PROTEIN ENGINEERING USING UNNATURAL AMINO ACIDS

Incorporation of Leucine Analogs into Recombinant Protein in Vivo

Thesis by

Yi Tang

In Partial Fulfillment of the Requirement

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2002

(Defended May 23rd, 2002)

To My Grandparents

ACKNOWLEDGEMENT

My life at Caltech has been truly wonderful. The last five years have helped me mature both as a person and as a scientist (or an engineer). Many people along the way have made the journey enjoyable and memorable. First and foremost, I thank my thesis advisor David Tirrell. Dave has been extremely supportive of my career in every aspect. He gave me freedom to pursue exciting projects and collaborations. His mentorship in this journey has given me confidence in my abilities and outlook. I cannot imagine a more rewarding research experience than what Dave has guided me through. I am forever indebted to his mentorship. I am looking forward to work with him further on a different level in the near future.

I thank Prof. Francis Arnold, Prof. Dennis Dougherty and Prof. David Chan for serving on my thesis committee and providing valuable advice. I am grateful for the encouragements provided by Prof. Arnold, Prof. Mark Davis, and Prof. Julia Kornfield during my academic job application. I am also extremely lucky to have interacted closely with Prof. John Baldeschwieler during the Caltech 10K Business Plan competition. Sitting through John's E102 class is definitely what every Techer should experience during graduate school.

Over the last five years, I have collaborated with some truly distinguished scientists and their coworkers. Prof. Bill Goddard, Dr. Vaidehi Nagarajan, and Dr. Jeremy Kua performed the simulation studies on fluorinated GCN4-p1. Vaidehi mentored me on my initial project while Dave was still at UMass. Prof. Bill DeGrado and Dr. Giovanna Ghirlanda at University of Pennsylvania Medical School helped us in the ultracentrifugation analysis of leucine zippers. Dr. Mona Shahgohli has been extremely helpful in MALDI-MS analysis of protein samples.

I thank the Tirrell group for assistance and friendship. I realize that the progress I have made so far would not have been possible without their advice and support. When I first joined the group, Bo, Jeff, Kristi, Wendy, Karen, Dave, Jill, and Kathy provided generous help in getting me settled in the laboratory. The same group of people also made the third floor a fun and enjoyable place to be; the effects of their presence can still be seen in the laboratory today. I especially thank another UMass member, Nandita Sharma, for teaching me skills of molecular biology and the art of patience. I owe a debt of gratitude to

Prof. Kiick for her advice during my academic job search. I am very fortunate to have interacted with other former members that passed through Caltech during the last five years, especially Toshi, Ralf, Jens, Kevin, Thierry, Hanna, Youngbaek, Yaz, and Angelika. I also like to thank the current members of the group, especially Isaac, Kent, Pin, Sarah, Paul, Jinsang, Wei, Jamie, Rebecca and others for making the entire Caltech experience more positive. I thank Isaac for making our office more lively and providing expert advice in synthetic chemistry. I thank Kent for deep and thought-provoking scientific discussions and advice on life. A special thanks goes to my good colleague and friend Pin Wang. It is truly wonderful to be able to work closely with such a genuine person. His enthusiasm for science is infectious and has influenced us all. Working with Pin has made the late night shifts much more fun and productive. Winning the 10K competition with our artificial company Duragene was one of the most exciting times of the last five years.

On a personal note, I like to thank the good friends who have cheered me on during the Caltech tenure, especially the first year group: Jason, Rhett, Pat, Tim, and Thanos. We have all survived the Caltech experience! I also like to thank Garnesh, Edgardo, Suzie, and Donhee for providing friendship and assistance along the way. Even though we did not do too well, I had a lot of fun playing on the IM basketball teams for three years. The Wednesday basketball group was especially fun and entertaining.

I thank my parents, who have made tremendous efforts and sacrifices towards my development as a person. My appreciation for what you have done for me is beyond words. I am truly fortunate to have such loving and supportive parents.

Last but definitely not least, I thank my wonderful wife Michelle. My biggest accomplishment at Caltech was meeting her and convincing her to become my wife. Her kindness, warmth and encouragement made the last five years wonderful. I thank her for everything she has done for me throughout the years.

ABSTRACT

The incorporation of unnatural amino acids into recombinant proteins is an important tool for understanding protein function, engineering robust proteins and introducing useful building blocks for protein-based materials biosynthesis. While site-directed mutagenesis using natural amino acids allows one to vary protein composition and protein functions, the scope of such manipulations is limited to the twenty naturally occurring amino acids. Important chemical functionalities such as alkenes, alkynes, ketones, halides, and azides are not present in the pool of amino acids specified by the genetic code. Developing methods to insert amino acids containing these orthogonal groups, either site specifically or residue specifically, can lead to new tools in protein chemistry and protein engineering. In this thesis, we will describe the incorporation of leucine analogs into recombinant proteins *in vivo*. Our objectives are to investigate and relax the substrate specificity of *E. coli* leucyl-tRNA synthetase towards nonproteinogenic amino acids. Our results show that manipulation of synthetase activity and specificity can provide new opportunities for stabilization and chemoselective modification of proteins.

Substitution of leucine residues by 5,5,5-trifluoroleucine at the ***d***-positions of the leucine zipper peptide GCN4-p1d increases the thermal stability of the coiled-coil structure. The midpoint thermal unfolding temperature of the fluorinated peptide is elevated by 13°C at 30 μM peptide concentration. The modified peptide is more resistant to chaotropic denaturants, and the free energy of folding of the fluorinated peptide is 0.5 to 1.2 kcal/mol larger than that of the hydrogenated form. A similarly fluorinated form of the DNA-binding peptide GCN4-bZip binds to target DNA sequences with affinity and specificity identical to those of the hydrogenated form, while demonstrating enhanced thermal stability. Molecular dynamics simulation on the fluorinated GCN4-p1d peptide using the Surface Generalized Born implicit solvation model revealed that the coiled-coil binding energy is 55% more favorable upon fluorination. These results suggest that fluorination of hydrophobic substructures in peptides and proteins may provide new means of increasing protein stability, enhancing protein assembly, and strengthening receptor-ligand interactions.

To make fluorination a general method of stabilizing protein structures, we studied *in vivo* incorporation of trifluoroleucine (Tfl) and hexafluoroleucine (Hfl) in place of leucine using leucine auxotrophic *E. coli* strains. The target protein is A1, which is a leucine zipper protein that has 74 residues and eight leucines. The leucine residues are buried at the dimer interface and stabilize the protein complex. Tfl supported protein synthesis efficiently and replaced up to 92% of leucines in the protein under normal expression conditions. The yield of fluorinated protein was reduced from 40 mg/L to 20 mg/L. We were able to tune the level of fluorination by altering the concentration of competing leucine in culture media. Tfl-A1 adopted the identical helical secondary structure and dimeric aggregation order. T_m of Tfl-A1 was elevated to 67°C, a 13°C increase over A1. The concentration of urea needed to denature 50% (C_m) of protein was elevated from 2.7 M to 7 M. In contrast to Tfl incorporation, the more hydrophobic amino acid Hfl did not support protein synthesis under similar conditions. From *in vitro* characterization of leucyl-tRNA synthetase (LeuRS) substrate specificity, Hfl was shown to be activated 4100 times slower than leucine (compared to the 240-fold rate attenuation of Tfl). The decreased rate of $tRNA^{Leu}$ aminoacylation by Hfl resulted in insufficient amounts of Hfl-$tRNA^{Leu}$ during protein synthesis. We raised the cellular LeuRS activity eightfold at the time of protein induction by overexpressing LeuRS under a constitutive promoter during cell growth. Under these conditions, Hfl was effectively incorporated into A1 at ~80% substitution rate. The presence of the nearly perfluorinated side chains in the protein core enhanced protein stability even further. T_m was increased to 76°C and ΔG_f decreased by 3.6 kcal/mol. More remarkably, C_m of Hfl-A1 was not observed within the urea solubility limit.

To further broaden the chemical functionality available for protein engineering, we investigated the proofreading mechanism of leucyl-tRNA synthetase (LeuRS). The aaRSs that activate the hydrophobic amino acids leucine, isoleucine and valine employ a proofreading mechanism that hydrolyzes noncognate aminoacyl adenylates and misaminoacylated tRNAs. Discrimination between structurally similar amino acids by these AARSs is believed to operate by a double-sieve principle, wherein a separate editing domain governs hydrolysis based on the size and hydrophilicity of the amino acid side chain. Leucyl-tRNA synthetase (LeuRS) relies on its editing function to correct misaminoacylation of $tRNA^{Leu}$ by isoleucine and methionine. Thr252 of *E. coli* LeuRS has been shown previously to be important in defining the size of the editing cavity. Here we report the isolation and characterization of three LeuRS mutants with point mutations at this position (T252Y, T252L, and T252F). The proofreading activity of the synthetase is

significantly impaired when an amino acid bulkier than threonine is introduced. The misaminoacylation rate of $tRNA^{Leu}$ by isoleucine and valine increases with increasing size of the amino acid substituent at position 252, and the noncognate amino acids norvaline and norleucine are inserted efficiently at leucine sites of recombinant proteins under conditions of constitutive overexpression of the T252Y mutant in *E. coli*. In addition, the unsaturated amino acids allylglycine, homoallylglycine, homopropargylglycine and 2-butynylalanine all support protein synthesis. These results demonstrate that programmed manipulation of the editing cavity can allow *in vivo* incorporation of novel protein building blocks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

General Features of Aminoacyl-tRNA Synthetases, Transfer-RNAs, and Recent Progress in Incorporation of Unnatural Amino Acids

1.1 Aminoacyl-tRNA Synthetases

General Features

The accuracy of translation of genetic information from messenger RNA (mRNA) to protein is maintained by a class of enzymes known as aminoacyl-tRNA synthetases (aaRSs). These enzymes are ubiquitous in all organisms in the three kingdoms of life. AARSs catalyze the ligation reactions between amino acids and their cognate tRNA molecules with high specificity (1). The product of the reaction is an aminoacyl-tRNA (aa-tRNA), which is recruited by the ribosomal machinery for protein synthesis. The accuracy of the ligation reaction is thus crucial for the faithful translation of genetic information. Nature has evolved separate aaRSs for each of the twenty naturally occurring amino acids. Each aaRS is specific for a single amino acid substrate and a set of cognate tRNA molecules corresponding to the particular amino acid. The modes of recognition between aaRS, amino acid and tRNA have been the subject of intense research for the last twenty years (2-6).

