure 3.3B). Both SR-A1 and SS-A1 displayed melting temperatures (T_m) 11 °C higher than that of Leu-A1. However, the equimolar mixture of the SR-A1 and SS-A1 heterodimer showed an additional 3 °C increase in T_m beyond that of the pure SS-A1 or SR-A1 homodimers. Although this further enhancement in thermostability was modest, it may be attributed to the ability of the coils to pack more compactly into dimers due to the stereochemical differences.[9] As a reference point, when **2** was used in the expression of A1, the ΔT_m was 13 °C.[7] This melting temperature was similar to the melting temperature observed here for the heterodimer.

3.4 Conclusions

Contrary to previous studies with isoleucine and valine analogs, this work demonstrates a limited stereochemical preference by LeuRS with respect to fluorination of C_γ.[14, 15] Both the SR and SS stereoisomers of TFL are incorporated into the protein where **3** is activated more rapidly than **4** by 9-fold. This is supported by evidence of higher protein yields for SS-A1 relative to SR-A1 and increased incorporation levels when A1 is expressed in media supplemented with **3** versus **4**. Neither stereoisomer appears to alter the coiled-coil structure, and substitution of leucine with either **3** or **4** equally enhances the thermostability of the homodimers. The heterodimer shows an additional increase in stability. Further exploration of stereochemical effects on protein stability is underway.

Table 3.1 Kinetic parameters for activation of **1-4** by *E. coli* LeuRS[a]

Substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (*rel*)
1	16.9 ± 4.5	4.22 ± 0.35	1
2	659 ± 103	0.40 ± 0.03	1/412
3	252 ± 92	0.59 ± 0.05	1/107
4	708 ± 280	0.19 ± 0.02	1/933

[a] Substrate **1** was used as the L-isomer; **2** as the mixture of (2*S*, 4*S*), (2*S*, 4*R*), (2*R*, 4*S*), and (2*R*, 4*R*) forms with the concentration referring only to the 2*S* forms; **3** as (2*S*, 4*S*) form and **4** as the (2*S*, 4*R*) form.

Figure 3.1 SDS-PAGE evidence of incorporation of **3** and **4** into recombinant protein A1. Lane **a**: uninduced, lane **b**: 20 amino acids (+L), lane **c**: 19 amino acids (-L), lane **d**: 19 amino acids (-L) plus **3** (250 μM), lane **e**: 19 amino acids plus **4** (250 μM).

Figure 3.2 MALDI mass spectra of the tryptic fragment corresponding to residues 46 through 67 from A1 expressed in media supplemented with either **1** A), **3** B), or **4** C).

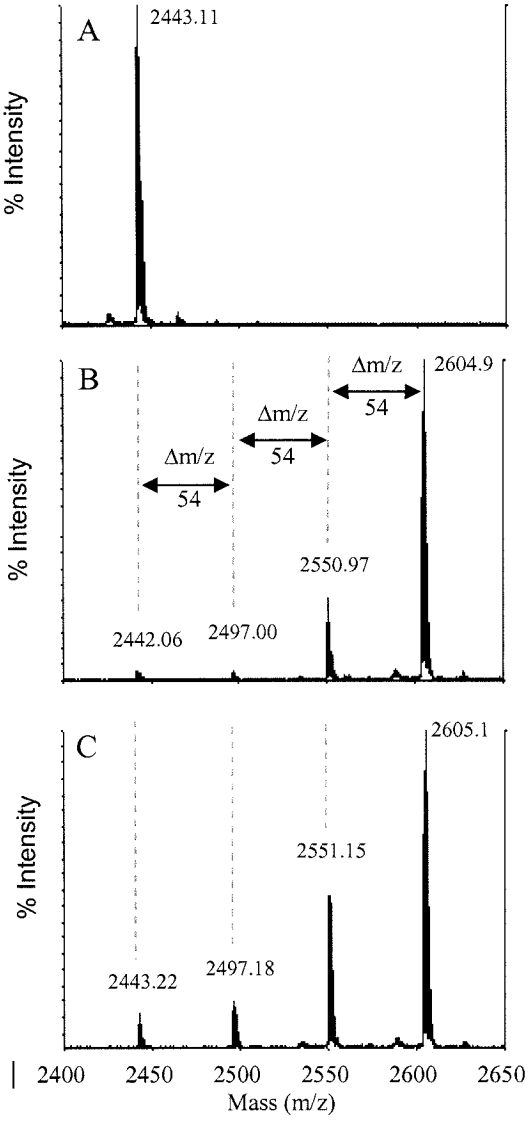
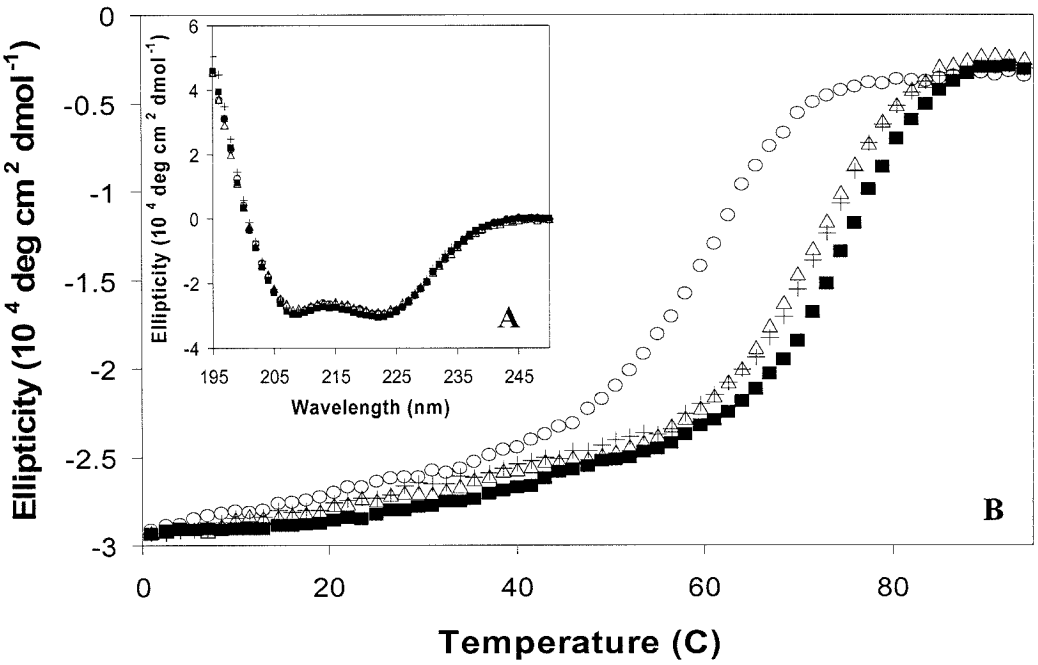


Figure 3.3 Circular dichroism spectra of Leu-A1 (o), SR-A1 (△), SS-A1 (+), and equimolar mixture of SR-A1 and SS-A1 (■). **A**: Wavelength scan at 1 °C, 10 μM protein concentration, PBS buffer, pH 7.4. **B**: Thermal denaturation (1.5 °C interval, 1 minute equilibration time, 10 second averaging time) at 10 μM protein concentration, PBS buffer, pH 7.4.



3.5 References

1. Link, A. J.; Mock, M. L.; Tirrell, D. A., *Curr. Opin. Biotechnol.* **2003,** 14, (6), 603-609.

2. Budisa, N., *Angew. Chem.-Int. Edit.* **2004,** 43, (47), 6426-6463.

3. Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P., *Annu. Rev. Biochem.* **2004,** 73, 147-176.

4. Bilgicer, B.; Kumar, K., *J. Chem. Edu.* **2003,** 80, (11), 1275-1281.

5. Rennert, O. M.; Anker, H. S., *Biochemistry* **1963,** 2, (3), 471-476.

6. Marsh, E. N. G., *Chem. Biol.* **2000,** 7, (7), R153-R157.

7. Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A., *Angew. Chem.-Int. Edit.* **2001,** 40, (8), 1494-1496.

8. Bilgicer, B.; Fichera, A.; Kumar, K., *J. Am. Chem. Soc.* **2001,** 123, (19), 4393-4399.

9. Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A., *Biochemistry* **2001,** 40, (9), 2790-2796.

10. Loftfield, R. B., *Biochem. J.* **1963,** 89, (1), 82-92.

11. Flossdorf J.; Kula, M. R., *Eur. J. Biochem.* **1973,** 36, (2), 534-540.

12. Loftfield R.B.; Eigner, E. A., *Biochim. Biophys. Acta* **1966,** 130, (2), 426-448.

13. Polet, H.; Conrad, M. E., *Proc. Soc. Exp. Biol. Med.* **1969,** 130, (2), 581-586.

14. Mock, M. L.; Michon, T.; van Hest, J. C. M.; Tirrell, D. A., *ChemBiochem* **2006,** 7, (1), 83-87.

15. Wang, P.; Fichera, A.; Kumar, K.; Tirrell, D. A., *Angew. Chem.-Int. Edit.* **2004,** 43, (28), 3664-3666.

16. Tang, Y.; Tirrell, D. A., *J. Am. Chem. Soc.* **2001,** 123, (44), 11089-90.

17. Petka, W. A.; Harden, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A., *Science* **1998,** 281, (5375), 389-392.

18. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A., *FEBS Lett.* **2001,** 502, (1-2), 25-30.

19. Son, S., Tanrikulu, I. C., and Tirrell, D.A., *ChemBioChem* **2006,** 7, (8), 1251-1257.

Chapter 4

Laboratory Evolution of Chloramphenicol Acetyltransferase Containing Trifluoroisoleucine

Portions of this chapter are adapted from a paper in preparation by Soojin Son and David A. Tirrell.

4.0 Abstract

Laboratory evolution was utilized to generate a fluorinated chloramphenicol acetyltransferase (CAT) whose activity was comparable to that of the wild-type enzyme. Upon *in vivo* incorporation of 5,5,5-trifluoroisoleucine into nine isoleucine positions within CAT, the enzyme's catalytic efficiency (k_{cat}/K_m) dropped from 10.2 $\pm$ 0.8 μM^{-1} min^{-1} to 3.9 $\pm$ 0.5 μM^{-1} min^{-1}. Four rounds of random mutagenesis, enrichment, and screening were performed, yielding a seven-amino acid fluorinated mutant (tfi-G4) whose activity was 2.8-fold higher than the fluorinated parent CAT (tfi-WT). The total number of isoleucine codons decreased only by one, resulting in an active, yet still highly fluorinated enzyme. Characterization of the wild-type hydrogenated and fluorinated (ile-WT, tfi-WT) counterparts as well as the G4 pair (ile-G4, tfi-G4) of enzymes showed that the fluorinated enzyme is less stable than its hydrogenated counterpart. The thermostability of the G4 pair was higher than that of the wild-type pair, resulting in a fluorinated enzyme (tfi-G4) with similar thermostability as the wild-type. The *pKa* of the catalytically essential histidine was not affected upon incorporation of 5TFI. This study deepens our understanding of protein behavior upon introduction of fluorinated amino acids and demonstrates the effectiveness of laboratory evolution as a tool complementary to incorporation of non-canonical amino acids.

4.1 Introduction and Background

The ability to incorporate non-canonical amino acids into proteins has expanded the realm of engineered proteins.[1-4] In particular, through residue-specific incorporation, where all-or a fraction-of a particular canonical amino acid is replaced by the non-canonical analog, substantial changes in the overall physical and chemical properties of proteins could be obtained.[5-10] Among the numerous non-canonical amino acids that have been incorporated *in vivo* through the use of auxotrophic *E. coli* hosts, fluorinated hydrophobic amino acids have been shown to be excellent analogs of their hydrogenated counterparts.[6, 9, 11] When trifluoroleucine, trifluoroisoleucine, and trifluorovaline, were incorporated into peptides derived from the yeast transcription factor GCN4, protein structure was not disturbed, while an increased resistance to chemical and thermal denaturation was observed. In addition, these fluorinated GCN4 proteins maintained their wild-type DNA-binding affinity as well as specificity. However, as expected, in other instances, the incorporation of non-canonical amino acids can lead to decreased activity and/or thermostability.[5, 12, 13] In these circumstances, we may turn to other protein engineering techniques to recover or even enhance the lost function or property.

Directed evolution was chosen as that tool.[14, 15] Mimicking natural evolution, directed evolution aims to generate proteins with new or improved properties by screening or selecting for the desired function from large pools of protein variants.[16-18] By exploring the sequence space provided by the 20 natural amino acids, directed evolution has been used successfully to engineer proteins with enhanced activity, thermostability, altered substrate specificity, and improved function in non-natural applications.[19-22] Recent study from our laboratory has successfully demonstrated using

laboratory evolution to recover thermostability of a protein, which was lost upon the incorporation of a non-canonical amino acid.[23] When 5,5,5-trifluoroleucine was incorporated into chloramphenicol acetyltransferase, the half life at 60 °C was reduced from 101 to 5 minutes. After two rounds of random mutagenesis and screening, a fluorinated mutant was found that exhibited similar thermostability to the wild-type enzyme.