AARS catalyzes the aminoacylation reaction in two steps as shown in Scheme 1.1.

Scheme 1.1

$$\text{ATP} + \text{aa} \xrightleftharpoons{\text{aaRS}} \text{aa-AMP} + \text{PP}_i \qquad (1)$$

$$\text{aa-AMP} + \text{tRNA} \xrightarrow{\text{aaRS}} \text{aa-tRNA} + \text{AMP} \qquad (2)$$

In the first step, the cognate amino acid is activated by the synthetase in the presence of adenosine 5'-triphosphate (ATP) to form the adenylated amino acid (aa-AMP), a mixed anhydride that is highly reactive. In the second step, tRNA binds to the synthetase and attacks the adenylate intermediate with either the 2'-OH or 3'-OH of the terminal ribonucleotide (7).

Although the twenty aaRSs perform essentially the same reaction with different substrates, the synthetases can be grouped into two unrelated classes, each with ten members. The existence of two classes of synthetases was suggested from aligning primary sequences of all known aaRSs and noting functional differences between them (8). The hypothesis has been subsequently confirmed by comparing the tertiary structures of the enzymes (9). Schimmel et al. first noted a conserved region in both the N-terminal sequences of isoleucyl-tRNA synthetase (IleRS) and methionyl-tRNA synthetase (MetRS, from

here on, synthetase specific for a particular amino acid is designated by the three letter amino acid abbreviation, followed by RS) (10). The region contains the tetrad HxGH, which is found in the N-termini of all class I synthetases, (11). Also conserved in class I synthetases is the motif KMSKS, which is involved in binding to the adenylate intermediate during catalysis (12). The aaRSs belonging to class I are GluRS, GlnRS, ArgRS, ValRS, IleRS, LeuRS, MetRS, TyrRS, and TrpRS. Three general motifs are found in class II synthetases, which include ProRS, SerRS, ThrRS, AspRS, AsnRS, HisRS, LysRS, PheRS, GlyRS, and AlaRS (8, 13). Motif 1 contains the pattern +G(F/Y)xx(L/V/I)xxPϕϕ, where + is a charged residue, x is any amino acid and ϕ is a hydrophobic residue. Motif 2 contains +ϕϕxϕxxxFRxE and motif 3 contains ϕGϕGϕGϕϕERϕϕϕϕ. These motifs are more variable than the conserved sequences found in class I (4). GlyRS and AlaRS do not contain the first two motifs (8). Functional differences were also noted between the two classes. Class I enzymes aminoacylate the terminal adenosines of tRNAs at the 2'-OH positions, while class II aaRSs aminoacylate the 3'-OH positions (14, 15), with the exception of PheRS. Amino acids with hydrophobic side chains are charged by class I enzymes, while neutral amino acids are substrates of class II enzymes. For a pair of amino acids with similar chemical functionalities, aaRS corresponding to the smaller amino acid always belongs to class II, while that specific for the larger amino acid is a class I enzyme (for example, LysRS belongs to I, while ArgRS belongs to II) (8).

The sizes of aaRSs vary from fewer than 350 residues (*Bacillus stearothermophilus* TrpRS, 325) (16) to well over 1000 (yeast LeuRS, 1080) (17), yet these enzymes all catalyze the same aminoacylation reaction. The aggregation states of aaRSs vary from monomers (class I, except TyrRS and TrpRS, which are dimers) to dimers (class II, except AlaRS, which is a tetramer; GlyRS and PheRS are $\alpha_2\beta_2$ heterotetramers). Jasin et al. constructed amino-terminal fragments of *E. coli* AlaRS (875 aa) and showed that the domain arrangement along the AlaRS primary sequence is modular (18). A fragment containing residues 1 through 385 is competent in *in vitro* adenylation. Addition of 80 more residues at the C-terminus restored the protein's capacity to correctly aminoacylate $tRNA^{Ala}$. The 468-residue fragment was also able to complement an *AlaS*⁻ *E. coli* mutant (gene of AlaRS removed by homologous recombination) (19) *in vivo*, despite a fivefold reduction in catalytic efficiency. The remaining portions of the synthetase were shown to facilitate self-assembly of the enzyme from monomers to tetramer.

While tetramerization of AlaRS was shown to be unnecessary for protein function in the previous work, dimerization of TyrRS is critical for enzyme function (20). Carter et al. assembled heterodimeric TyrRS complexes in which one monomer contained a point mutation that disabled amino acid activation (first step), and the other monomer truncated in the C-terminus and was defective in aminoacylation (second step) (20). Carter showed that $tRNA^{Tyr}$ was able to bind to the heterodimeric complex and was aminoacylated correctly with tyrosine. However, when one subunit that was defective in both functions was reassembled with an intact monomer, the complex was completely inactive. From these studies, the authors concluded that a single $tRNA^{Tyr}$ binds across one TyrRS dimer and interacts with both subunits. Using similar approaches, Bedouelle et al. identified residues on the surface of the dimer that are in direct contact with tRNA (21, 22).

Structural Features of aaRS

The crystal structures of all of the aaRSs have been elucidated with the exception of AlaRS (23-48). The structural features of the two classes of enzymes are distinct, suggesting different evolutionary roots. The three-dimensional structures of ValRS (Class I) (37) and SerRS (Class II) (33) complexed with their cognate tRNAs are shown in Figures 1.1 and 1.2, respectively.

Class I enzymes are elongated molecules that contain large N-terminal and C-terminal domains (43). The catalytic cores located at the N-termini adopt characteristic dinucleotide folds which consist of five parallel β-strands (Figure 1.1, middle of the structure). This structural fingerprint, named the Rossmann fold (49), is found in all dehydrogenases and is a common structure in other ATP utilization enzymes, including kinases. The β-strands are interrupted by flanking α-helices, which make significant contacts with ATP, amino acid and the acceptor stem of incoming tRNA. The histidine groups in the conserved HxGH sequence interact with the phosphates directly or through hydrogen bonded water molecules (28, 43). The second lysine in the KMSKS motif also directs the binding of ATP by interacting with the α and γ phosphates. Residues in both motifs help to stabilize the pentacoordinated transition state of the phosphate group during catalysis and secure the activated amino acid in the binding pocket, preventing hydrolysis of the reactive adenylate by water molecules (4). The large number of hydrogen bonds between Tyr-AMP and the TyrRS active site result in a dissociation constant of 12 pM for Tyr-AMP

(50-53). The interactions also rigidify the conformation of the aa-AMP and allow in-line attack by the 2'-OH moiety on the terminal adenosine of the tRNA. Amino acid specificities of aaRSs are determined by residues lining the terminal walls of the active sites. For example, in TyrRS, the side chains of an aspartic acid and a tyrosine hydrogen bond to the phenolic -OH and serve to prevent phenylalanine from binding (28). Similarly in yeast ArgRS, Asp351 forms two hydrogen bonds with the unique guanidinium side chain of arginine (30). For hydrophobic amino acids, such as leucine (35) and phenylalanine (39), the binding pockets are mostly hydrophobic. The side chains of three tryptophan residues (Trp495, 456 and 462) tightly surround the aliphatic side chain of a bound valine-adenylate in the *T. thermophilus* ValRS active site (54), excluding larger residues from entering the pocket (37). Also found in aaRSs associated with aliphatic hydrophobic amino acids, such as leucine, valine and isoleucine, are large insertion domains that proofread against structurally similar, incorrectly activated amino acids (35, 37, 41). The connective peptide domains (CP1) are approximately two hundred amino acids in size and their primary sequences interrupt the Rossmann fold of the catalytic domains. The functional significance of the CP1 domains will be discussed in detail in section 1.3.

The C-terminal residues of class I enzymes interact with the anticodon loops of tRNAs and play important roles in recognizing cognate tRNAs (Figure 1.1) (40, 43). The structures of these domains vary significantly between class I enzymes. For example, in GluRS, the C-terminus folds into a β-barrel structure (43). The domain distorts the anticodon loop of $tRNA^{Glu}$ and interacts with all three nucleotides in the anticodon loop. In contrast, the GlnRS C-terminus consists entirely of α-helices (40). The helices assemble into a bundle, driven by a large number of leucines in the core. GlnRS uses a cluster of positively charged amino acids at the tip of the helical bundle to interact with $tRNA^{Gln}$.

The characteristic Rossmann fold is not found in any of the Class II aaRSs (31). Instead, class II structures are centered around a seven-strand, anti-parallel β-sheet fold (Figure 1.2). The β-sheets are flanked by at least four conserved α-helices. The three conserved motifs found in class II synthetases contribute to different aspects of enzyme structure and function. Motif 1 is often found at the dimer interfaces of α2 dimers (44). Motif 2 and motif 3 span three of the seven β-strands and are intimately involved in ATP binding and catalysis. Unlike class I enzymes which fix ATP in an extended form, the active-site residues of class II synthetases distort the conformation of ATP (55). The γ–phosphate group is

bent back over the adenine base, exposing the α-phosphate for nucleophilic attack by the amino acid, which is buried deeper in the active site. The adenine is always stacked against the phenyl ring of a conserved Phe residue in the active site (44). The phosphate groups are fixed in position by invariable arginine residues in motif 2. The transition state formed during adenylate formation is also stabilized by conserved residues in motifs 2 and 3 (9).

The mechanisms of tRNA anticodon recognition are significantly different for enzymes in this class as well. Based on the modes of aaRS-anticodon interactions, class II aaRS can be further divided into two subgroups (4). Class IIa consists of ProRS, ThrRS, HisRS, GlyRS and SerRS (56-58). With the exception of SerRS (27), aaRSs in this group have C-terminal domains that interact with two of the three nucleotides (N35 and N36) in the anticodon loops (34). These two nucleotides determine the identities of the tRNAs that correspond to Thr, Gly, and Pro (each of these amino acids is specified by one codon group, in which the third nucleotide is a "wobble"). The C-termini of class IIa enzymes fold into α/β structures. In ProRS, a proline and a phenylalanine form a hydrophobic patch where G35 and G36 are supported (34). The wobble nucleotide N34 forms no specific contacts with residues in this domain. In contrast, class IIb enzymes (LysRS, AsnRS and AspRS) contact the anticodon loops via N-terminal domains that adopt β-barrel conformations (32, 59). For these amino acids, all three anticodon nucleotides constitute identity elements in their respective tRNAs. All three bases interact with one surface of the β-barrel, which permits a larger contact area compared to the α/β structure found in class IIa aaRSs

Because of their structural differences, class I and class II synthetases approach the acceptor stems of tRNAs differently. Class I enzymes bind to the acceptor stem from the minor groove side, while class II enzymes approach from the major groove side (44). A consequence of the different aaRS-tRNA interactions is the conformation of the bound tRNA acceptor stem. In a class II aaRS-tRNA complex, the acceptor arm does not undergo major conformation change upon docking on aaRS. The helical orientations of the last two base pairs and the unpaired *N*CCA tip are not interrupted and the 3'-OH of the terminal adenosine is juxtaposed to the α–phosphate of the bound aa-AMP intermediate. On the other hand, the first two base-pairs in a class I tRNA are disrupted completely and the single stranded *N*CCA arm enters the binding pocket in a hairpin conformation (43, 44). The bent orientation of the acceptor arm results in an attack on the aa-AMP by the 2'-OH of the terminal adenosine (8).

AARSs of M. jannaschii

Most organisms contain twenty different aaRSs for each of the natural amino acids, with a few striking exceptions. The recently sequenced genome of the methanogenic archaeon *Methanococcus jannaschii* contained only 16 aaRS ORFs (60). The genes for AsnRS, GlnRS, ProRS and LysRS were not found by homology search. The lack of AsnRS and GlnRS proteins were explained by the dual functions of AspRS and GluRS (5, 61). The abilities of AspRS to aspartylate both $tRNA^{Asp}$ and $tRNA^{Asn}$, and the abilities of GluRS to glutamylate both $tRNA^{Glu}$ and $tRNA^{Gln}$ have been noted in other prokaryotic organisms, such as *Thermus thermophilus* (62) and plant organelles (63). The misaminoacylated tRNAs are recognized by specific amidases, which convert the noncognate carboxylic acid moieties into amides. Translation fidelity is maintained by ribosomal accessory proteins, such as the elongation factor Tu (EF-Tu). Binding to EF-Tu is tolerated only when the conversion of Asp-$tRNA^{Asn}$ to Asn-$tRNA^{Asn}$ or that of Glu-$tRNA^{Gln}$ to Gln-$tRNA^{Gln}$ is completed by the amidases (62).

A protein with lysine-dependent ATP-pyrophosphate (PPi) exchange activity was purified from *M. jannaschii* whole cell extracts (64). N-terminal sequencing of the protein, followed by Southern blot hybridization led to the discovery of *M. jannaschii LysS*. Surprisingly, the protein encoded by this gene shows no sequence homology to any of the known class II LysRSs and lacks all three of the motifs shared by members of this class. Instead, signature sequences HxGH and KSMSK were found and the protein shares strong sequence homology with class I aaRSs. Comparison of *M. jannaschii* LysRS with other euryarchaeal LysRS sequences shows that a number of archaea contain class I LysRSs (64, 65).

Perhaps even more surprising about *M. jannaschii* aaRSs organization is the discovery of a ProRS that is able to charge proline to $tRNA^{Pro}$ and cysteine to $tRNA^{Cys}$ in separate reactions (66, 67). The additional cysteinylation activity of ProRS compensates for the lack of a separate CysRS in the genome of this organism. The two aminoacylation activities are comparable and no cross-aminoacylation between amino acids and noncognate tRNAs is observed. Analysis of the ProRS sequence showed there are no additional active sites or tRNA binding domains. The mechanism whereby ProRS maintains specificity in two distinct reaction is not known. The consensus view is that upon amino acid activation, accessory proteins are recruited by ProRS to assist in the discrimination between $tRNA^{Pro}$ and $tRNA^{Cys}$.

Other functions of AARSs

In addition to supplying the building blocks of protein biosynthesis, several aaRSs are involved in other cellular processes (68), including transcription regulation (69, 70), RNA splicing (71), DNA binding (70) and tRNA proofreading (72), etc. Putney and Schimmel showed that *E. coli* AlaRS binds to a palindromic sequence upstream of the *AlaS* gene and controls transcription of its own gene (69). The suppression of *AlaS* transcription is further enhanced at elevated concentrations of alanine, possibly due to stronger protein-DNA interaction upon binding of alanine to AlaRS. Similarly, PheRS from *T. thermophilus* binds to the ORF of one of the PheRS subunits and blocks transcription (70). In a different mode of autoregulation, ThrRS binds to its own mRNA and prevents ribosome binding and protein translation (45). Molecular mimicry between the secondary structures of the mRNA leader region and the anticodon loop of $tRNA^{Thr}$ promotes the idiosyncratic protein-RNA interactions. Deletion of a ThrRS domain that interacts with the acceptor stem of $tRNA^{Thr}$ significantly decreased its affinity for the tRNA, but had minimal affect on mRNA recognition (45).

AARSs of eukaryotic organisms contain nuclear localization signaling sequences that facilitate their transport into the nuclei, where eukaryotic tRNAs are proofread and aminoacylated by aaRSs before export to the cytoplasm (73). Wakasugi and Schimmel recently showed that the C-terminus of human TyrRS contains a domain highly homologous to the endothelial monocyte-activating polypeptide II (EMAP II). When the 169-residue polypeptide was cleaved from the rest of TyrRS, the fragment behaves as interleukin-8 and stimulates production of TNF-α and other cytokines. The authors speculated that human TyrRS, when secreted and cleaved, is involved in the signal transduction of cell apoptosis and macrophage recruitment (72, 74).

1.2 AARS:tRNA Interactions

AARSs maintain translational fidelity by joining the correct amino acids with correct tRNA molecules. Each of the twenty aaRSs has evolved to recognize only one amino acid (with the exception of ProRS from *M. jannaschii*). In contrast, the degeneracy of the genetic code (more than one codon per amino acid) requires that each synthetase must recognize a set of tRNA isoacceptors, which are different

tRNAs containing anticodons corresponding to the same amino acid (2). Because each tRNA is processed in an identical fashion on the ribosome regardless of its amino acid association, tRNAs species are similar to each other in both sequences and structures. Hence, the mode in which each aaRS selects a small set of substrates in a pool of nearly-identical tRNAs involves subtle differences in protein-RNA interactions (75-77).

The primary sequences of tRNAs from different organisms are highly homologous. The lengths of tRNAs vary from 76 to 93 residues depending on the sizes of the variable loops. Figure 1.3 shows the base-pairing observed in the secondary structure of yeast $tRNA^{Phe}$ in a cloverleaf representation (78, 79). Figure 1.4 shows the tertiary structure of $tRNA^{Phe}$ determined using X-ray crystallography (80, 81). The acceptor stem, located on the opposite end of the anticodon loop, is where an amino acid is ligated. The tip of the stem contains four unpaired nucleotides of which the trinucleotide CCA is universally conserved. The first unpaired base (N73) plays a critical role in establishing tRNA identity, and is thus named the discriminator nucleotide (77). The tRNA folds into the L-shaped structure shown in Figure 1.4, stabilized by tertiary interactions between bases in the D-loop and the T-loop, and N15-N48 (Figure 1.3) (82). Because of their roles in stabilizing the tRNA structures, D-loop and T-loops bases are well conserved and are usually not involved in aaRS recognition (exceptions include *E. coli* $tRNA^{Leu}$ and yeast $tRNA^{Phe}$, see below). The interactions between N8 and N14 induce a hairpin turn in the D-stem region of tRNA. The D-stem, anticodon loop and acceptor stem form the inside surface of the L-shape tRNA and interact with most of the aaRS molecule, as suggested first by Rich and Schimmel (82). Bases in these three regions hence play determining roles in aaRS recognition. Schulman and Pelka modified adenines and cytosines in $tRNA^{Met}$ using chloroacetaldehyde (83). Active tRNA fractions were separated from inactive fractions and the locations of modified bases were compared. Modifications to the anticodon loop bases C34 and A35 inactivated $tRNA^{Met}$ completely, while altering the D-loop bases had no major effect on MetRS recognition. The variable loops are less well conserved among tRNAs (Figure 1.3). The length of this region is usually 6-10 bases for class I tRNAs. Class II tRNAs, which include isoacceptors of $tRNA^{Ser}$, $tRNA^{Leu}$ and prokaryotic $tRNA^{Tyr}$, contain long (up to 20 bases), based-paired variable arms (Figure 1.10) (76, 84). The unique arms of these tRNAs isoacceptors are important for aaRS recognition (33).

Two experimental approaches have been used to study tRNA identity elements (85, 86). The first method uses genetically modified tRNAs containing the amber suppressor anticodon CUA in their anticodon loops. A target protein carrying a stop codon (TAG) at the N-terminus is coexpressed with the amber suppressor tRNA gene. The identity of the amino acid specified by TAG is checked with N-terminal sequencing, and is used to correlate with any changes made to rest of the tRNA. Normanly and Abelson have used this method to elucidate the identity elements of *E. coli* $tRNA^{Ser}$ and successfully switched a $tRNA^{Leu}$ to $tRNA^{Ser}$ by introducing the $tRNA^{Ser}$-specific elements (85).