In this study, we explore the effects of 5,5,5-trifluoroisoleucine (5TFI) on the activity of chloramphenicol acetyl transferase (CAT). Following the biochemical characterization of the fluorinated enzyme, we use directed evolution in the presence of 5TFI to engineer fluorinated CAT with enhanced activity. Through the use of error-prone PCR to generate random mutants and an enrichment step that relies on the ability of mutants to express functional CAT using 5TFI, we found a fluorinated mutant whose activity is almost 3-fold higher than the fluorinated enzyme with the wild-type sequence. In addition, the effects of fluorination on sensitivity to thermal and chemical denaturation as well as to changes in pH were determined in order to expand our understanding of behavior of the proteins modified with non-canonical amino acids.

4.2 Materials and Methods

4.2.1 Cloning of CAT Expression Plasmid

The CAT expression plasmid was constructed from pQE 80L (Qiagen). In pQE 80L, a silent copy of a CAT gene is present between two transcription terminators. Although this gene was not being expressed, in order to prevent any complications that

might have resulted during future experiments, a portion of this gene was eliminated. The plasmid was digested with restriction enzymes *Bpu*I and *Nco*I, removing 528 base pairs from the silent CAT gene. The incompatible ends were blunted using DNA polymerase I Large (Klenow) Fragment, and a blunt end ligation was performed to regenerate a circular plasmid.

Into the modified pQE80L, the isoleucyl-tRNA synthetase (IleRS) gene was inserted using *Nhe*I restriction sites. The IleRS gene was cloned directly from *E. coli* genomic DNA. Within this IleRS gene exists a *BamH*I site. Since *BamH*I and *Hind*III restriction sites within the multiple cloning region of pQE80L will later be used to clone in the target CAT gene as well as its library members, the *BamH*I site in the IleRS gene needed to be removed. This was accomplished through a silent mutation using Quickchange mutagenesis with the following primers: 5'-GGCGTGCAGTGGATTCCGGACTGGGGCCAG-3' and its complement. The resulting plasmid is referred to as pQ80E-IRS (7335 bp).

The plasmids pQ80E-IRS and previously prepared pCCAT (3921 bp)[23] were both digested with *Eco*RI and *Hind*III to release part of the multiple cloning region C-terminal to the hexahistidine tag and CAT gene, respectively. The 663 bp CAT gene released from pCCAT and 7293 bp vector released from pQE80-IRS were isolated, purified, and ligated together, resulting in pQ80E-CAT-IRS. The T5 promoter controls the inducible expression of N-terminally 6xHis-tagged CAT.

4.2.2 Construction of Protein Expression System

An isoleucine auxotrophic derivative of *E. coli* strain BL21(DE3), designated AIV (*E. coli* B F$^-$ *ompT hsdS*(r_B^- m_B^-) *gal dcm* λ(DE3) *ilv*D691), constructed in our laboratory, was used as the host strain for protein expression.[6] The final expression system AIV[pQ80E-CAT-IRS] was obtained by transformation of pQ80E-CAT-IRS into AIV.

4.2.3 Generation of CAT Library of Mutants

The libraries of mutants were generated by error-prone PCR using Mutazyme II (Stratagene) according to manufacturer's protocols. The purified insert library was digested with the *BamH*I and *Hind*III and ligated into *BamH*I/*Hind*III digested pQ80E-CAT-IRS. The ligation mixture was transformed into AIV cells through electroporation and transformants were recovered in 2xYT liquid medium supplemented with ampicillin (200 μg/mL). Approximately 10,000 transformants were generated in the first library while 100,000-300,000 were generated for each of the next three libraries. Based on the sequences of 15 randomly chosen transformants, the average mutation rate of the first library was 0.2 nucleotides per gene while the following three libraries exhibited higher mutation rates of approximately 2.9 to 3.8 nucleotide mutations per gene.

4.2.4 Enrichment Scheme

M9 minimal medium used in the selection process was always supplemented with 1 mM $MgSO_4$, 1 mM $CaCl_2$, 0.4 wt% glucose, 1 mg/L thiamine, and ampicillin (100

μg/mL). Other supplements are described in detail in the following procedure. The selection scheme described below is also illustrated in Scheme 4.1.

Following transformation, mutants were grown in M9 minimal medium supplemented with 20 natural amino acids (40 μg/mL each). After the cultures had grown to OD_{600} of 0.8, the cells were pelletted through centrifugation (6000 g, 7 min). Following two washes with 0.9 % NaCl, the cell pellet was resuspended in M9 medium supplemented with 19 natural amino acids, reduced amounts of isoleucine (1 to 4 μg/mL), and 5TFI (250 μg/mL). After 10 minutes, IPTG (0.1 mM) was added to induce expression of CAT. After 15 minutes, chloramphenicol (CAM) was added. The concentration of CAM was varied from 35 μg/mL to 150 μg/mL depending on the progress of the enrichment. After 5 hours, a second media shift procedure was applied to remove IPTG and CAM from the medium. The cell pellets were resuspended in fresh M9 medium containing 20 natural amino acids. A small aliquot of the culture was added to fresh M9 medium containing 20 amino acids, without CAM or IPTG, resulting in 1 to 200 dilution based on volume, and allowed to grow overnight. Fresh M9 medium containing 20 natural amino acids was inoculated with a small aliquot of the overnight cultures, resulting in 1 to 10 dilution based on volume. Once these cultures reached OD_{600} of 0.8, the whole cycle was repeated. Following the second round CAT expression period, a small aliquot of the expression cultures was plated onto 2xYT plates containing ampicillin (200 μg/mL). In the first generation, 176 clones were picked from the enriched library and 880 to 1056 clones were picked from each of the following three libraries. The kinetics of CAT mutants were analyzed in 96-well plate format.

4.2.5 Protein Expression in 96-Well Plates

M9 minimal medium used in the protein expression process was always supplemented with 1mM $MgSO_4$, 1mM $CaCl_2$, 0.4 wt % glucose, and 1mg/L thiamine. Other supplements are described in detail in the following procedure. All growth stages and protein induction were performed in a humidified shaker apparatus (Kumer ISF-1-W) at 30 °C and 80% humidity.

Following two rounds of enrichment as described above, selected clones were transferred to 96 deep-well plates containing M9 medium supplemented with 20 amino acids (40 μg/mL) and antibiotics (ampicillin 200 μg/mL). To serve as controls, eight colonies of AIV[pQE80-CAT-IRS] were also picked and transferred into each deep-well plate under the same conditions. Cultures were grown overnight. The resultant starter culture plates were then used to inoculate new 96 deep-well plates containing M9 minimal medium supplemented with 20 amino acids (40 μg/mL) and antibiotics (ampicillin 200 μg/mL) via pin transfer. After 12 hours, the media shift procedure was applied through centrifugation (5000 rpm, 10 min, 4 °C) and resuspension of cells in 0.9% NaCl using Multimek 96-Channel Automated Pipettor (Beckman). Following two wash steps, the cell pellet was resuspended in M9 minimal medium supplemented with 19 amino acids (40 μg/mL each, minus Ile), plus TFI (250 μg/mL), and 1.0 mM IPTG. After 6 hours, cells were harvested by centrifugation (5000 rpm, 10 min, 4 °C). The cell pellets were stored at –80 °C overnight and thawed at room temperature in 100 mM Tris, pH 7.8, and 1.0 mg/mL lysozyme (Sigma). Cells were lysed through repeated pipetting using Multimek 96. Cellular debris was pelleted by centrifugation (5000 rpm, 10 min, 4 °C) and aliquots of the supernatant were used for analysis.

4.2.6 Protein Expression and Purification in Larger Scale

M9 minimal medium (200 mL) supplemented with 1mM $MgSO_4$, 1mM $CaCl_2$, 0.4 wt% glucose, 1 mg/L thiamine, and ampicillin (100 μg/mL) was inoculated with 4 mL of fresh overnight culture (M9) of the expression strain. After the culture had grown to OD_{600} of 1.0, the media shift procedure was applied. The cell pellet was then resuspended in 200 mL M9 medium without isoleucine. Following resuspension, 50 mL of the culture was added to a flask containing isoleucine (40 μg/mL) and the remaining 150 mL was aliquoted into a flask containing 5TFI (250 μg/mL). IPTG (1 mM) was added after 10 minutes to induce expression of CAT. Cells were collected after 3 hours by centrifugation (6000 g, 8 min, 4°C) and frozen at –80 °C.

The cell pellets were thawed on ice and resuspended in 50 mM Tris (pH 7.8) plus lysozyme (0.5 mg/mL, Sigma). Following sonication for 1 minute, cell debris was sedimented (10,000g, 30 min, 4 °C). β-mercaptoethanol (10 mM, Sigma) was added to the supernatant, which was subjected to Ni-NTA chromatography according to manufacturer's protocols under native conditions (Qiagen). After binding the protein to the resin and loading the sample onto the column, the column was subjected to wash buffer 1 (50 mM Tris pH7.8, 10 mM imidazole, 10 mM β-mercaptoethanol), wash buffer 2 (50 mM Tris pH7.8, 10 mM imidazole, 10% ethanol v/v), wash buffer 3 (50 mM Tris pH 7.8, 20 mM imidazole), and wash buffer 4 (50 mM Tris pH 7.8, 40 mM imidazole). Purified CAT was eluted with elution buffer (50 mM Tris, pH 7.8, 200 mM immidazole). CAT concentrations were determined by Bradford assay (BioRad) with bovine serum albumin (BSA) as the standard.

4.2.7 CAT Activity Assay

The CAT activity assays were performed according to literature procedures in buffer containing 50 mM Tris pH 7.8, 1 mM 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, Pierce), 0.4 mM acetyl coenzyme A (MP Biomedicals), 0.25 mg/mL BSA, and varying concentrations of CAM (0.78 to 100 μM).[24] All kinetic measurements were performed with saturating acetyl coenzyme A. The initial rates were measured by monitoring the absorbance change at 412 nm as a function of time using a 96-well spectrophotometer (Safire 2, TECAN).

4.2.8 Effect of Temperature on CAT Activity

Purified CAT samples were incubated at various temperatures for 30 minutes using a Thermomixer (Eppendorf), then immediately cooled on ice. CAT activity was assayed at 25 °C in buffer containing 50 mM Tris pH 7.8, 1 mM DTNB, 0.4 mM acetyl coenzyme A, 0.25 mg/mL BSA, and 100 μM CAM.

4.2.9 Effect of Chemical Denaturant on CAT Activity

Purified CAT samples were equilibrated at a given concentration of GuHCl (99.8% UltraPure guanidine hydrochloride, Sigma) for 30 minutes at room temperature. CAT activity was assayed in buffer containing the given concentration of GuHCl, 50 mM Tris pH 7.8, 1 mM DTNB, 0.4 mM acetyl coenzyme A, 0.25 mg/mL BSA, and 100 μM CAM.

4.2.10 pKa of the Catalytic Histidine Residue

In order to examine the pH-dependence of CAT activity, the following buffers (50 mM) were used over the specified pH ranges: Mes/NaOH pH 5.1-6.7; Mops/NaOH, pH 6.7-7.5; Hepes/NaOH, pH 7.5-8.5. Activity was determined in both buffers at pH values where the buffer ranges overlap. No significant buffer effects were observed. The assay mixture contained 0.4 mM acetyl coenzyme A, 0.2 mg/mL BSA, and 100 μM CAM. At each time interval, 40 μL of the reaction mixture was removed and added to 120 μL of 6 M GuHCl, 50 mM Tris, pH 8.0, which immediately stopped the reaction. Each time point mixture was assayed at room temperature in the presence of 1 mM DTNB. Control incubations were performed concurrently at each pH value in the absence of CAT in order to correct for any background rate.

4.2.11 Circular Dichroism Spectroscopy

In order to perform structural studies of CAT proteins using circular dichroism spectroscopy, purified proteins were subjected to dialysis (Pierece, Slide-A-Lyzer Dialysis Cassette, 3500 MWCO) for the removal of imidazole. Concentrations of all dialyzed proteins in 50 mM Tris, pH 7.8, were determined using the Bradford assay (BioRad) with BSA as the standard. Circular dichroism spectra were recorded on an Aviv 62DS spectropolarimeter. All experiments were performed in a rectangular cell with path length of 1 mm with a thermostatically controlled cuvette. All measurements were completed at 5 μM protein concentration.

Wavelength scans were performed at 25 °C from 250 to 195 nm. Data points were collected every 1 nm with an averaging time of 5 seconds at each wavelength.