This technique is limited to the study of tRNAs that do not involve anticodon loops in aaRS:tRNA interactions, such as $tRNA^{Leu}$ and $tRNA^{Ser}$. Hou and Schimmel constructed libraries of $tRNA^{Ala}_{(CUA)}$ mutants and showed a single mismatched base pair on the acceptor stem (G3-U70) is the sole identity determinant in AlaRS recognition (87). Transplanting this pair into other *E. coli* tRNAs, including $tRNA^{Cys}$, $tRNA^{Phe}$ and $tRNA^{Lys}$, switched the identities of these tRNAs to that of $tRNA^{Ala}$. Alanines were found at TAG-coded positions in target proteins when these hybrid tRNAs are coexpressed (87).

Recent *in vitro* tRNA aminoacylation studies have benefited from the runoff transcription method developed by Uhlenbeck et al. (86, 88-90). The tRNA gene is placed under the control of a bateriophage T7 promoter and transcribed *in vitro* in this method. The acceptor stem CCA trinucleotide is revealed through restriction digestion of an engineered *BstNI* (5' CC/AGG) site prior to T7 polymerase transcription. This technique allows synthesis of large quantities (milligram scale) of unmodified tRNAs. Although tRNAs produced via these steps do not undergo base modifications as when they are transcribed *in vivo*, most unmodified tRNAs show minimal loss of activity in aminoacylation when compared to the posttranscriptionally modified tRNAs (86, 91).

Using tRNAs generated *in vitro*, Uhlenbeck and coworkers showed that five bases contribute to the identity of yeast $tRNA^{Phe}$ (Figure 1.3, bottom) (92). All three of the anticodon bases, a base in the D-loop (G20) and the discriminating nucleotide in the acceptor stem (A73) are sufficient to render a number of tRNAs substrates of yeast PheRS. Inserting these identity elements into *E. coli* $tRNA^{Phe}$ (which is a poor substrate of yeast PheRS), yeast $tRNA^{Met}$, $tRNA^{Arg}$ and $tRNA^{Tyr}$ all resulted in hybrid tRNAs capable of being aminoacylated by yeast PheRS (92). The aminoacylation rate for each of the engineered tRNAs is comparable to that of unmodified yeast $tRNA^{Phe}$. Uhlenbeck was able to show that each of the five

elements contributes equally to the identity of yeast $tRNA^{Phe}$, and the loss of any one of the elements can result in a tenfold decrease in PheRS specificity. Similarly, Schulman and Pelka switched the identities of *E. coli* $tRNA^{Met}$ and $tRNA^{Val}$ by constructing hybrid tRNAs with exchanged anticodon loops (93). The aminoacylation efficiency of $tRNA^{Val}$ (UAC → CAU) in MetRS catalyzed methionylation was comparable to that of wild type $tRNA^{Met}$ (CAU). Soll and coworkers identified the anticodon and acceptor stem nucleotides as the major recognition elements of *E. coli* $tRNA^{Glu}$ by generating a comprehensive library of $tRNA^{Glu}$ mutants using the *in vitro* transcription method (94).

Posttranscriptional modifications of tRNA bases do establish additional identity elements for some tRNAs (95). One of the isoleucine isoacceptors, $tRNA^{Ile}_2$, contains a modified cytidine base, lysidine (L34) (Figure 1.5) (96). This processed tRNA is essential for the translation of the codon AUA, which encodes isoleucine. The modified base prevents misaminoacylation of $tRNA^{Ile}_2$ by MetRS (methionine and isoleucine are all encoded by AU*N*, where *N* is G for methionine). Unmodified $tRNA^{Ile}_{2,(CAU)}$ is efficiently recognized and methionylated by MetRS (97). C34 is a critical identity element for $tRNA^{Met}$ and the installment of a lysidine base at this position in $tRNA^{Ile}$ serves as a negative determinant against MetRS recognition *in vivo*. In another example, methylation of a single base in yeast $tRNA^{Asp}$ (G37 → m^1G37) prevents ArgRS misaminoacylation of $tRNA^{Asp}$ with arginine (98). The methylated guanine base reduces the rate of arginylation of $tRNA^{Asp}$ by more than 430-fold.

Both *in vivo* amber suppression and *in vitro* aminoacylation techniques have generated a large compilation of identity elements of tRNAs from different organisms (76). In section 1.4, I will discuss the essential features of $tRNA^{Leu}$ isoacceptors and their interactions with LeuRS.

1.3 AARS Proofreading Mechanisms

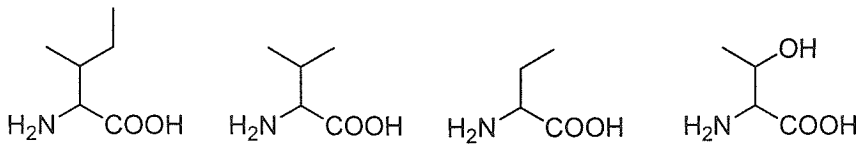
The multiple tRNA identity elements suppress errors in tRNA selection during translation to a frequency of 10^{-6} *in vivo* . DNA replication has an error frequency of 10^{-6}-10^{-10} and is maintained by a proofreading mechanism in which incorrectly inserted nucleotides are removed (99). Proofreading mechanisms have also been observed in the aminoacylation reactions catalyzed by aaRSs (100). After the initial misactivation of a noncognate, structurally similar amino acid, the aaRS corrects the error and prevents the release of a misaminoacylated tRNA. Proofreading by aaRS significantly improves

translation fidelity. The error rates of protein synthesis have been determined to be between 10^{-3} and 10^{-4} (101, 102).

Editing activities are found in numerous aaRSs, including MetRS (103), PheRS (104), AlaRS (105), ThrRS (106), ValRS (107), IleRS (108), and LeuRS (109). MetRS edits against homocysteine, an immediate metabolic precursor of methionine, through formation of a homocysteine thiolactone (103). Proofreading decreases MetRS specificity toward homocysteine by a factor of 60. Using different rejection pathways, PheRS and AlaRS edit for tyrosine (104) and serine (105), respectively. The most robust proofreading machineries are present in ValRS, IleRS, and LeuRS. Each of these enzymes must strictly discriminate among the aliphatic amino acids valine, isoleucine and leucine to maintain cell viability. More than thirty years of biochemical and structural studies on the editing mechanisms of IleRS and ValRS have revealed their highly sophisticated (and analogous) modes of maintaining aminoacylation fidelity. I will discuss the editing mechanisms of IleRS and ValRS in this section, and provide a description of LeuRS proofreading in section 1.4.

Discriminating between valine and isoleucine is an especially challenging task for ValRS and IleRS. These two amino acids differ by one methylene unit. Pauling proposed that the misincorporation frequencies for valine and isoleucine in proteins should be 1 out of every 5 residues, based on his calculation that one methylene group contributes to 1 kcal/mol of hydrophobic binding energy (110). The binding energy in substrate/protein interactions was recalculated to be 3.4 kcal/mol per methylene group (111), which suggests that the minimal error rate of inserting a valine in isoleucine positions should be 1 in 200. However, amino acid composition analysis of cellular proteins shows that the actual misincorporation rate of valine at isoleucine codon positions is 1/3000 (102, 112) and that of isoleucine at valine codons is less than 1/60,000 (113). Similarly, the isosteric threonine is activated 250-fold less efficiently than valine by ValRS, while the *in vivo* misincorporation rate is less than 1/45,000 (100, 114).

Scheme 1.2



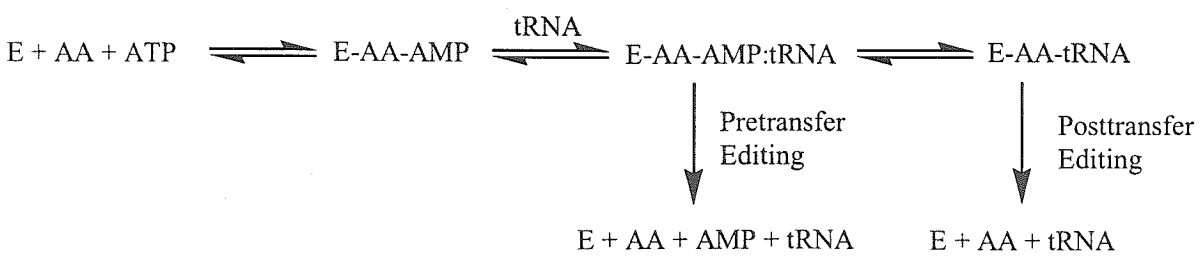
Isoleucine Valine α-aminobutyric acid Threonine

Evidence of an editing mechanism in IleRS came from the work of Baldwin and Berg in 1966 (115). The authors purified both IleRS-Val-AMP and IleRS-Ile-AMP and demonstrated that both isoleucine and valine were activated by the synthetase. However, when $tRNA^{Ile}$ was added to either complex, valine was rapidly hydrolyzed from the enzyme, whereas Ile-$tRNA^{Ile}$ was recovered as the only aminoacylated product. A hydrolytic mechanism specific for valine must be present in IleRS. Hopfield theorized that an incorrectly processed intermediate has access to a rejection pathway in addition to the main pathway (116), and he verified the proposal experimentally (112, 114).

Fersht and coworkers suggested a double-sieving mechanism is present in both ValRS and IleRS, and may contribute to the improved substrate selectivity (107, 113). The model, which has been verified in both synthetases through structural studies (37, 41), states that IleRS and ValRS each contain two sites of amino acid discrimination. The first site (the catalytic synthetic site) serves as a coarse sieve and strictly precludes amino acids larger than the intended substrate from binding and activation (113). However, amino acids smaller than the cognate substrate are tolerated in this site, albeit with decreased affinities. Misactivated amino acids are rejected by a second site (the editing site) which serves as a fine sieve. This sieve hydrolyzes the noncognate amino acid, either as an adenylated intermediate or as a tRNA-bound product, based on the goodness of fit of the amino acid side chain. Cognate substrate is forbidden from binding to the second site and is released as an aa-tRNA for protein synthesis. For example, in IleRS and ValRS, the coarse sieves preclude the binding of leucine and isoleucine, respectively, while the fine sieves reject valine and threonine, respectively.