Temperature scans were performed at 222 nm from 25 °C to 89.5 °C with points taken every 1.5 °C. Each data point was collected after 1 minute of thermal equilibration at the desired temperature with an averaging time of 10 seconds at each temperature. Chemical denaturation curves were obtained from guanidine hydrochloride titration experiments. Protein samples were prepared in varying concentrations of GuHCl, 50 mM Tris, pH 7.8 and allowed to equilibrate for 30 minutes. Data was collected at 222 nm, 25 °C, with averaging time of 30 seconds. The pH unfolding experiment was performed using the following buffers: 50 mM glycine (pH 2.0), 50 mM citric acid (pH 3.0), 50 mM acetic cid (pH 4.0, pH 5.0), 50 mM MES (pH 6.0), 50 mM NaH_2PO_4 (pH 7.0, pH 8.0), 50 mM H_3BO_3 (pH 9.0), 50 mM CAPS (pH 10.0, pH 11.0). Data was collected at 222 nm, 25 °C, with averaging time of 30 seconds.

4.3 Results and Discussion

4.3.1 Characterization of ile-CAT and tfi-CAT

Purified CAT activities were assessed colorimetrically by monitoring the absorbance at 412 nm in the presence of the enzyme, acetylcoenzyme A, DTNB, and CAM.[24] BSA was also added to the reaction mixture in order to prevent adhesion of the proteins to the microwells, which has been shown to be a significant source of protein loss in low concentrations of proteins, and also is more problematic in the case of the fluorinated proteins.[25] The level of 5TFI replacement was determined to be 85 $\pm$ 4 % based on MALDI mass spectrometry and quantitative amino acid analysis. Steady-state kinetic analysis was used to calculate K_m and k_{cat} values CAT expressed in media

containing 20 canonical amino acids (ile-CAT) and fluorinated CAT containing 5TFI (tfi-CAT). Based on the crystal structure, none of the nine isoleucine residues is near the substrate binding pocket or participates directly in catalysis (Figure 4.2). The recognition of and the affinity to the substrate, CAM, were not disturbed upon the incorporation of 5TFI as indicated by similar K_m values, 17.1 $\pm$ 1.2 μM for ile-CAT and 16.9 $\pm$ 1.5 μM for tfi-CAT. However, due to a lowered turnover number, k_{cat}/K_m was reduced by more than two-fold of the wild-type upon fluorination, 10.2 $\pm$ 0.8 μM^{-1} min^{-1} to 3.9 $\pm$ 0.5 μM^{-1} min^{-1}, resulting in a decreased overall efficiency of the enzyme. The measured k_{cat} and K_m values were within the range reported in literature.[26, 27]

4.3.2 Enrichment and Screening

In order to recover the activity of the tfi-CAT, mutants were generated using random mutagenesis, and enriched and screened for those exhibiting increased activity after incorporation of 5TFI. In the first generation, approximately 10,000 mutants with an average mutation rate of 0.2 were generated and subjected to the enrichment process. The enrichment process, shown in Scheme 4.1, was designed to enrich for those mutants that are able to express functional CAT in the presence of 5TFI. With CAM in the medium during the protein induction period, only those mutants with functional fluorinated CAT will grow. The rate of cell growth is limited in medium containing 5TFI (less than 2-fold increase in OD_{600} over the course of 5 hours). Therefore, to increase the viability of the cells and their growth rates, small amounts of isoleucine were added to the medium. Since CAM is a bacteriostatic antibiotic,[28] the enrichment process does not kill those cells with nonfunctioning CAT. Following the protein induction step in the

presence of 5TFI and CAM, which leads to a 2.5- to 3-fold increase in OD_{600}, the cells are returned back to rich medium lacking IPTG. During this growth phase in rich medium, further enrichment takes place of those mutants that had been expressing functioning fluorinated CAT. Presumably, those cells that were stalled in their protein synthesis and cell division due to the lack of functioning fluorinated CAT require additional time to return to normal activity and start dividing. To quantify the extent of enrichment, a mixture (1:1) of wild-type CAT and L116S mutant, whose activity is less than 1% of that of the wild-type, was put through two rounds of enrichment. Plasmids isolated from 25 resulting clones were sequenced. Based on the sequences, approximately 10-fold enrichment of the wild-type CAT was realized.

Following the enrichment process, individual clones were screened in a 96-well format. After a screen and a rescreen of the best 24 clones, those clones showing the highest activity were expressed on a larger scale and kinetics were determined on purified proteins. The best one or two clones, i.e., those showing the highest increase in activity, were subjected to the next round of evolution.

4.3.3 Progress of Evolution: Identification of Improved Clones

Figure 4.1 shows the progress of four rounds of evolution based on the kinetics of purified proteins while Table 4.1 provides more detailed characterization of the CAT mutants. The catalytic activities of the fluorinated enzymes steadily increase through four generations to finally arrive at tfi-G4, which exhibits a slightly higher activity, 11.0 $\pm$ 1.1 μM^{-1} min^{-1}, than the wild-type CAT. Another important trend is that the gap between the isoleucine and fluorinated forms of the same sequence continued to decrease,

resulting in relative activity of tfi-G4 to ile-G4 of 0.86. This indicates that the protein sequence has become more accommodating for 5TFI. The average replacement level of 5TFI for all fluorinated mutants of CAT was 85 $\pm$ 5%, ensuring that the observed CAT activities are contributed by those proteins that are highly fluorinated.

Because more than one mutation was found in clones resulting from generations 2 through 4, it is difficult to make conclusions about the individual mutation effects. However, some possible explanations can be provided for a few of the mutations that are depicted in Figure 4.2. First, it is important to note that the total number of isoleucine positions only decreased by one, with the removal of two (I41V, I119V) and the addition of one (V94I). Presumably, the two positions were less tolerant to fluorination, and therefore, upon mutation to valine, activity was enhanced. However, in general, it appears that accommodating mutations could be found that allow for proper functioning in the presence of 5TFI. The residue flanked by two of the remaining isoleucine positions, H61, was a hotspot in the evolution process, where it was first mutated to a tyrosine in the first generation, and then again to a tryptophan in the second generation. Since it was the only mutation in the first generation, it is clear that H to Y mutation enhanced the fluorinated activity. H61 resides between two isoleucine positions where the two isoleucines are pointing in toward the core and the H61 faces the solvent-exposed surface. We could imagine that because 5TFI exhibits exaggerated hydrophobicity, it may pack more tightly inwards, creating a bigger space for H61 to fill. Therefore, a larger tyrosine residue may allow for better filling of the space, which may explain the further mutation to tryptophan. Three of the mutations, V94I, F95L, and F134L, are in proximity to the catalytic site, with F134 actually lining the intersubunit CAM binding

pocket. F134 is a well conserved residue; among 14 variants of CAT, the position corresponding to residue 134 is occupied by phenylalanine in 12 of them, glutamic acid in the other two.[26, 29] Consequently, F134L is likely to cause some change in catalytic activity, presumably enhancing it in the presence of 5TFI.

4.3.4 Secondary Structure Analysis via Circular Dichroism Spectroscopy

In order to determine the effect of fluorination on the secondary structure of CAT, circular dichroism spectroscopy was performed. At 5 μM protein concentration, the secondary structures of ile-WT, tfi-WT, ile-G4, and tfi-G4 are similar as evidenced by overlapping curves in Figure 4.3. Although slight differences may exist, the overall secondary structure remains undisturbed by the incorporation of 5TFI for both the wild-type and G4 proteins. Furthermore, the mutations introduced into G4 appear to have limited effects on the secondary structure.

4.3.5 Thermostability of CAT Mutants

The thermostability of the final mutant tfi-G4 along with its isoleucine counterpart, ile-G4, were compared to the wild-type pair of enzymes (ile-WT and tfi-WT) through both structural and catalytic analysis. The trends observed from monitoring the remaining activity following incubation at elevated temperatures mirrored those obtained from circular dichroism spectroscopy (Figure 4.4). Both of the isoleucine variants are more thermostable than their fluorinated counterparts. A similar trend was observed when 5,5,5-trifluoroleucine was incorporated into CAT.[23] However, the difference in thermostability between tfi-G4 and ile-G4 was reduced as compared to the wild-type pair.

In addition, the G4 pair exhibited higher thermostability as a whole, resulting in comparable thermostability between the wild-type enzyme (ile-WT) and the fluorinated mutant (tfi-G4). Although the evolution process was not designed to select or screen for enhanced thermostability, the increase in structural stability may contribute towards more efficient catalysis, and therefore be favorable. Furthermore, the increased stability may have led to enhanced evolvability of the mutants, a concept demonstrated both through simulations and experiments in recent work by Arnold and coworkers.[30]

4.3.6 Stability toward Chemical Denaturant

The sensitivity of CAT proteins towards chemical denaturant, specifically guanidine hydrochloride, was similarly analyzed, through both circular dichroism and activity assays. Trends observed in thermostability experiments were again observed in these set of experiments. As shown in Figure 4.5, the fluorinated CAT mutants were more sensitive than their hydrogenated counterparts toward guanidine hydrochloride. The wild-type pair of enzymes was more sensitive than the G4 pair. At 1 M GuHCl, both of the fluorinated variants had lost essentially all activity, while 2 M GuHCl was necessary to abolish the activity of their hydrogenated counterparts. At 2.5 M GuHCl both of the fluorinated variants were essentially unfolded as reported by CD spectroscopy, while 3 M GuHCl was required to unfold the hydrogenated CATs. Furthermore, the evolved G4 pair exhibited higher resistance toward denaturation by GuHCl than the wild-type pair. Based on these results, as well as those from thermostability experiments, it appears that compromised protein folding is a significant factor in the lower catalytic efficiency of tfi-WT. Consequently, improving the stability

of the protein fold may account for much of the observed 3-fold enhancement in the activity of tfi-G4.

4.3.7 pH Effect on CAT Mutants

The effects of pH on folding as well as on activity were explored. Circular dichroism spectroscopy was performed over a pH range of 2 through 11. As expected, all proteins showed the most intense [θ_{222}] near physiological pH (Figure 4.6B). However, the hydrogenated CATs exhibited a greater tolerance to both lower and higher pH, with similar [θ_{222}] over the pH range of 5 to 10. Again, these results indicate that the incorporation of 5TFI into CAT destabilizes the protein, and that evolution recovers part of the loss in stability.

Circular dichroism spectroscopy reports the effect of pH on protein folding. However, the pH effect on catalytic activity is somewhat different. The mechanism of catalysis involves His-193, which acts as a base, abstracting a proton from the primary hydroxyl group of CAM.[26] Therefore, the pK_a of His-193 limits the range of pH where CAT is active. To determine the pK_a, the pH range from 5.1 to 8.5 was explored for the wild-type and G4 enzymes. Since the DTNB (which is used to detect the release of free thiol during catalysis) is sensitive to pH, reactions were stopped by the addition of 6 M GuHCl, 50 mM Tris, pH 8.0 and analyzed at pH 8.0. As shown in Figure 4.6A, relative activity decreases with pH as an ionizable group with an apparent pK_a between pH 6.2 and 6.6. Similar pK_a values between fluorinated and hydrogenated forms suggest that fluorination does not significantly affect the local environment around the catalytically essential histidine (Table 4.1). In addition, the mutations introduced into G4 slightly

lower the pK_a value. As a reference, the pK_a of His-195 in type III CAT, which corresponds to His-193 in type I, has been determined to be 6.62 $\pm$ 0.13, while the pK_a of a free histidine is 6.0.[31]

4.4 Conclusions

Through four rounds of laboratory evolution, a fluorinated mutant CAT (tfi-G4) was found whose activity is 2.8 times that of the fluorinated parent enzyme, tfi-WT. In all, seven amino acid mutations were found that include I41V, H61W, V94I, F95L, R114Q, I119V, and F134L. The number of isoleucine sites decreased only by one, yielding a more active, yet still fluorinated CAT. The k_{cat}/K_m of tfi-G4 was slightly higher than that of the wild type enzyme, ile-WT, while the difference in enzyme efficiency between the fluorinated and hydrogenated forms decreased. These results indicate that at least some of the accumulated mutations act specifically to accommodate the remaining eight 5TFI residues in the enzyme.

The wild-type and the G4 pairs of enzymes were characterized for their folding and activity in response to thermal, chemical, and pH denaturation. Under physiological conditions, the secondary structures of all four proteins were similar as evidenced by circular dichroism. However, their sensitivity towards denaturants differed. In all cases, the hydrogenated enzymes were more resistant than the fluorinated forms, exhibiting higher levels of activity and folding. The evolved pair of enzymes also exhibited increased stability as compared to the wild-type pair, resulting in similar thermostabilities for tfi-G4 and the wild-type enzyme (ile-WT). Based on these stability results, it appears

that the initial decrease in activity was in part due to the reduced stability. The pK_a of the catalytically essential histidine was not affected by the incorporation of 5TFI.

Fluorinated amino acids are complex with respect to their effects on protein folding and function. Unlike simpler coiled-coil motifs, CAT suffers marked reductions in activity and stability upon fluorination. However, these altered properties were successfully recovered using laboratory evolution. The technique of combining non-canonical amino acid incorporation with directed evolution promises to be a powerful tool for engineering functional protein-like macromolecules of novel composition.

Scheme 4.1 The enrichment process is designed to enrich for mutants that express functional CAT in the presence of 5TFI in order to minimize the number of non-functional mutants in the screening procedure.