Scheme 1.3 shows two possible pathways in which editing of noncognate amino acid can occur, both are tRNA dependent (100). In the pretransfer mode, the activated amino acid adenylate is hydrolyzed by aaRS upon tRNA binding. In the posttransfer mode, the tRNA is aminoacylated by the aaRS with the wrong amino acid, followed by deacylation to yield the free tRNA and amino acid.

Scheme 1.3



Both pathways have been independently established *in vitro*, although their individual contributions to the overall rate of editing have been difficult to separate. Berg showed the first evidence of tRNA-dependent pretransfer editing of IleRS-Val-AMP (115). The pretransfer editing activity is highly dependent on tRNA identity and integrity (117). $tRNA^{Ile}$ is believed to induce a conformational change in IleRS, which allows the translocation of the adenylate to the fine sieve. Schimmel and coworkers stimulated hydrolysis of IleRS-Val-AMP using a DNA aptamer, which was isolated from a library of DNA oligomers (65-mers) with randomized sequences (118, 119). Multiple rounds of selection for DNAs that bind to immobilized IleRS-Val-AMP, followed by elution and amplification, converged to one specific DNA sequence. The DNA aptamer is predicted to have a tertiary structure completely different from that of $tRNA^{Ile}$. The lack of a 2'-OH, which is aminoacylated in the posttransfer pathway, shows that acceptor function is not necessary for pretransfer editing and the adenylated amino acid shuttles between the two sieves via a diffusive process. Crosslinking between aptamer and IleRS showed that parts of the DNA interact with the CP1 domain of the synthetase (119).

Posttransfer editing was first established by Schimmel and coworkers using the misaminoacylated, Val-$tRNA^{Ile}$ (108). Synthesis of Val-$tRNA^{Ile}$ was first performed with wild type IleRS in 20% DMSO, which induced errors in aminoacylation (108). Subsequently, isolation of editing-impaired IleRS mutants has greatly facilitated the synthesis of misaminoacylated tRNAs for the study of posttransfer editing (120). While IleRS slowly deacylated Ile-$tRNA^{Ile}$ (0.8 s^{-1}), it hydrolyzed Val-$tRNA^{Ile}$ with a rate constant of 10 s^{-1} (108). Fersht and coworkers detected misaminoacylated α-aminobutyric acid (αbu–$tRNA^{Val}$) and Thr-$tRNA^{Val}$, followed by hydrolysis of these transient species, using rapid quenching and sampling experiments (107). However, similar experiments performed with valine and IleRS did not result in the detection of Val-$tRNA^{Ile}$ (121), suggesting that posttransfer editing might not be the dominant proofreading pathway for IleRS. Posttransfer editing is unlikely to serve as the main editing mechanism, as suggested by critics of the model, mainly because the dissociation rate of a misaminoacylated tRNA from aaRS is comparable to the rate of deacylation (for $tRNA^{Ile}$, the deacylation and dissociation rates were determined to be 10 s^{-1} (108) and 12.4 s^{-1} (122), respectively). Once dissociated, the tRNA is likely to be recruited and protected by EF-Tu (123), preventing it from reassociating with the synthetase.

The aminoacylation and editing sites were shown to be separate in IleRS through site-directed mutagenesis (124). Substitution of Gly56 in the active site of *E. coli* IleRS to an alanine caused a substantial decrease in IleRS discrimination against valine in the adenylation reaction ($k_{cat}/K_m(rel)_{, Ile/val}$ deceased from 180 to ~1). The same mutation, however, had no effect on the posttransfer editing rate of Val-tRNA^{Ile}. On the other hand, certain mutations in the CP1 domain of IleRS disrupted the ability of IleRS to proofread *in vitro*, but had no effect on rates of isoleucine activation (120). In fact, the CP1 domain of IleRS has been shown to be dispensable, to a certain extent (125). IleRS mutants carrying ~50 amino-acid deletions in the CP1 domain (231-276 or 276-329, but not both) can synthesize Ile-tRNA^{Ile} efficiently *in vitro*. Reassembly of IleRS fragmented at the CP1 domain restored its ability to aminoacylate tRNA^{Ile} (126).

The CP1 domains that interrupt the Rossmann folds in IleRS, ValRS and LeuRS are considerably larger than that of MetRS. CP1 of *E. coli* MetRS contains 118 amino acids, that of IleRS contains 316 amino acids. The CP1 domains of IleRS and ValRS have been cloned and expressed as stand-alone, 6xHis tagged proteins (127). IleRS-CP1 deacylated Val-tRNA^{Ile} with the same rate as that of intact IleRS, suggesting that it is entirely responsible for (at least) posttranslational proofreading. ValRS-CP1 was also shown to be equivalent to ValRS in deacylating Thr-tRNA^{Val}.

Crystal structures of *T. thermophilus* IleRS (41, 46), *Staphylococcus aureus* IleRS complexed with tRNA^{Ile} (46), and *T. thermophilus* ValRS complexed with tRNA^{Val} (37) provide convincing structural evidence for the double-sieve model and posttransfer proofreading. Valine was bound in both the active site and the editing site of *Tt* IleRS (41), while isoleucine was bound only in the active site. Gly45 in the *Tt* IleRS structure (Gly56 in *Ec*) is part of the active site as predicted by Schimmel (124). In the *Tt* ValRS structure, the corresponding amino acid is replaced by proline, resulting in a smaller substrate binding pocket which excludes isoleucine (37). The IleRS editing cavity is too small to accommodate Ile, but fits the valine side chain perfectly (41). In contrast to that of IleRS, the ValRS editing pocket recognizes amino acid side chains based on hydrophilicity (37). An aspartate residue in the editing pocket (Asp328) can hydrogen bond to the hydroxyl moiety of a threonyl side chain and facilitate recognition. The isosteric side chain of valine is composed of entirely hydrocarbon units and is thus not energetically favored to enter the hydrophilic editing site. The mode of recognizing αbu by ValRS the editing pocket is not clear. An

alternative binding groove may exist in the editing site where the linear aliphatic side chain can be identified.

The CP1 domains in *Tt* and *Sa* IleRS were crystallized in different orientations. The *Sa* CP1 is bent toward the Rossmann domain by 47° with respect to the rest of the molecule (orientation II) (46). In this orientation, the acceptor stem of $tRNA^{Ile}$ can adopt the hairpin conformation also seen at the $tRNA^{Glu}$ acceptor stem in the $tRNA^{Glu}$-GluRS complex (43), and can enter the catalytic active site. In the *Tt* structures, the CP1 domain and the Rossmann domain are in nearly a linear relationship (orientation I) (41). The tRNA can no longer be inserted into the synthetic site in orientation I due to steric clashes at the pocket entrance. Instead, the acceptor stem adopts a more helical conformation and inserts into a cavity in the editing domain, which is located 30 Å away from the synthetic site (Figure 1.1). In this conformation, the KMSKS helix that covers the active site tightly in orientation II is lifted away from the active site, revealing an opening in which the tRNA-tethered amino acid can be translocated to the editing site (37). Steitz and coworkers proposed that in posttransfer proofreading, the $tRNA^{Ile}$ molecule is anchored to IleRS and the acceptor stem can swing the amino acid from the active site to the editing site, facilitated by conformation changes in both the Rossmann domain and the CP1 domain (46). This "swinging" model is analogous to the editing of incorrectly inserted bases during DNA replication (128, 129). DNA polymerase I swings the nucleotide from the polymerase domain to the exonuclease domain 30 Å away, where proofreading can be performed.

In the pretransfer mode, the tRNA does not interact directly with the aminoacyl adenylate (118). Instead of acting as a covalent tether, tRNA participates in the formation of a translocation channel, as proposed by Yokoyama and coworkers based on the *Tt* ValRS structure (37). To induce translocation (130), the CP1 domain adopts orientation II. The tRNA acceptor stem, along with parts of the synthetase, form a channel that spans the active and editing sites. The adenylate diffuses to the editing site and binds in a different orientation compared to aa-tRNA, in order to accommodate the AMP moiety (37). The proofreading-inducing DNA aptamer studied by Schimmel and coworkers may participate in the formation of the translocation channel and play a role similar to that of tRNA during pretransfer proofreading (119). Although no direct structural evidence of the channel is available, this model provides the most satisfactory explanation for the dependency of pretransfer editing on tRNA binding.

Schimmel and coworkers proposed a "postpre-prepre" editing model based on studies using a translocation mutant (130, 131). A key aspartate residue (D342) in *Ec* IleRS, which hydrogen bonds to the α-amino group of Val-tRNA^{Ile} during posttransfer editing, was mutated. Even though this residue is not involved in the pretransfer pathway (37), both posttransfer and pretransfer proofreading capacities of the mutant are impaired (131). Studies using a fluorescent analog of ATP, *N*-methylantraniloyl dATP (132), showed that the translocation mechanisms of both Val-tRNA^{Ile} and Val-AMP from the active site to the editing site are severely damaged. The "postpre-prepre" model suggests that posttransfer "primes" the synthetases for pretransfer editing (131). After translocation of Val-tRNA^{Ile} from the aminoacylation site to the editing site, tRNA^{Ile} is kept bound at the editing site. The presence of a tRNA^{Ile} acceptor stem in the editing site fixes the conformation of IleRS and forms the translocation channel that allows aa-AMP diffusion. All subsequently misadenylated valines are edited via the pretransfer mechanism. This model explains the dependency of the pretransfer pathway on an intact posttransfer mechanism. The "postpre-prepre" model also suggests that under *in vivo* conditions, posttransfer editing is the dominant proofreading pathway for IleRS. Considering the two hundred-fold attenuation in the activation rate of valine compared to that of isoleucine, the occurrence of two successive rounds of valine misactivation by the same IleRS molecule is very rare.

1.4 Leucyl-tRNA Synthetase (LeuRS) and tRNA^{Leu}

General Features

LeuRS is much less well studied compared to both IleRS and ValRS. LeuRSs from more than 260 different organisms have been identified and sequenced, including *S. cerevisiae* (133) (both mitochondrial (134) and cytosolic (17)), higher plant (135), *E. coli* (136), *Bacillus subtilis* (137), and human(138). Alignment of LeuRS sequences from several organisms is shown in appendix 1.

LeuRS is a large, monomeric, class I synthetase. *E. coli* LeuRS contains 860 amino acids (136, 139), while yeast cytosolic LeuRS contains 1080 amino acids (17). Overexpressed *Ec* LeuRS has been purified to >95% purity using two chromatographic steps (140). LeuRS has also been expressed with a 6xHis affinity tag and the fusion protein can be purified using single-step nickel affinity chromatography

(141). The activities of tagged and wild-type synthetases are nearly identical. The K_m and k_{cat} of *Ec* LeuRS in the leucine adenylation reaction are 15 μM and 3.0 s^{-1}, respectively (141).

Figure 1.6 shows the sequence alignment between *T. thermophilus* LeuRS, IleRS and ValRS, adopted from Cusack et al. (35). The primary sequences of the three enzymes are highly similar and most of the catalytic motifs are strictly conserved. Mutations to conserved residues in LeuRS have detrimental effects on activity (142). LeuRS shares 22% sequence identity with IleRS, and 21% with ValRS. The CP1 domain of LeuRS spans residues 228 to 405, comparable in size to the CP1 domains found in both IleRS and ValRS. The location of the CP1 domain in the primary sequence of a prokaryotic LeuRS differs slightly from that of IleRS or ValRS (35, 143). In *Tt* LeuRS, CP1 is inserted after a unique stretch of 50 amino acids, while in *Tt* IleRS and ValRS, regions of similar sizes are inserted immediately after their CP1 domains (Figure 1.6). This domain has been associated with zinc binding in all three synthetases. Overall, two zinc binding motifs are present in this trio of enzymes. Both zinc atoms are critical to the correct folding of these enzymes. Mutations of residues in either zinc binding region of IleRS destroyed the ability of the enzyme to coordinate zinc and rendered the enzymes unstable and completely inactive (143, 144). Interestingly, in eukaryotic and archaeal LeuRSs, CP1 domains are inserted before the 50 residue zinc binding region, similar to the orders observed in IleRS and ValRS.

LeuRS also contains domains that are not found in ValRS and IleRS (35). A 60-residue leucyl-specific domain located immediately before the KMSKS signature motif is unique in the LeuRS sequence. This domain contains predominantly β-sheets as seen from the crystal structure. The role of this domain is not known. It is mostly likely involved in the specific recognition of $tRNA^{Leu}$.

LeuRS from yeast mitochondria (134), encoded by the *NAM2* gene, has been shown to exhibit an RNA splicing function (145, 146). LeuRS acts as an intron-specific splicing factor for mRNAs that encode essential respiratory proteins in yeast. The mechanism by which LeuRS recognizes specific mRNA sequences is not known. Martinis and coworkers showed that LeuRS from mycobacteria and from humans can essentially substitute for yeast mitochondria LeuRS (147). LeuRSs from these evolutionarily distant organisms are not involved in mRNA splicing in their original hosts, but are able to gain splicing capability in the presence of pre-mRNAs and to complement a *NAM2*$^-$ yeast strain.

Structural features of LeuRS

The structure of *Tt* LeuRS has been determined to 2 Å resolution and is shown in Figure 1.7 (35). As expected from sequence alignment, LeuRS displays an overall structure nearly identical to those of IleRS (41) and ValRS (37). The enzyme is highly elongated with three major domains. The catalytic domain adopts the Rossmann dinucleotide fold and is situated at the center of the structure. The signature motifs HxGH and KMSKS (in *Tt* LeuRS, the first lysine in KMSKS is replaced by a valine) are positioned at the entrance to the active site. Also observed in the structure is a tightly bound, nonhydrolyzable, sulfamoyl analog of Leu-AMP (LeuAMS). An α-helix rich domain at the C-terminus of LeuRS interacts with the anticodon loops of $tRNA^{Leu}$, and is similar to those present in other Class I synthetases, including IleRS, ValRS and MetRS. Superpositioning the C_α traces of the Rossmann domain and the C-terminus domains of LeuRS and IleRS (total of 437 amino acids) shows a RMS deviation of 2 Å (35). The editing domain of LeuRS is slightly slanted with respect to the rest of the enzyme when compared to that of IleRS. This domain is connected to the Rossmann domain via a pair of antiparallel β-ribbons. The connective region between the editing and the active sites is highly flexible, a structural feature that is required for the observed conformation change during proofreading. Superimposing 129 amino acids in the editing domains of IleRS and LeuRS results in a RMS deviation of 1.38 Å. Absent in the IleRS, but clearly visible in the LeuRS structure, is the leucyl-specific domain. This domain stems outward from the active site and is positioned to interact directly with the acceptor stem of incoming tRNAs. This unique domain is also connected to the rest of the synthetase via a flexible, β-ribbon linker. No crystal structure of LeuRS complexed with $tRNA^{Leu}$ is available at this time.

The synthetic active site of *Tt* LeuRS with bound LeuAMS is shown in Figure 1.8. The residues that interact with the aliphatic, branched side chain of leucine are shown in Figure 1.9. The pocket is formed by residues Met40, Phe41, Tyr43, Phe501, Ser504 and His545. The side chains of these residues are all within 5 Å of the isobutyryl side chain of leucine. These residues are well conserved among different LeuRSs. Surprisingly, this pocket is not entirely hydrophobic: Ser504, along with Asp80, bind a single water molecule in the active site (Figure 1.9). The water molecule is 3.5 Å away from one of the C_δ atoms of leucine. The peculiar positioning of the water molecule may contribute to the misactivation of γ-hydroxyl-leucine (109) (see below). Phe41, Phe501 and Met40 surround the C_β and C_γ carbons of the side

chain tightly, while His545 seals off the terminal wall of the binding pocket with its imidazole ring. The α-amino moiety hydrogen bonds to Asp80 and the carbonyl oxygen hydrogen bonds with the side chain of His541. The VMKSK sequence in LeuRS interacts directly with the adenosine ring of LeuAMS and fixes the analog tightly, as is observed in the TyrRS active site (28).

LeuRS/tRNA^{Leu} Interactions

tRNA isoacceptors of leucine and serine are class II tRNAs. Figure 1.10 compares the sequences and proposed secondary base-pairing of *E. coli* tRNA^{Leu} and tRNA^{Ser}. They contain long, based paired, variable arms that are not found in class I tRNAs. Prokaryotic tRNA^{Tyr} isoacceptors also possess the long arm and belong to the class II family (76). Specific interactions between tRNA^{Leu} and LeuRS are unknown because the crystal structure of LeuRS complexed with tRNA^{Leu} has not been determined. Phosphate mapping between tRNA^{Leu} from a number of species and bean cytoplasmic LeuRS with ethylnitrosourea showed base-paired tertiary elements such as N15, N48 and the D-loops are strongly protected (148). Recognition elements of tRNA^{Leu} isoacceptors from *E. coli* (149, 150), yeast (151), human (152) and the archae *Haloferax volcanii* (153) have been elucidated using *in vitro* generated tRNAs. While the long variable arms are well conserved in the different sets of tRNA^{Leu} isoacceptors, the identity elements of tRNA from these four organisms differ significantly. Figure 1.10 and Figure 1.11 show conserved tRNA^{Leu} isoacceptor elements from *E. coli*, yeast, and *H. volcanii* .

The shared positive identity element among tRNA^{Leu} isoacceptors is the invariable adenine occupying the discriminator base. The single mutation A73G in *E. coli* tRNA$^{Leu}_{(CAG)}$ reduces aminoacylation efficiency (k_{cat}/K_m of LeuRS) by a factor of 150 (149). The same mutation in human tRNA^{Leu} turned the tRNA into a serine acceptor *in vitro* (154). In contrast, guanine is conserved at the discriminator base among all tRNA^{Ser} acceptors (155). Thus, A73 in tRNA^{Leu} serves not only as an positive identity element for LeuRS, it also protects the tRNA from being serylated by SerRS.

Leucine is encoded by six codons (CUN and UUA/G) in most organisms. Five tRNA^{Leu} isoacceptors are present in *E. coli* (156-160). The codon CUU is translated via a "wobble" base pair by one of the NAG containing tRNAs. The occurrences of each codon in *E. coli* have been determined: CUG: 79%; CUC: 8%; CUU: 6%; UUG: 4%; UUA: 3% and CUA: 1% (161). Because of the large number of

leucine codons, the anticodon loop is not expected to play an important role in LeuRS recognition; a similar situation characterizes the $tRNA^{Ser}$ isoacceptors (6 codons) (162). The invariable, central base A35 found in all *E. coli* $tRNA^{Leu}$ isoacceptor anticodon loops can be changed without noticeable effects on LeuRS aminoacylation (CAG → UGA substitution decreased relative V_{max}/K_m by 25%) (149). Human LeuRS also does not use this position as an identity element (152). In contrast, yeast LeuRS relies heavily upon this nucleotide for $tRNA^{Leu}$ recognition (151). The A35G mutation in yeast $tRNA^{Leu}$ resulted in >100-fold reduction in recognition by yeast LeuRS (151).

E. coli LeuRS also specifically recognizes its cognate tRNAs based on tertiary structures of the tRNA, which are strongly influenced by the positions of residues G18 and G19 in the D-loop, and the tertiary interactions between N15 and N48 (149). *E. coli* $tRNA^{Leu}$ isoacceptors contain the conserved A15-U48 base pair and position bases G18 and G19 differently from yeast and human $tRNA^{Leu}$ (Figure 1.11). On the contrary, yeast and human LeuRS are less selective with respect to tRNA tertiary structures. As a result, unilateral aminoacylation has been observed in cross-species experiments (163): *E. coli* LeuRS is completely inactive towards yeast $tRNA^{Leu}$, while yeast LeuRS efficiently leucylates *E. coli* $tRNA^{Leu}$ (with a fourfold decrease in the aminoacylation rate). Human LeuRS is also able to aminoacylate *E. coli* $tRNA^{Leu}$ (138).