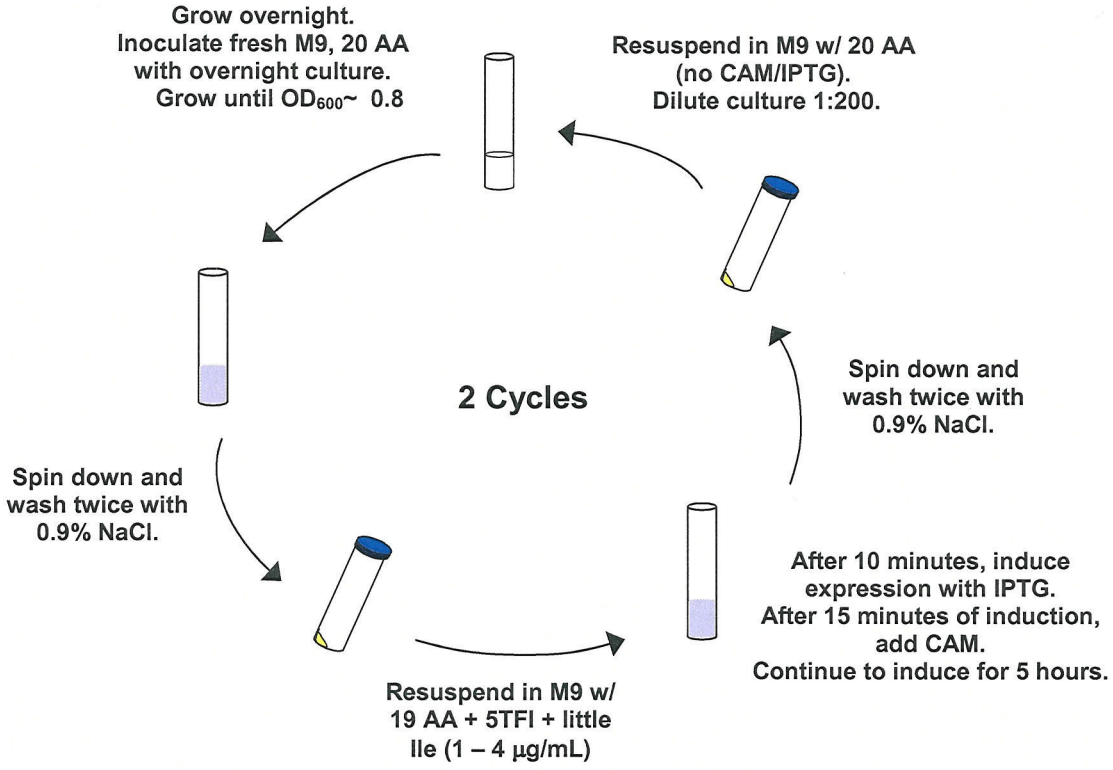


Figure 4.1 Evolutionary progression of catalytic activity of CAT mutants expressed in medium supplemented with either isoleucine (open squares) or 5TFI (closed squares). The dotted horizontal line indicates the activity level of ile-WT.

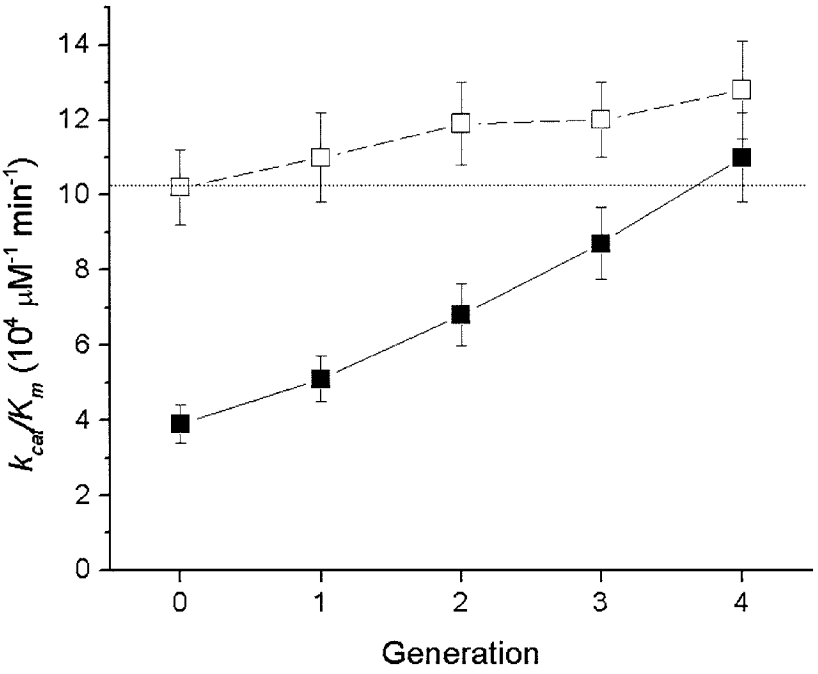
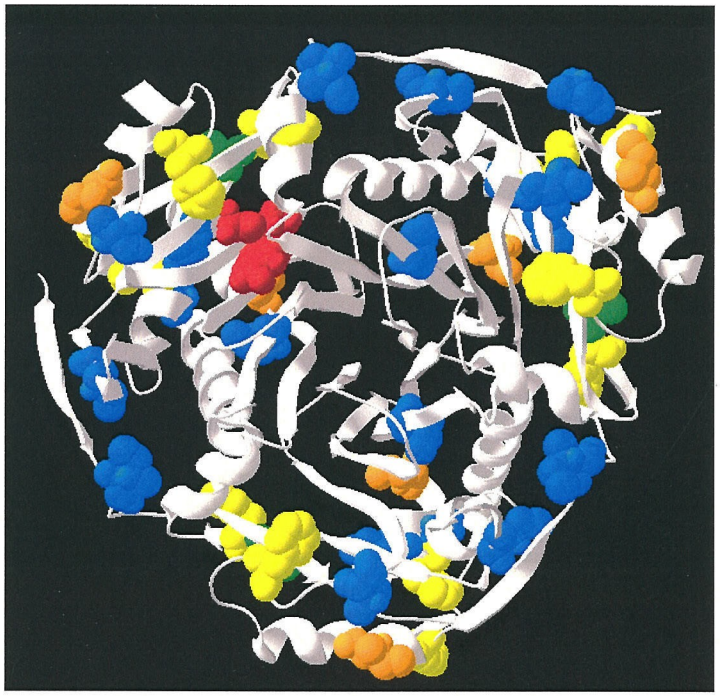


Table 4.1 Characterization of wild-type CAT and resulting mutants through four generations. The levels of 5TFI substitution in fluorinated CAT from generation 0 to 4 were determined by MALDI mass spectrometry and quantitative amino acid analysis. The yields are determined based on the purified protein amounts. All errors represent the standard deviations of at least three trials. The pK_a of the catalytically essential histidine was determined only for the wild-type and G4 variants.

Generation	Mutations	Yield (mg/L) Ile, 5TFI	% Sub. of 5TFI	pK_a Ile, 5TFI	k_{cat} (10^5 min^{-1}) Ile, 5TFI	k_{cat}/K_m (10^4 $\mu M^{-1}min^{-1}$) Ile, 5TFI
0	-	41 ± 18, 1.4 ± 1.0	85 ± 4	6.5 ± 0.1, 6.6 ± 0.1	17.4 ± 0.7, 6.6 ± 0.3	10.2 ± 1.0, 3.9 ± 0.5
1	H61Y	48 ± 22, 1.6 ± 1.3	86 ± 3	-	17.3 ± 0.6, 8.7 ± 0.3	11.0 ± 1.2, 5.1 ± 0.6
2	H61W, V94I, F134L	50 ± 25, 1.2 ± 1.1	84 ± 4	-	19.3 ± 0.8, 10.8 ± 0.5	11.9 ± 1.1, 6.8 ± 0.8
3	I41V, H61W, V94I, R114Q, F134L	47 ± 28, 2.0 ± 1.6	85 ± 5	-	20.2 ± 1.0, 14.1± 0.6	12.0 ± 1.0, 8.7 ± 1.0
4	I41V, H61W, V94I, F95L, R114Q, I119V, F134L	49 ± 26, 2.8 ± 1.9	83 ± 4	6.2 ± 0.1, 6.3 ± 0.1	22.0 ± 1.1, 16.9 ± 0.7	12.8 ± 1.3, 11.0 ± 1.2

Figure 4.2 Structural model of type I CAT trimer with one of three CAM (depicted in red) molecule placed in the binding pocket. Isoleucine/5TFI residues are highlighted in blue; mutations from valine to isoleucine/5TFI are indicated in green; orange balls represent isoleucine/5TFI to valine mutations; the remaining mutations are highlighted in yellow.

Front



Back

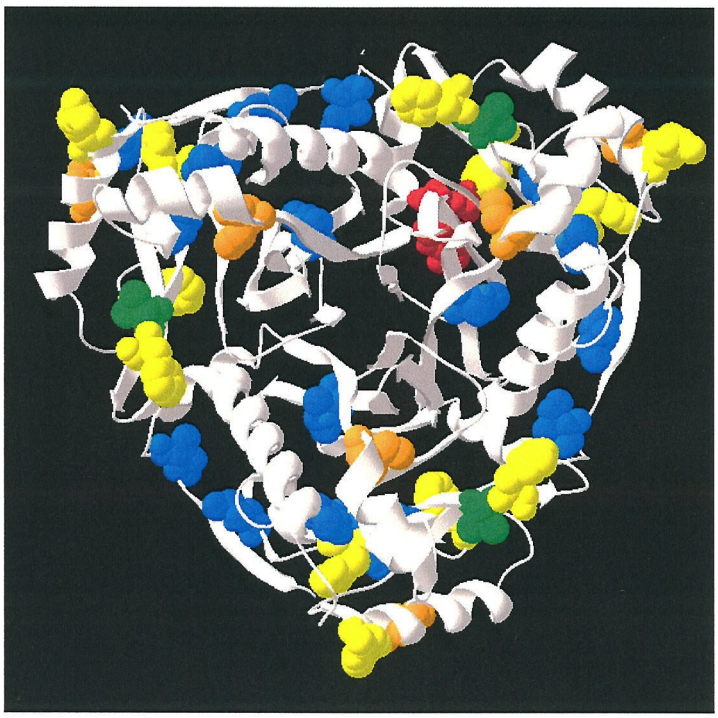


Figure 4.3 Secondary structures of ile-WT (open triangles), tfi-WT (closed triangles), ile-G4 (open circles), and tfi-G4 proteins (closed circles). All proteins are at 5 μM concentration, in 50 mM Tris, pH 7.8, 25 °C. The curves indicate the average of three scans.

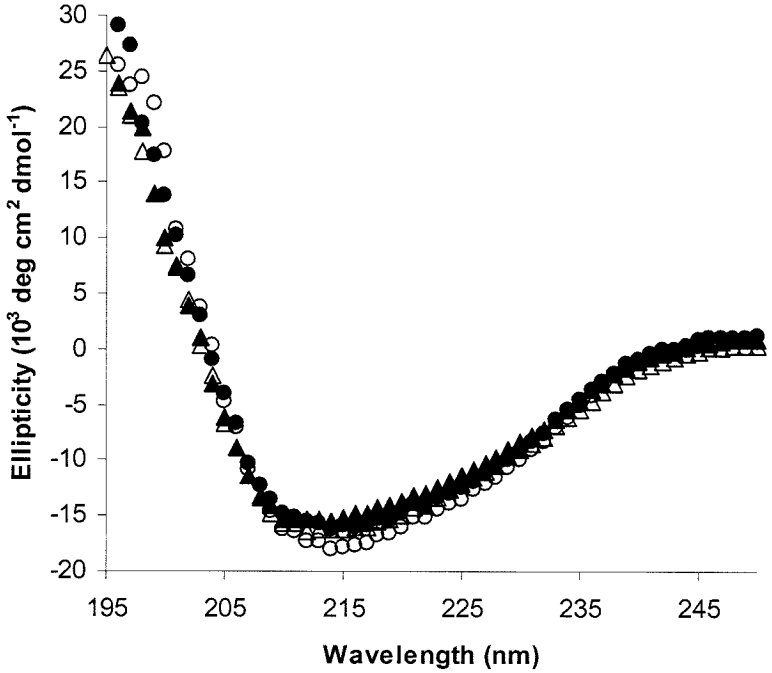


Figure 4.4 Thermostability of CAT variants: ile-WT (open triangles), tfi-WT (closed triangles), ile-G4 (open circles), and tfi-G4 proteins (closed circles). A) Residual activity versus temperature profiles. Residual activity is the ratio of the activity after incubation for 30 minutes at the designated temperature divided by the initial activity at 30 °C. Each point represents an average of three trials. B) Thermal denaturation monitored by circular dichroism spectroscopy (average of three scans, 1.5 °C interval, 1 minute equilibration time, 10 second averaging time).

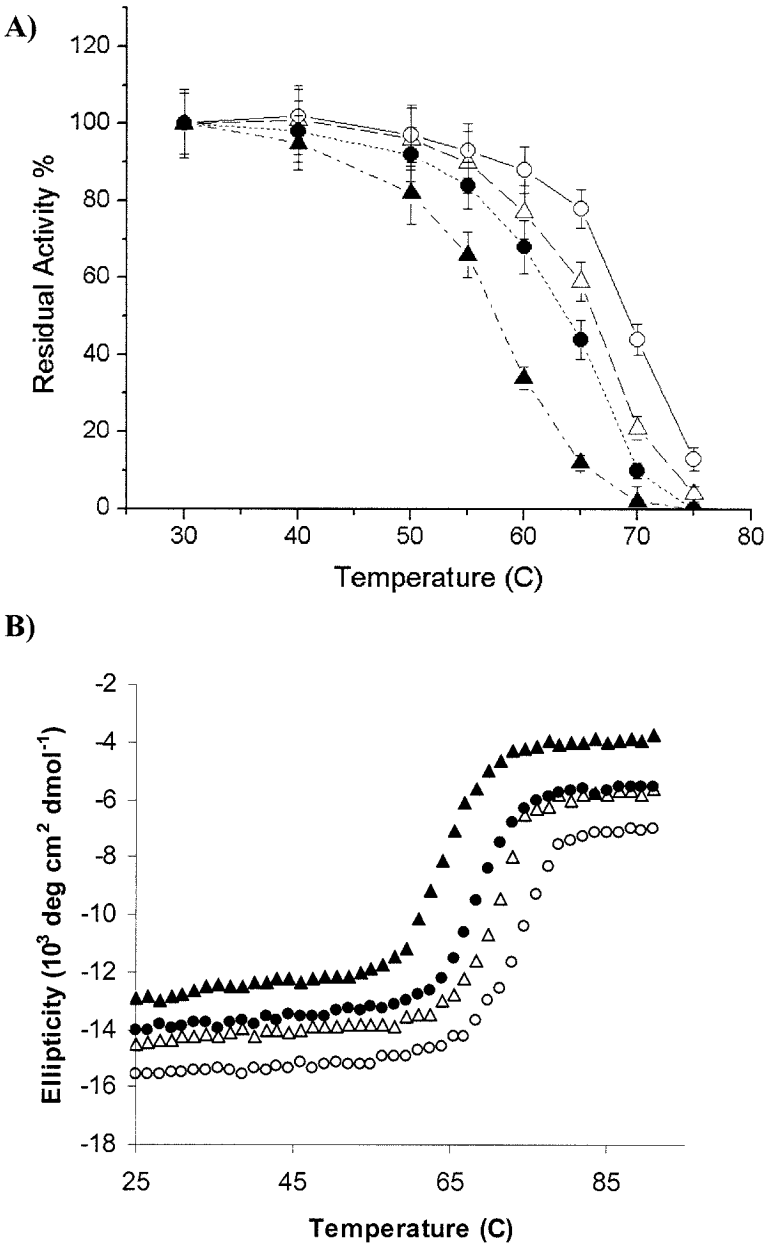


Figure 4.5 Stability of CAT variants against chemical denaturation by GuHCl: ile-WT (open triangles), tfi-WT (closed triangles), ile-G4 (open circles), and tfi-G4 proteins (closed circles). A) Residual activity versus GuHCl concentration profiles. Relative activity is the ratio of the activity at designated [GuHCl] divided by the activity at 0 M GuHCl. Each point represents an average of three trials. B) GuHCl titration monitored by circular dichroism spectroscopy (average of three scans, 1 minute equilibration time, 20 second averaging time).

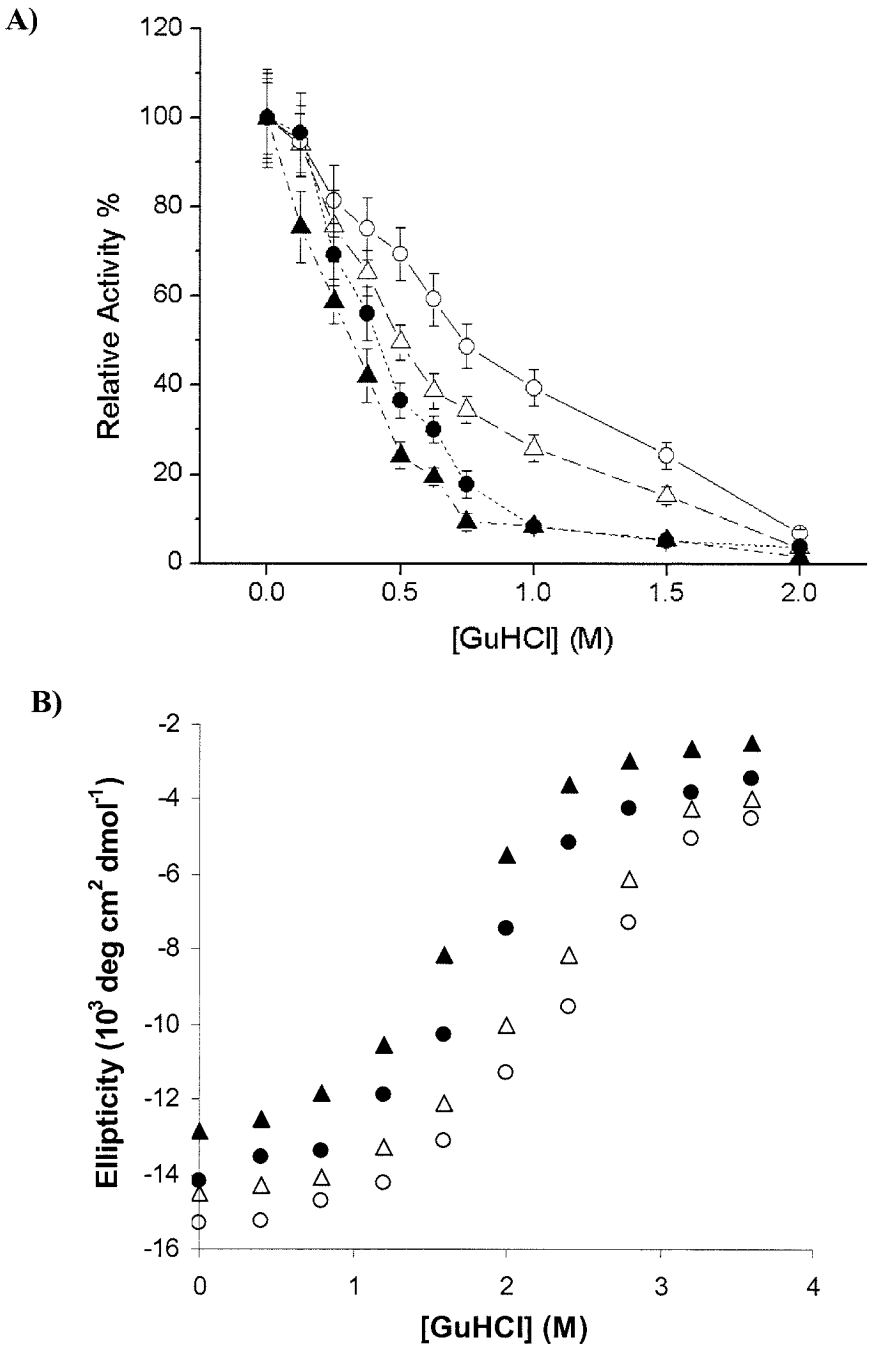
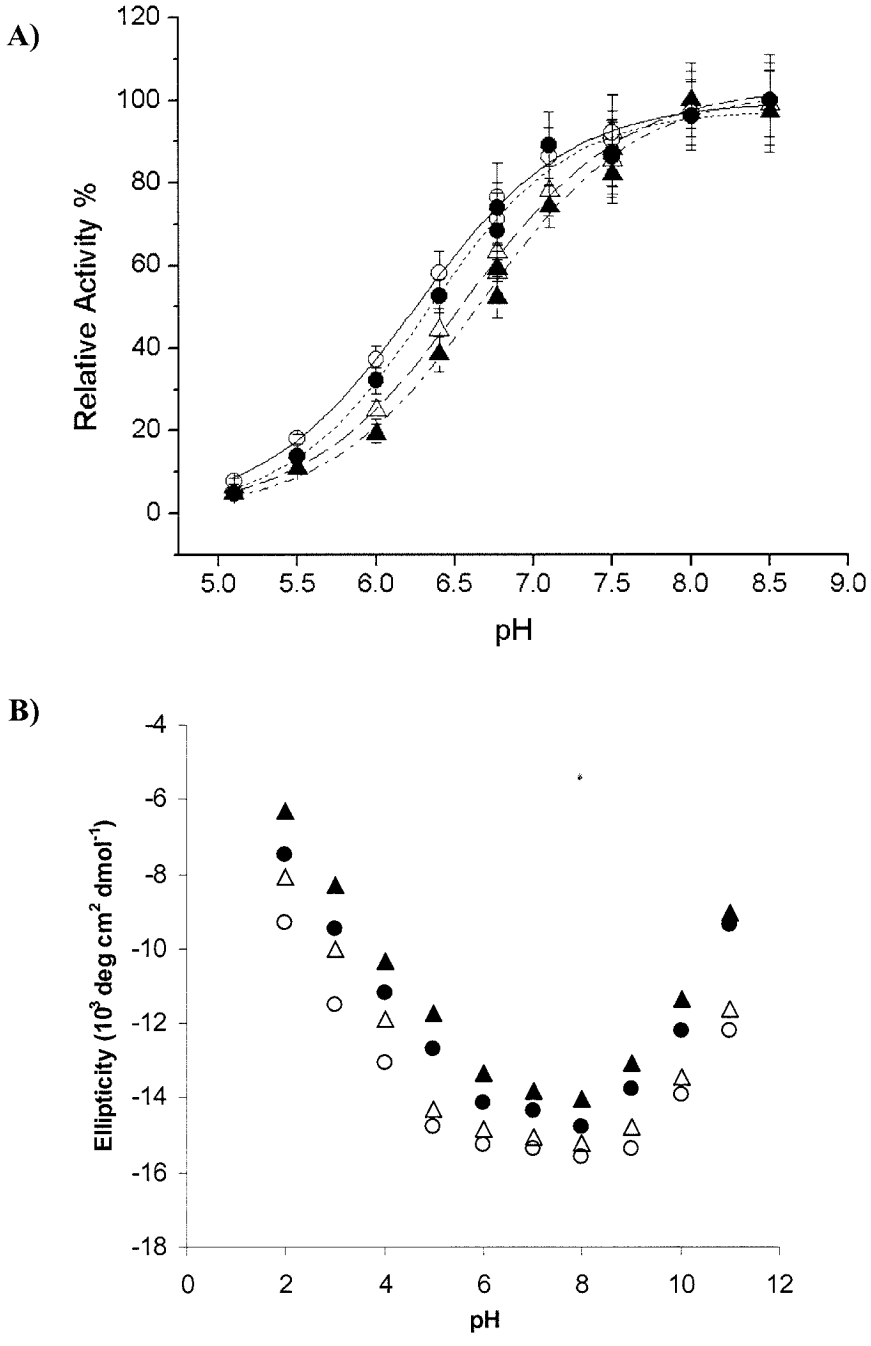


Figure 4.6 Response of CAT variants to changes in pH: ile-WT (open triangles), tfi-WT (closed triangles), ile-G4 (open circles), and tfi-G4 proteins (closed circles). A) Relative activity versus pH profiles. Relative activity is the ratio of the activity at designated pH divided by the highest observed activity. Each point represents an average of three trials. The lines describe best-fit theoretical curves[30]. B) pH effect on protein folding as monitored by circular dichroism spectroscopy (average of three scans, 1 minute equilibration time, 20 second averaging time).



4.5 References

1. Link, A. J.; Mock, M. L.; Tirrell, D. A., *Curr. Opin. Biotechnol.* **2003,** 14, (6), 603-609.

2. Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P., *Annu. Rev. Biochem.* **2004,** 73, 147-176.

3. Budisa, N., *Angew. Chem.-Int. Edit.* **2004,** 43, (47), 6426-6463.

4. Wang, L.; Schultz, P. G., *Angew. Chem.-Int. Edit.* **2005,** 44, (1), 34-66.

5. Cirino, P. C.; Tang, Y.; Takahashi, K.; Tirrell, D. A.; Arnold, F. H., *Biotechnol. Bioeng.* **2003,** 83, (6), 729-734.

6. Wang, P.; Tang, Y.; Tirrell, D. A., *J. Am. Chem. Soc.* **2003,** 125, (23), 6900-6906.

7. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A., *J. Am. Chem. Soc.* **2002,** 124, (20), 5652-5653.

8. Kirshenbaum, K.; Carrico, I.; Tirrell, D., *ChemBioChem* **2002,** 3, 235-237.

9. Tang, Y.; Tirrell, D. A., *J. Am. Chem. Soc.* **2001,** 123, (44), 11089-90.

10. Kiick, K. L.; Tirrell, D. A., *Tetrahedron* **2000,** 56, (48), 9487-9493.

11. Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A., *Biochemistry* **2001,** 40, (9), 2790-2796.