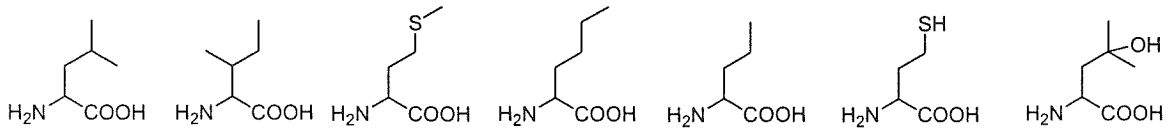
The long variable arms in $tRNA^{Ser}$ isoacceptors are the principal identity elements for SerRS (162). The loops interact specifically with a long, N-terminal coiled-coil domain in SerRS (Figure 1.2) (27). The variable arms, however, do not play important roles in LeuRS-$tRNA^{Leu}$ interactions in either prokaryotic or eukaryotic systems (149, 151). Deletions of up to two base pairs in the variable arm of either *E. coli* or yeast $tRNA^{Leu}$ had minimal effect on the rate of aminoacylation. *In vitro* selection of random tRNA sequences that are aminoacylated by *E. coli* LeuRS resulted in tRNAs that contained fixed identity elements as discussed above (A15-U48, A73, etc) and large differences in the variable loops (164). In sharp contrast to *E. coli* and yeast LeuRSs, LeuRS from the archae organism, *H. volcanii*, does use the variable loop as an identity element (153). The cloverleaf representation of conserved bases among *H. volcanii* $tRNA^{Leu}$ isoacceptors is shown in Figure 11. The unique trinucleotide $U^{47A}A^{47B}C^{47C}$ at the tip of the variable arm, as well as the conserved nucleotides at the bases of the loop, U44 and U47H, are absolutely critical for recognition between LeuRS and $tRNA^{Leu}$. Mutations to any of these residues, with

the exception of U^{47A}, are especially detrimental to the aminoacylation rate (> 1000-fold attenuation in rates have been observed for a single mutation) (153). As a result of this unique recognition element, LeuRS from *H. volcanii* is completely inactive towards *E. coli* and yeast $tRNA^{Leu}$. The tertiary structural determinants G18G19 and G15-C48 in *H. volcanii* $tRNA^{Leu}$ are the same as those in yeast, suggesting that *E. coli* LeuRS is also unlikely to recognize and aminoacylate the archaeal $tRNA^{Leu}$ (163). Lack of cross-species aminoacylation between *E. coli* and *H. volcanii* would make the LeuRS/$tRNA^{Leu}$ pairs from these organisms mutually orthogonal.

Proofreading Mechanisms of LeuRS

While the need for IleRS to proofread against valine is essential, it is not obvious which of the amino acids LeuRS must edit against. A large CP1 domain in the crystal structure clearly points to a double-sieve mechanism in LeuRS (35). None of the other nineteen natural amino acids is activated rapidly enough to pose any serious threat to LeuRS aminoacylation fidelity. Both methionine and isoleucine have been shown to be edited by LeuRS *in vitro* (165), even though their adenylation rates are significantly slower than to that of leucine (isoleucine and methionine are activated by LeuRS ~3000-fold and ~5000-fold more slowly, respectively) (166). Several nonproteinogenic, naturally present amino acids are activated by LeuRS *in vitro* (scheme 1.4) (167). Homocysteine was shown to be cyclized by LeuRS to a thiolactone independent of $tRNA^{Leu}$ (109). Proofreading by LeuRS reduces its specificity toward homocysteine by a factor of 25. Norvaline differs from leucine by one branched methylene unit, while norleucine is a linear isomer of leucine. These amino acids are synthesized by *E. coli* under stress conditions and have been suggested to be proofread by LeuRS, although no direct biochemical evidence is available to date (168).

Scheme 1.4



Englisch et al. have demonstrated that both pre- and posttransfer editing are employed by LeuRS in rejecting the noncognate amino acid γ-OH-leucine (109). This amino acid is activated by LeuRS, likely through displacement of the bound water molecule in the synthetic site. The –OH group would be able to form alternative hydrogen bonds with both Ser504 and Asp80. The authors showed that while both *E. coli* and yeast LeuRS were capable of proofreading against γ–OH-leucine, the dominant editing pathway of each enzyme is different. *E. coli* LeuRS possesses high initial discrimination towards noncognate substrates, and proofreads against misactivated amino acids mainly through a posttransfer mechanism. On the other hand, yeast LeuRS was shown to discriminate among substrates poorly in the first step, but contains a highly robust pretransfer editing mechanism.

It is reasonable to hypothesize that LeuRS proofreads using a substrate recognition mechanism based on both size and hydrophilicity, similar to that found in the ValRS editing domain (37, 107). Without crystal structures of a bound tRNA and a noncognate amino acid in the editing site, the molecular basis of LeuRS proofreading still remains obscure. A limited number of mutagenesis experiments in the CP1 domain has provided some insight into the editing activities of LeuRS.

Chen et al. constructed a mutant *E. coli* LeuRS in which the residues between amino acids 328 and 368 in the CP1 domain are duplicated (165, 169). The insertion mutant had 50% of the leucylation activity compared to the wild type enzyme (K_m was unchanged, k_{cat} decreased from 3.0 s^{-1} to 1.5 s^{-1}) and retained selectivity toward methionine and isoleucine in the adenylation reaction. However, the editing function of the mutant was significantly disrupted, as evident from the accumulation of Met-$tRNA^{Leu}$ and Ile-$tRNA^{Leu}$ during aminoacylation assays *in vitro*. The authors also expressed and purified the CP1 domain separately and found that unlike the CP1 domains of IleRS and ValRS (127), the editing domain of LeuRS cannot function properly without being covalently attached to the rest of the enzyme.

The same authors also identified alanine 293 to be a critical residue for the proper functioning of both synthetic and editing sites (170, 171). Bipartite assembly of LeuRS fragments generated at different positions yielded functioning LeuRSs, with the exception of two fragments isolated by cleaving the peptide bond between residues 292 and 293 (170). Alanine 293 is located within an α-helix in the editing domain close to the flexible linker region, and faces the active site. Chen et al. proposed that the stability of the helix is important for the flexibility of the active site and the substrate translation processes. LeuRS

variants carrying mutations at the A293 position have impaired editing capacities (171). Furthermore, the K_m for ATP binding at the active site is significantly decreased (e.g., the A293D mutant has K_m of 42 μM compared to 280 μM observed for the wild type). The increased affinity for ATP can lead to decreases in leucylation activity since ATP is an inhibitor of amino acid activation (2).

Nureki and coworkers identified an essential threonine-rich motif in the CP1 domain of IleRS (41) that is also present in LeuRS: $T^{247}TRDT^{252}$ in LeuRS and $T^{228}TTPWT^{233}$ in IleRS. Mursinna et al. have subsequently shown that T252 is intimately involved in substrate recognition during proofreading (172). Using alanine scanning mutagenesis, the authors showed that the mutation T252A significantly impaired the rates of leucine aminoacylation. Furthermore, this mutant was able to hydrolyze the correctly charged Leu-$tRNA^{Leu}$ rapidly. In the crystal structure, the threonine side chain hydrogen bonds to a water molecule and is positioned near the surface of the editing domain (35). The T252-water complex may interact with the side chains of the translocated substrates directly. In Chapter 4, I will discuss in detail the role of T252 during proofreading and our approach to disruption of LeuRS proofreading by mutating this residue to bulkier amino acids.

1.5 Unnatural Amino Acid Incorporation

The incorporation of unnatural amino acids into recombinant proteins is an important tool for understanding protein function, engineering robust proteins and introducing useful building blocks for protein-based materials biosynthesis (173). While site-directed mutagenesis using natural amino acids allows one to vary protein composition and protein functions, the scope of such manipulations is limited to the twenty naturally occurring amino acids. Important chemical functionalities, such as alkenes, alkynes, ketones, halides, and azides, are not present in the pool of amino acids specified by the genetic code. Developing methods to insert amino acids containing these orthogonal groups, either site specifically or residue specifically, can lead to new tools in protein chemistry and protein engineering .

Several strategies are currently employed for introducing noncanonical amino acids. Synthetic and semisynthetic methods are capable of producing gram-scale proteins and are unlimited by the side chain functionalities present in the amino acids. Semisynthetic protein synthesis involves the enzymatic coupling of synthetic, short peptides (174). The power of solid phase peptide synthesis (SPPS) is severely

restricted by the length of the protein, currently limited to under fifty residues. The native chemical ligation technique (175) developed by Kent et al. allows the sequential coupling of peptides containing C-terminal thioesters and is a useful strategy in circumventing the size restrictions imposed by SPPS. This method suffers from poor yields (< 25% per coupling) and the requirement of strongly denaturing reaction conditions (6 N HCl). Raines and coworkers recently reported improved peptide ligation routes utilizing C-terminal thioesters and N-terminal azides (176). The higher coupling yields and milder reaction conditions are promising for the synthesis of large proteins containing unnatural amino acids.