12. Budisa, N.; Huber, R.; Golbik, R.; Minks, C.; Weyher, E.; Moroder, L., *Eur. J. Biochem.* **1998,** 253, (1), 1-9.

13. Minks, C.; Huber, R.; Moroder, L.; Budisa, N., *Biochemistry* **1999,** 38, (33), 10649-10659.

14. Arnold, F. H., *Acc. Chem. Res.* **1998,** 31, (3), 125-131.

15. Tao, H. Y.; Cornish, V. W., *Curr. Opin. Chem. Biol.* **2002,** 6, (6), 858-864.

16. Stemmer, W. P. C., *Nature* **1994,** 370, (6488), 389-391.

17. Moore, J. C.; Arnold, F. H., *Nat. Biotechnol.* **1996,** 14, (4), 458-467.

18. Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. C., *Nat. Biotechnol.* **1996,** 14, (3), 315-319.

19. Miyazaki, K.; Arnold, F. H., *J. Mol. Evol.* **1999,** 49, (6), 716-720.

20. Arnold, F. H.; Volkov, A. A., *Curr. Opin. Chem. Biol.* **1999,** 3, (1), 54-59.

21. Joo, H.; Lin, Z. L.; Arnold, F. H., *Nature* **1999,** 399, (6737), 670-673.

22. Merz, A.; Yee, M. C.; Szadkowski, H.; Pappenberger, G.; Crameri, A.; Stemmer, W. P. C.; Yanofsky, C.; Kirschner, K., *Biochemistry* **2000,** 39, (5), 880-889.

23. Montclare, J. K.; Tirrell, D. A., *Angew. Chem.-Int. Edit.* **2006,** 45, (27), 4518-4521.

24. Shaw, W. V., *Sci. Prog.* **1992,** 76, (301-02), 565-580.

25. Suelter, C. H.; DeLuca, M., *Anal. Biochem.* **1983,** 135, 112-119.

26. Shaw, W. V.; Leslie, A. G. W., *Annu. Rev. Biophys. Biophys. Chem.* **1991,** 20, 363-386.

27. Ellis, J.; Bagshaw, C. R.; Shaw, W. V., *Biochemistry* **1991**, 30, (44), 10806-10813

28. Yunis, A. A., *Annu. Rev. Pharmacol. Toxicol.* **1988,** 28, 83-100.

29. Schwarz, S.; Cardoso, M., *Antimicrob. Agents Chemother.* **1991,** 35, (8), 1551-1556.

30. Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H., *Proc. Natl. Acad. Sci. U.S.A.* **2006,** 103, (15), 5869-5874.

31. Lewendon, A.; Shaw, W. V., *Biochem. J.* **1993,** 290, 15-19.

Chapter 5

The Effect of Fluorinated Amino Acids on the Rate of Proteolytic Degradation

The work with mutant GCN4 proteins were performed with Blair Benham-Pyle.

5.0 Abstract

The effect of fluorinated amino acids on the sensitivity of recombinant proteins toward proteolytic degradation was explored. Two types of proteins, coiled-coil proteins based on GCN4 and mutants of chloramphenicol acetyltransferase (CAT), were subjected to proteolysis by elastase and trypsin. 5,5,5-trifloroisoleucine (5TFI) and 4,4,4-trifluorovaline (4TFV) were incorporated into the 4 **a**-positions of INL and VNL, yielding tfi-INL and tfv-VNL, respectively. The progress of proteolysis was monitored by Tris-Tricine gel electrophoresis and densitometry analysis. The half life of tfi-INL in the presence of elastase was 4-fold longer than that of its hydrogenated counterpart. Less significant effects were seen with tfv-VNL as well as upon exposure to trypsin. On the contrary, upon incorporation of 5TFI and 5TFL into CAT variants G1 and L2A1, respectively, the proteins became more sensitive to proteolysis by both elastase and trypsin.

5.1 Introduction and Background

For many applications proteins must remain active for an extended period of time and under a wide range of conditions.[1] Especially when the protein is used for *in vivo* purposes, such as in orally administered therapeutics, the ability of the protein to resist cleavage by proteases is crucial.[2] Most current techniques for stabilization of proteins and peptides aim to make them somewhat "unnatural" to avoid recognition by naturally occurring enzymes. Several chemical modification strategies are possible, including glycosylation, PEGylation, alkylation of amide nitrogen, and peptide cyclization.[3] In addition, α-amino acids have been replaced by or mixed in with β-amino acids, resulting in compounds with increased stability toward proteolytic degradation.[4, 5] The introduction of D-amino acids has also shown promise.[6, 7]

Despite their success, these techniques still suffer from some limitations. One of the most crucial challenges to overcome is the loss of biological activity. Because the protein sequences have been evolved using the naturally occurring 20 amino acids, such dramatic changes to the amino acids or further modifications to the proteins often result in altered activity. Also, most of these techniques require either the chemical synthesis of the polypeptide or chemical modifications following translation. Chemical synthesis of peptides limits the size of the protein to less than 50 amino acids while yielding low amounts. Also, further modifications to proteins could be time and cost intensive.

In vivo incorporation of non-canonical amino acids has the potential to address these limitations and serve as an effective tool for rendering proteins more resistant to proteolysis. Since non-canonical amino acids contain atoms and functional groups that

are not found in the 20 canonical amino acids, the recognition and the catalysis by proteases may be reduced. Furthermore, because the *E. coli* cellular machinery could be utilized for protein expression *in vivo*, large amounts of proteins could be accessed with relative ease. Finally, because the non-canonical amino acids that are incorporated *in vivo* still maintain the same stereochemistry and configuration at the backbone, biological activity could be more easily preserved.[8-10] Based on these hypotheses, some preliminary studies have been conducted in our laboratory and shown promising results. Specifically, when 5,5,5-trifluoroisoleucine was incorporated into artificial extracellular matrix proteins in place of isoleucine, its susceptibility to proteolysis by elastase was greatly reduced.[11]

In light of these studies, we were interested in investigating the effects of fluorination on the susceptibility of two types of proteins to proteolysis, mutants based on the *bzip* protein GCN4 and mutants of the enzyme chloramphenicol acetyltransferase. The introduction of trifluoromethyl groups into the side chains of valines, isoleucines, or leucines could potentially alter the proteolysis rate. We explored this hypothesis by monitoring the status of *bzip* and CAT variants at various time intervals in the presence of either elastase or trypsin. The trends observed for *bzip* and CAT variants were different. Fluorinated *bzip* mutants, especially those containing 5,5,5-trifluoroisoleucine (5TFI) , exhibited an increase in stability with respect to proteolysis by elastase, while susceptibility toward trypsin was unaltered. In contrast, fluorinated CAT variants containing either 5TFI or 5,5,5-trifluoroleucine, were more sensitive to proteolysis than their hydrogenated counterparts. These results could be understood in the context of the two main factors determining the rate of proteolysis of folded proteins: the unfolding rate,

either globally or locally, and the catalytic rate of hydrolysis by the proteases. Both of these factors are likely to be affected by the incorporation of fluorinated amino acids and therefore, depending on the magnitude of change in each of these factors, one observes different net changes in the rate of proteolysis.

5.2 Materials and Methods

5.2.1 Protein Expression and Purification

bzip and CAT proteins were expressed and purified according to the protocols outlined in Chapters 2 and 4, respectively. The purified protein concentrations were determined by the Bradford protein assay (BioRad) with bovine serum albumin (Sigma) as the standard.

5.2.2 Proteolysis of bzip Proteins

Four *bzip* proteins (INL, tfi-INL, VNL, and tfv-VNL) were analyzed for digestion with elastase (Sigma, from porcine pancreas) and trypsin (Promega, sequencing grade modified, porcine). The reported K_m values for the susbstrates of porcine elastase and trypsin are 1.15 mM and 2.76 mM.[12, 13] Protein samples (400 μg/mL) in phosphate buffered saline, pH 7.4, were incubated at 25 °C for 30 minutes. Proteases were added to the protein solutions according to the concentrations listed in Table 5.1. Aliquots were removed at various time intervals and phenylmethylsulfonyl fluoride (PMSF, Sigma, 500 μg/mL) was immediately added to inhibit further proteolysis. Samples were analyzed on a 15% Tris-Tricine gel, visualized using Coomassie Blue stain, and the progress of

proteolysis was quantified through densitometry using ImageJ (National Institutes of Health freeware image analysis program). Control curves indicated a linear relationship between protein concentration and stain intensity for all proteins within the relevant protein concentration range. To calculate the half lives, $t_{1/2}$, log of the remaining protein concentrations was plotted against time and linear fitting was performed.

5.2.3 Proteolysis of CAT Variants

Four CAT variants (ile-G2, tfi-G2, leu-L2A1, tfl-L2A1) were analyzed for proteolytic degradation using elastase (Sigma, from porcine pancreas), and trypsin (Promega, sequencing grade modified, porcine). The digestion buffer (50 mM Tris, 100 μg/mL BSA, pH 7.8) was incubated at 37 °C for 30 minutes. BSA was added to the digestion buffer to prevent protein adsorption to the tube wall during the incubation period, which has been shown to be significant with dilute protein solutions. CAT variants were added to the digestion buffer to a final concentration of 0.02 μM (0.55 μg/mL) and incubated for an additional 10 minutes at 37 °C. Proteases were added to the protein solutions to a final concentration of 5 and 2 μg/mL for elastase and trypsin, respectively. Controls were also performed where no protease was added. Samples were taken at selected time intervals and PMSF (Sigma, 100 μg/mL) was immediately added to prevent further proteolysis. The remaining CAT activity was assayed at room temperature according to procedures outlined in Chapter 4. The initial rates were measured by monitoring the absorbance change at 412 nm as a function of time using a 96-well spectrophotometer (SPECTRAmax, Molecular Devices).

5.3 Results

5.3.1 Proteins of Interest: bzip and CAT Variants

Two types of proteins were of interest in this study. The first is a leucine zipper protein based on the yeast transcription factor GCN4, which is composed of a disordered DNA-binding domain and a coiled-coil dimerization domain. Four **a**-position residues in the hydrophobic core of the coiled-coil domain, which are isoleucines in INL and valines in VNL, have been replaced with 5,5,5-trifluoroisoleucine and 4,4,4-trifluorovaline, respectively, yielding tfi-INL and tfv-VNL. There are no other sites of substitution within the DNA-binding domain. The sequence, as well as the properties, of these proteins is described in Chapter 2. The other type of protein is the type I chloramphenicol acetyltransferase, a trimeric enzyme containing 9 isoleucine and 11 leucine residues dispersed throughout each monomeric unit. Four variants of CAT were explored: leu-L2A1, tfl-L2A1, ile-G1, and tfi-G1. L2A1 is a three amino acid mutant of CAT which was evolved for higher thermostability in the presence of 5,5,5-trifluoroleucine.[14] As described in Chapter 4, G1 is a single amino acid mutant of CAT, the first generation mutant in the evolution of CAT evolved for higher activity in the presence of 5,5,5-trifluoroisoleucine. The progress of digestion of *bzip* proteins was monitored by following the amount of intact protein remaining via gel electrophoresis. CAT variants were monitored by quantifying the remaining catalytic activity.

5.3.2 Proteolysis of bzip Proteins by Elastase and Trypsin

To examine the time course of degradation of *bzip* proteins by proteases, gel electrophoresis was performed and protein bands were visualized using Coomassie Blue stain. Densitometry analysis yielded the fraction of intact protein. Figure 5.1 displays the progress of proteolysis by elastase of the INL/tfi-INL as well as the VNL/tfv-VNL pairs of proteins. Both of the fluorinated proteins exhibit higher resistance toward cleavage by elastase. The difference is greater when 5TFI replaced isoleucine residues in INL. The time required to cleave half of the protein, $t_{1/2}$, was 4 times longer for tfi-INL as compared to INL, while the difference was only 1.7-fold between tfv-VNL and VNL. Similarly, the extent of proteolysis by trypsin was monitored. Although some increased resistance to trypsin was detected in the fluorinated proteins, the extent was much smaller than that observed with elastase (Figure 5.2). The $t_{1/2}$ of tfi-INL was 1.6 times longer than that of its hydrogenated counterpart. No significant difference was seen between VNL and tfv-VNL.

5.3.3 Proteolysis of CAT Variants by Elastase and Trypsin

Tfl-L2A1 and tfi-G1, as well as their hydrogenated counterparts, leu-L2A1 and ile-G1, were subjected to proteolysis by elastase and trypsin. The extent of cleavage was monitored over a period of 180 minutes during which time the samples without any protease showed a drop in activity of less than 20%. As compared to their hydrogenated counterparts, both fluorinated mutants undergo higher levels of cleavage by elastase as evidenced by a sharper drop in activity (Figure 5.3A). All the data points have been normalized to remove background losses in activity. A similar trend was observed when the proteins were subjected to trypsin (Fig. 5.3B). However, the differences between the

hydrogenated and fluorinated forms of CAT were smaller in response to trypsin. The hydrogenated CAT variants exhibited similar sensitivity to proteolysis by both elastase and trypsin.

5.4 Discussion

Not only is amino acid sequence important, but the structural stability of the protein becomes an important factor in determining the proteolysis rate. It is known that upon denaturation, proteins become more susceptible to proteolysis *in vitro*.[15, 16] In fact, it is believed that effective proteolytic digestion of a protein requires unfolding of the substrate, either globally or locally.[17, 18] Global unfolding of protein depends on the thermodynamic stability of the folded structure, and therefore, conformational stability and proteolytic sensitivity have been inversely correlated.[19-21] Similarly *in vivo*, thermodynamic stability has been identified as a major determinant of the rate of intracellular degradation.[22] However, global unfolding of proteins isn't necessary for proteolysis. Proteolysis can also occur under native conditions through local fluctuation, rather than global unfolding, with minimal change in solvent accessible surface area.[18, 23] Based on this mechanism, it has been shown that increasing the flexibility of a protein enhances its susceptibility to proteases *in vitro* and *in vivo*.[23-26]

The introduction of fluorinated amino acids into proteins can affect both of the steps required for proteolysis. First, thermodynamic stability can be shifted in either direction. In the case of *bzip* proteins, the introduction of 5TFI and 4TFV increased stability, although to different extents. On the other hand, the introduction of either 5TFI

or 5TFL into CAT decreased the overall stability, as indicated by thermostability measurements. Less quantifiable is the effect on local fluctuations. However, it is likely that, in addition to the global effect, the introduction of fluorinated amino acids can impact the local fluctuations and instabilities. The enzymatic reaction can also be affected if the recognition motif and/or the substrate involves the amino acid that is being substituted by its non-canonical analogs. In the case of elastase, which cuts after small hydrophobic residues including valine, isoleucine, and leucine, replacement of these residues by their fluorinated analogues has the potential to affect the binding to the protease as well as its catalysis. These two effects combine together to give rise to the observed results.

Among the set of experiments with *bzip* proteins, the largest difference in the proteolysis rate between the fluorinated and hydrogenated forms of proteins was found in the INL/tfi-INL pair when exposed to elastase. $t_{1/2}$, the time required to cleave half of the protein, was 4 times longer for tfi-INL than INL. Perhaps this is a result of both factors, stability and catalysis, contributing towards the observed resistance towards elastase. As shown in Chapter 2, the thermal melting temperature (T_m) of tfi-INL is 27 °C higher than INL; the increased stability hinders access to the protein by the protease. In addition, 4 **a** position residues in the leucine zipper have been replaced with 5TFI. Since isoleucines and leucines are the largest substrates that elastase recognizes, perhaps the additional bulk provide by the trifluoromethyl group decreases the efficiency of the catalysis. These two effects combined to give rise to the observed result. In the case of tfv-VNL, the increase T_m is 4 °C. Furthermore, since elastase efficiently recognizes isoleucine and leucine, the presence of bulkier trifluromethyl group on valine may not hinder its fitting

into the larger binding pocket. Therefore, the proteolysis rate is only decreased by 1.7-fold in the case of tfv-VNL.

Trypsin is a serine protease that cuts after arginine and lysine. Therefore, the incorporation of 5TFI and 4TFV in place of isoleucine and valine does not directly affect the chemistry of the proteolysis. The observed changes in the proteolysis rate by trypsin would be mainly attributed to the stability effect. As hypothesized, the changes are much smaller as compared to proteolysis by elastase. The largest changes were again observed between tfi-INL and INL, but this difference was only 1.6-fold in $t_{1/2}$. The difference between VNL and tfv-VNL was insignificant. In light of these results, the 4-fold difference in $t_{1/2}$ between tfi-INL and INL in the presence of elastase could not be solely explained by stability effect; substrate recognition and catalysis must be disturbed somewhat by 5TFI.

The results from the CAT experiments were quiet different. In response to both elastase and trypsin, fluorinated CAT variants showed higher susceptibility to degradation. The difference was greater for degradation by elastase, especially for tfi-G1. This increased sensitivity to proteases may be attributed to the lower stability of fluorinated CAT variants. T_{50}, the temperature at which half of the activity is lost after 30 minutes, was 6 °C lower for tfi-G1 as compared to ile-G1 (unpublished data). However, global stability alone may not fully explain the observed results since tfl-L2A1 exhibits comparable thermostability to leu-L2A1 (ΔT_{50} = 0.9 °C).[14] Local instabilities and fluctuations, especially near the site of fluorinated amino acid incorporation, could lead to increased solvent accessible area. This could further contribute to a faster rate of proteolysis for the fluorinated CAT variants. The fact that a greater difference was seen

in response to elastase as compared to trypsin also supports this hypothesis. Since elastase cuts after small hydrophobic residues, the presence of fluorinated amino acids near these hydrophobic areas could disturb the local packing, allowing faster proteolysis. As for *bzip* proteins, because of their simple structure, local fluctuations may be less significant.

Previous studies in our laboratory showing that incorporation of non-canonical amino acids reduces the rate of proteolysis are intrinsically different from the current study. The artificial extracellular matrix proteins used in previous studies are denatured and/or unstructured. Therefore, the only factors influencing the rate are substrate recognition and catalysis. Furthermore, artificial extracellular matrix proteins are cleaved by human leukocyte elastase primarily at isoleucine sites. Incorporation of 5TFI into those isoleucine sites suppresses the rate of proteolysis.

5.5 Conclusion

Incorporation of non-canonical amino acids affects both steps involved in proteolysis- unfolding as well as catalysis. The unfolding step is greatly influenced by the stability of the protein- both by global thermodynamic stability and by local fluctuations. Incorporation of fluorinated amino acids can increase stability as observed in *bzip* proteins or decrease it found for CAT. Furthermore, catalysis by the protease can be affected directly by the fluorinated amino acid, since the side chains are different from those of any natural substrate. The additional bulk of the trifluoromethyl group may further affect catalysis. Therefore, in some cases, incorporation of fluorinated amino

acids increases susceptibility to proteolysis, while in others, fluorination allows proteins to be more resistant to degradation by proteases. Balancing these factors in designing proteins can allow us to take better advantage of non-canonical amino acids in protein engineering.

Table 5.1 Protease concentrations used in the digestion of *bzip* proteins.

Protein	Protease	[Protease] (μg/mL)
INL and tfi-INL	Elastase	10
VNL and tfv-VNL	Elastase	2.5
INL and tfi-INL	Trypsin	1.6
VNL and tfv-VNL	Trypsin	0.4

Figure 5.1 Elastase degradation of *bzip* proteins as monitored by Tris-Tricine gel electrophoresis and quantified using densitometry analysis. A) INL (open triangle) and tfi-INL (closed triangle) and B) VNL (open circle) and tfv-VNL (closed triangle).

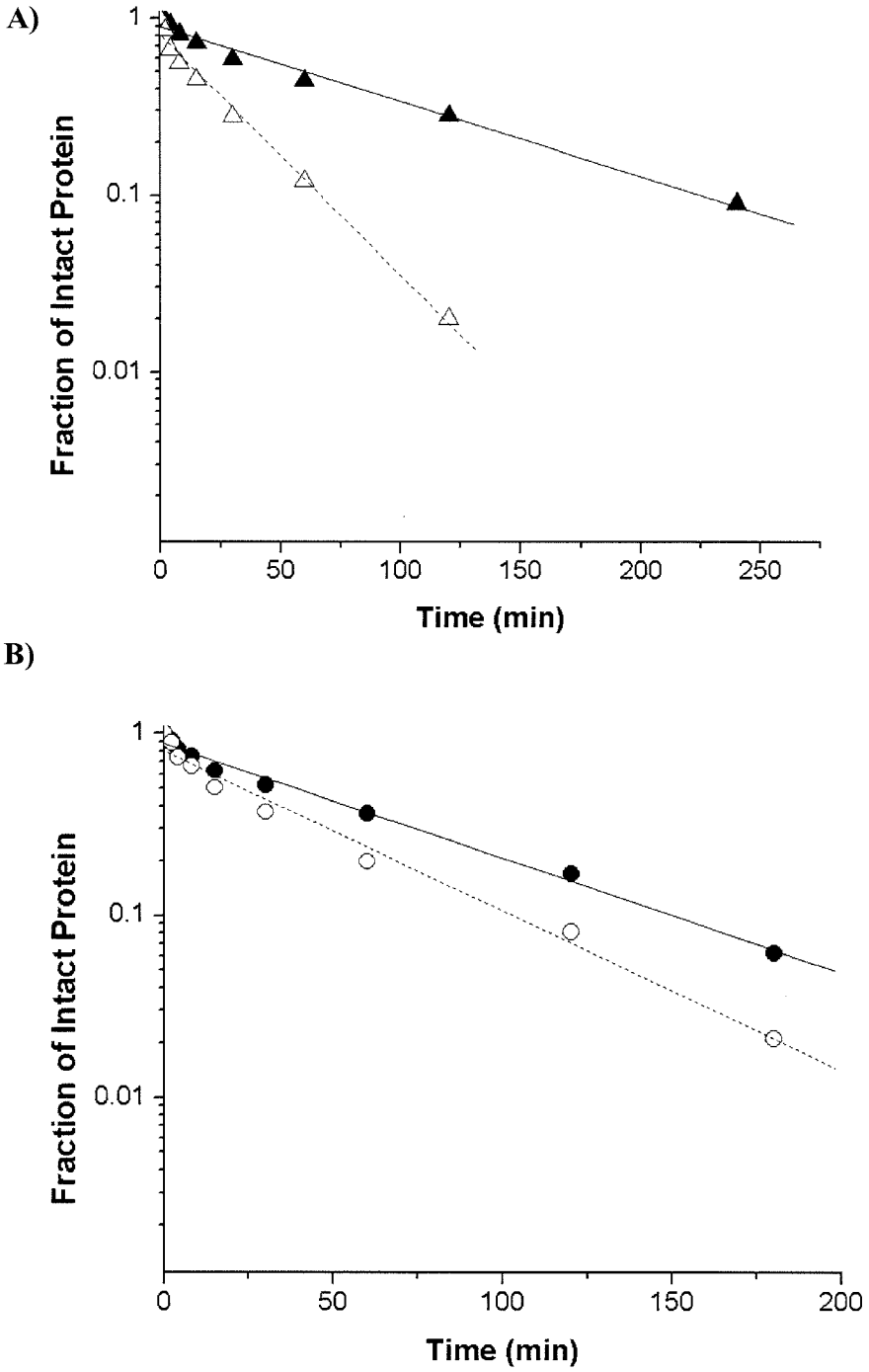


Figure 5.2 Trypsin degradation of *bzip* proteins as monitored by Tris-Tricine gel electrophoresis and quantified using densitometry analysis. A) INL (open triangle) and tfi-INL (closed triangle) and B) VNL (open circle) and tfv-VNL (closed triangle).

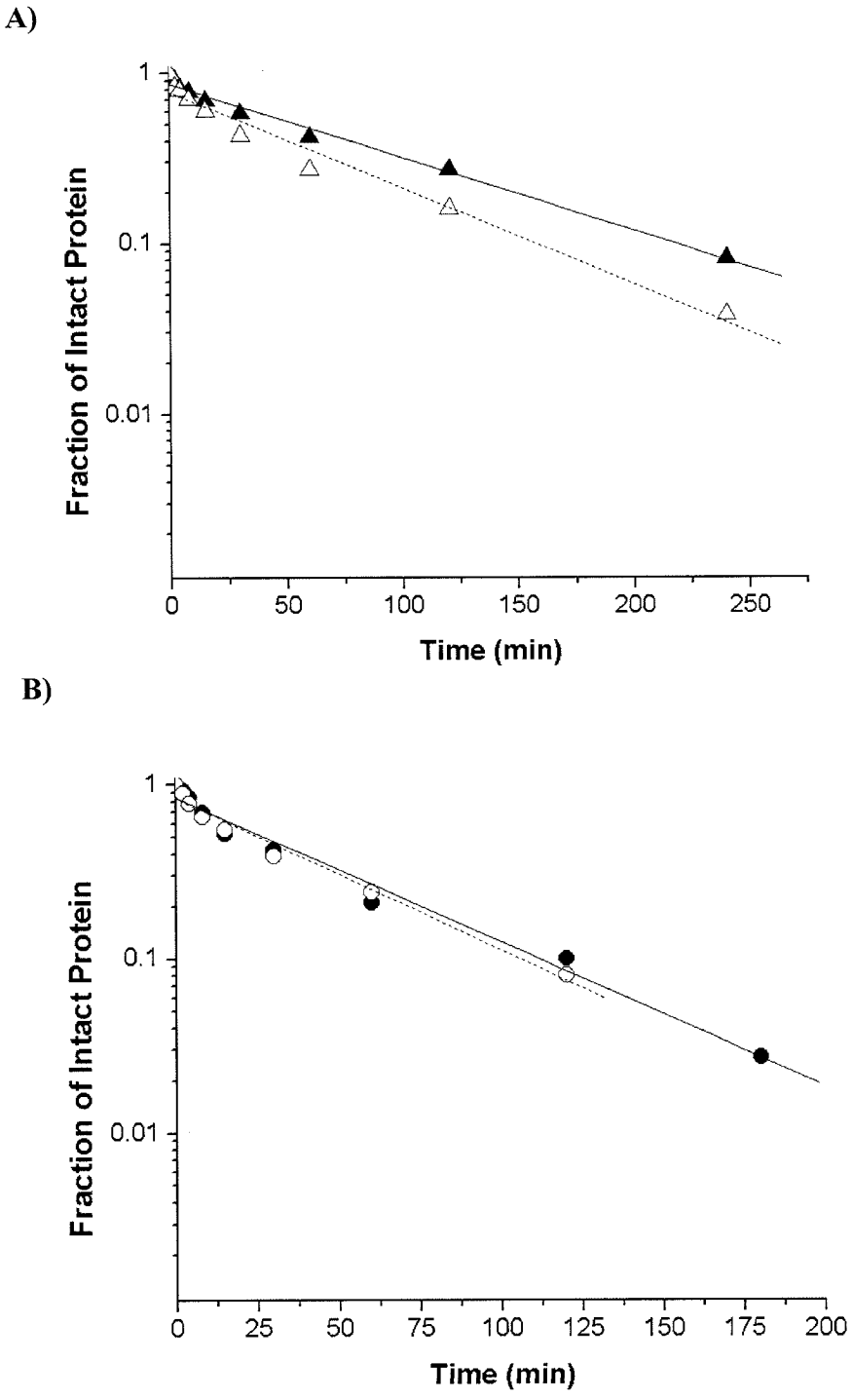
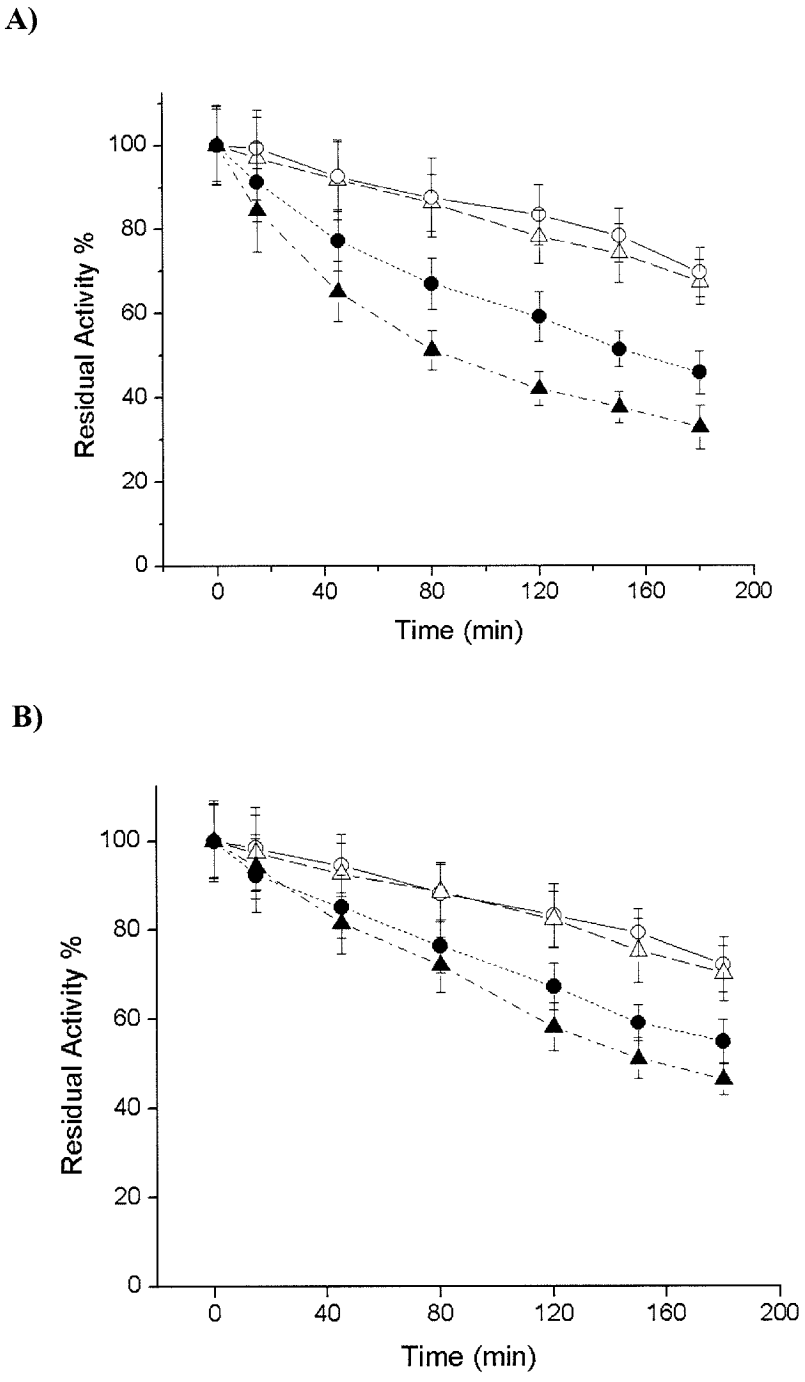


Figure 5.3 Proteolysis of CAT variants determined by residual activity: ile-G1 (open triangle), tfi-G1 (closed triangle), leu-L2A1 (open circle), and tfl-L2A1 (closed circle). A) Elastase degradation. B) Trypsin degradation.



5.6 References

1. Inouye, M. a. S., Raghupathy, *Protein engineering: applications in science, medicine, and industry*. Academic Press: Orlando, **1986**.

2. Reid, R. E., *Peptide and Protein Drug Analysis*. Marcel Dekker, Inc.: New York, **2000**.

3. Adessi, C.; Soto, C., *Curr. Med. Chem.* **2002**, 9, (9), 963-978.

4. Hintermann, T.; Seebach, D., *Chimia* **1997**, 51, (5), 244-247.

5. Hook, D. F.; Bindschadler, P.; Mahajan, Y. R.; Sebesta, R.; Kast, P.; Seebach, D., *Chem. Biodiversity* **2005**, 2, (5), 591-632.

6. Chorev, M.; Goodman, M., *Trends Biotechnol.* **1995**, 13, (10), 438-445.

7. Schumacher, T. N. M.; Mayr, L. M.; Minor, D. L.; Milhollen, M. A.; Burgess, M. W.; Kim, P. S., *Science* **1996**, 271, (5257), 1854-1857.

8. Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A., *Biochemistry* **2001**, 40, (9), 2790-2796.

9. Wang, P.; Tang, Y.; Tirrell, D. A., *J. Am. Chem. Soc.* **2003**, 125, (23), 6900-6906.

10. Son, S., Tanrikulu, I. C., and Tirrell, D. A., *ChemBioChem* **2006**, 7, (8), 1251-1257.

11. Mock, M. L. Protein modification through in vivo incorporation of noncanonical amino acids. California Institute of Technology, Pasadena, **2006**.

12. Lorand, L.; Rule, N. G., *Nature* **1961**, 190, (477),722-723.

13. Bieth, J.; Spiess, B.; Wermuth, C. G., *Biochem. Med.* **1974**, 11, (4), 350-357.

14. Montclare, J. K.; Tirrell, D. A., *Angew. Chem.-Int. Edit.* **2006**, 45, (27), 4518-4521.

15. Hagihara, B.; Nakayama, T.; Matsubara, H.; Okunuki, K., *J. Biochem.* **1956**, 43, (4), 469-481.

16. Harrington, W. F.; Vonhippel, P. H.; Mihalyi, E., *Biochim. Biophys. Acta* **1959**, 32, (1), 303-304.

17. Wilkinson, J. M., *Fragmentation of Polypeptides by Enzymic Methods*. John Wiley & Sons: New York, **1986**.

18. Park, C.; Marqusee, S., *J. Mol. Biol.* **2004**, 343, (5), 1467-1476.

19. Sauer, R. T., *Fold. Des.* **1996**, 1, (2), R27-R30.

20. Kolmar, H.; Frisch, C.; Gotze, K.; Fritz, H. J., *J. Mol. Biol.* **1995**, 251, (4), 471-476.

21. Fontana, A.; deLaureto, P. P.; DeFilippis, V.; Scaramella, E.; Zambonin, M., *Fold. Des.* **1997**, 2, (2), R17-R26.

22. Parsell, D. A.; Sauer, R. T., *J. Biol. Chem.* **1989**, 264, (13), 7590-7595.

23. Rote, K. V.; Rechsteiner, M., *J. Biol. Chem.* **1986**, 261, (33), 5430-5436.

24. Perry, S. T.; Lee, K. L.; Kenney, F. T., *Arch. Biochem. Biophys.* **1979**, 195, (2), 362-367.

25. Rosa, J. J.; Richards, F. M., *J. Mol. Biol.* **1981**, 145, (4), 835-851.

26. Allende, J. E.; Richards, F. M., *Biochemistry* **1962**, 1, (2), 295-304.

Chapter 6

Conclusions and Future Directions

Protein engineering, since first demonstrated more than 25 years ago, has contributed tremendously to our understanding of protein structure and function.[1-3] In addition to the originally developed technique of site-directed mutagenesis, numerous other techniques have emerged, providing us with a wide variety of powerful tools to not only to study proteins, but also to engineer proteins of industrial and medical importance.[3-7] Among these techniques, the ability to incorporate non-canonical amino acids has allowed us to introduce novel chemical and physical properties into proteins. In this thesis, the *in vivo* incorporation of fluorinated amino acids has been explored. The results found through these studies have provided further insights into how we can better control protein structure, stability, and function through fluorination.

In the right context, the incorporation of fluorinated aliphatic residues has the power to enhance protein stability against thermal and chemical denaturation as well as proteolytic degradation. To further this effect, multiple fluorinated amino acids could be incorporated simultaneously, for example, 5TFI into the **a** and 5TFL into the **d** positions of coiled-coil proteins. Since both 5TFI and 5TFL alone gave rise to significant increases in thermal denaturation temperature, the combined effect could be dramatic. Furthermore, the stability against proteases, especially elastase, could be further reduced. Since elastase cleaves after small hydrophobic residues, fluorinating, if possible, all such residues has the potential to drastically reduce proteolysis. Incorporating multiple fluorinated amino acids can also create fluorine-fluorine interactions that can further stabilize protein folds. Even in cases where fluorination of one amino acid decreases the stability of a protein, incorporating other fluorinated amino acids could result in favorable interactions that can reverse the effect on stability.

The most important conclusion of this work may be that there are no all-encompassing rules concerning fluorinated amino acids. The replacement of a methyl group with a trifluoromethyl group, even in the same position within a protein, can produce different effects depending on which amino acid is fluorinated. In addition, the stereochemistry of the fluorinated amino acid may or may not be important, again depending on the amino acid, in determining the efficiency of the incorporation as well as the effect on protein structure and stability. Furthermore, while the incorporation of a fluorinated amino acid stabilizes one protein while maintaining function, its incorporation into another could render it less stable and less active.

Clearly, the effect of fluorinated amino acid incorporation on proteins is complex and, presently, we lack the ability to accurately predict changes in protein behavior upon fluorination. This is expected given our limited understanding of the molecular origins of protein structure and function. The results described here show that directed evolution can be applied to the introduction of fluorinated amino acids. Therefore, in instances when a property of the protein is adversely affected upon non-canonical amino acid incorporation, laboratory evolution can be utilized to recover this loss while maintaining the advantages that the non-canonical amino acid provides. Through studies explored in this thesis and those to come in the future, we will continually learn more about the impact of fluorinated amino acids upon incorporation into proteins.

References

1. Winter, G.; Fersht, A. R.; Wilkinson, A. J.; Zoller, M.; Smith, M., *Nature* **1982**, 299, (5885), 756-758.

2. Ulmer, K. M., *Science* **1983**, 219, (4585), 666-671.

3. Brannigan, J. A.; Wilkinson, A. J., *Nat. Rev. Mol. Cell Biol.* **2002**, 3, (12), 964-970.

4. Bornscheuer, U. T.; Pohl, M., *Curr. Opin. Chem. Biol.* **2001**, 5, (2), 137-143.

5. Link, A. J.; Mock, M. L.; Tirrell, D. A., *Curr. Opin. Biotechnol.* **2003**, 14, (6), 603-609.

6. Collins, C. H.; Yokobayashi, Y.; Umeno, D.; Arnold, F. H., *Curr. Opin. Biotechnol.* **2003**, 14, (4), 371-378.

7. Gustafsson, C.; Govindarajan, S.; Minshull, J., *Curr. Opin. Biotechnol.* **2003**, 14, (4), 366-370.