Site-specific Incorporation in vitro

Site-specific incorporation of unnatural amino acids using coupled transcription/translation *in vitro* has enabled the methodical insertion of a large array of amino acids at a designated position in target protein. The method, based on work by Hecht and others (177, 178), and developed by Schultz and Chamberlin and their respective coworkers (179, 180), uses chemically aminoacylated suppressor tRNAs. An unnatural amino acid is first chemically appended to the acceptor stem of a tRNA generated *in vitro*. The anticodon loop of the tRNA (yeast $tRNA^{Phe}$) is modified to an amber suppressor anticodon CUA. The engineered, aminoacylated aa-$tRNA^{Phe}_{(CUA)}$ can then decode a nonsense codon UAG, and insert the unnatural amino acid in the target protein during translation. The yeast $tRNA^{Phe}$ is not recognized by *E. coli* PheRS and is thus not deacylated or misacylated when mixed with *E. coli* cellular extracts. Most amino acid analogs tested were inserted at the intended positions in the protein of interest. Using this method, Schultz and other have investigated 1) the contribution of individual residues to protein stability (181); 2) the role of charged residues during enzyme catalysis (182); 3) the effects of replacing amide linkages with ester linkages in proteins (180); and 4) methods of protein self-splicing (183). Hecht and coworkers have used this method to induce site-specific proteolysis through an engineered allylglycine residue (183). The technique has also been extended to the incorporation of noncanonical amino acids in intact living cells (184). Dougherty and coworkers injected both the suppressor tRNA chemically acylated with an unnatural amino acid, and the nicotinic acetylcholine receptor mRNA that contains the stop codon UAG, into *Xenopus laevis* oocytes (185). The presence of unnatural residues in the ion channel induced

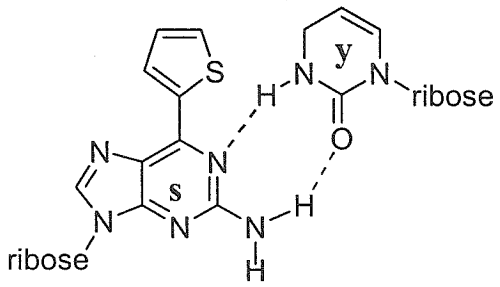
significant changes in its function *in vivo* and provided insights into the roles of individual amino acids in the channel (186, 187) and the arrangement of different receptor subunits in the membrane (188).

Competition between translation termination and amber codon suppression decreases the efficiency of this method. Intrinsic release factors detect in-frame stop codons present in mRNAs and terminate protein translation. In particular, release factor 1 (RF1) binds strongly to the UAG stop codon commonly employed for unnatural amino acid incorporation. The concentration of RF1 in whole cell extracts thus effects the relative amounts of full-length and truncated proteins synthesized. In vitro translation using cellular extracts with diminished amounts of RF1 increased the yield of full-length protein fivefold, as demonstrated by Short and coworkers (189).

Four-base codons (190) and unnatural base pairs (191) have also been used to increase the efficiency of unnatural amino acid incorporation *in vitro*. The four-base codon strategy relies on the recognition between an engineered tRNA containing four bases in the anticodon loop and a complementary four-base codon in a mRNA. Translation of the four-base codon with a tRNA appended with the unnatural amino acids leads to full length protein (190). A frame shift occurs when the first three bases in the four-base codon are recognized by a cognate tRNA. Frame-shift suppression can be achieved by placing a stop codon in frame with the first three bases. Mistranslation of the four-base codon by a cognate tRNA thus leads to a truncated protein, which can be separated from the desired, full-length protein. Streptavidins containing a variety of unnatural amino acids at different positions have been synthesized using this technique (190, 192, 193). Sisido and coworkers have used this method to study the site-to-site photoinduced electron transfer in proteins using streptavidin containing L-*p*-nitrophenylalanine, and *N*-biotinyl-L-1-pyrenylalanine (194). The same authors also investigated the tolerance of ribosome towards unnatural amino acids containing large aromatic side chains (193). The authors showed that incorporation efficiencies were higher for amino acids contain straight aromatic groups, such as 2-anthrylalanine than those with expanded groups, such as 9-anthrylanaline. These authors further increased the versatility of four-base codon translation by demonstrating the incorporation of two different amino acids independently into a single protein: 7-nitrobenz-2-oxa-1,3-diazol-4-yl)-L-lysine and L-2-naphthylanaline were inserted into positions 54 and 57 of streptavidin, respectively (192).

Unnatural base-parings also facilitate anticodon-codon recognition orthogonal to those specified by the natural bases (195). Hirao and coworks used the unnatural pair pyridin-2-one (y) and 2-amino-6-(2-thienyl)purine (s) towards the site-specific incorporation chlorotyrosine (ClTyr) in Ras protein (191).

Scheme 1.5



These two bases hydrogen bond through a pattern unique to those present among A-T and G-C pairs. During transcription, yTP is preferentially recruited by the T7 polymerase to complement a nucleotide containing s on the template DNA. Transcription of entire template was not effected by the presence of an unnatural base. Aminoacylated tRNA containing CAs in the anticodon loop recognized the yUG codon specifically and inserted ClTyr during *in vitro* translation. Full length proteins were obtained as the only translation product, in contrast to the near equal amounts of full-length and truncated proteins synthesized using the four-base codon strategy. The insertion of ClTyr at the yUG position was nearly quantitative, as determined by LC-MS.

Site-specific incorporation of unnatural amino acids *in vitro* requires synthetic joining of amber tRNAs with the target amino acid prior to coupled transcription/translation. Chemical synthesis of aminoacylated tRNA is difficult to scale up and once the amino acid is incorporated, the freed suppressor tRNA cannot be reused. Consequently, a major limitation of the *in vitro* method is poor yield of target protein recovery (microgram scale). To obtain proteins in large quantities suitable for NMR analysis, X-ray crystallography and materials applications (196), *in vivo* methods of incorporating nonproteinogenic amino acid have been investigated.

Site-specific Incorporation in vivo

To site-specifically incorporate an unnatural amino *in vivo*, several requirements must be satisfied (197). First, to prevent misincorporation of the twenty-first amino acid into positions in the target protein other than that specified by the amber codon, the analog must not be a substrate of the endogenous translation machinery. Second, an aaRS orthogonal to all host tRNAs must be engineered to avoid misaminoacylation. The aaRS must also activate the unnatural amino acid much more rapidly than the natural amino acids. Third, an amber suppressor tRNA orthogonal to host aaRSs, but is able to be aminoacylated by the orthogonal aaRS is required (198, 199). AARS/tRNA pairs from different kingdoms of life often contain vastly different recognition elements and are therefore frequently orthogonal (92, 153). Uhlenbeck and coworkers demonstrated that $tRNA^{Phe}$ from yeast and *E. coli* are orthogonal because they differ by three identity elements (92). Furter showed that an *E. coli* PheRS (ePheRS) temperature-sensitive strain can survive at the nonpermissive temperature only when both yeast PheRS (yPheRS) and yeast $tRNA^{Phe}$ are introduced (200). Transformation of this strain with either yPheRS alone, or yeast $tRNA^{Phe}$ alone failed to complement the lethal mutation. Furter then chose an *E. coli* strain that is resistant to the analog of interest, *p*-F-phenylalanine (fPhe) as the expression host (200). The chromosomal copy of ePheRS contained a mutation at the active site that rejected binding of fPhe (201). Wild type yPheRS was shown to activate fPhe efficiently in the host strain.

Equipped with an orthogonal yPheRS/$tRNA^{Phe}_{CUA}$ pair, and a host strain resistant to the unnatural amino acid, Furter successfully incorporated fPhe site-specifically into the target protein *in vivo* (200) (Figure 1.12). Less than 7% of phenylalanine positions in the target protein were substituted with fPhe. The incorporation rate of fPhe at positions specified by the amber codon was over 75%, as determined by N-terminal sequencing. The incomplete substitution was because unmodified yPheRS aminoacylated $tRNA^{Phe}_{CUA}$ more efficiently with phenylalanine than with fPhe.

Wang and coworkers from our laboratory have subsequently optimized the above system by reversing the specificities of yPheRS to one that preferentially activates unnatural surrogates (202). Mutation of Thr415 at the active site of yPheRS to a smaller glycine residue decreased its specificity for phenylalanine by 100-fold, while gained specificity towards the bulkier 2-naphthylalanine. Surprisingly, 2-naphthylalanine was activated 9-fold faster than phenylalanine as a result of the point mutation. The

chromosomal copy of ePheRS was unmodified and was nearly inactive towards the analog. When the mutant yPheRS was overexpressed, quantitative incorporation of 2-naphthylalanine at amber position was observed.

Schultz and coworkers have developed a similar approach towards the site specific incorporation of *O*-methyl-tyrosine (203) and 2-naphthylalanine (204). A TyrRS/$tRNA^{Tyr}$ pair from the archaeal organism *M. jannaschii* was introduced into an *E. coli* expression host. *Mj* TyrRS is completely orthogonal to the pool of *Ec* tRNAs, while *Ec* TyrRS aminoacylates *Mj* $tRNA^{Tyr}$ slowly. Mutagenesis of eleven nucleotides in *Mj* $tRNA^{Tyr}$ reduced its recognition by *Ec* TyrRS to below background levels (205). To translate amber codons, CUA was inserted in the anticodon loop of the modified *Mj* $tRNA^{Tyr}$. *Mj* TyrRS was engineered to gain specificity towards the unnatural amino acid, and lose specificity towards tyrosine simultaneously. In order to create a "blank" background during selection, five active-site residues in *Mj* TyrRS were mutated to alanine, which completely inactivated the synthetase (203). Combinatorial mutagenesis of these five residues generated a library of mutant synthetases. A two-step selection procedure was used to rapidly isolate mutants that regained aminoacylation activities towards only the analog. In the first step, an amber codon was placed at a nonessential position within the chloramphenicol acetyltransferase (CAT) gene. Cells were grown in the presence of all twenty natural amino acids and one unnatural amino acid. Strains that contained active mutant *Mj* TyrRS are able to translate the amber codon, express the full length CAT protein, and survive in chloramphenicol-supplemented culture. The second selection step was performed in the same way, except that cells were grown in media without the unnatural amino acid. A strain that survived the positive selection (first step), but failed to grow during the negative selection (second step), contains a mutant TyrRS active towards only the unnatural amino acid. The stringencies of the selection steps can be varied by changing the concentration of chloramphenicol in the growth media. From the same library, the authors isolated TyrRS mutants that incorporated either *O*-methyl-tyrosine (203) or 2-naphthylalanine (204) at the amber codon. The *Mj* TyrRS mutant that incorporated *O*-methyl-tyrosine activated the unnatural surrogate 100-fold faster than towards tyrosine (203).

Residue-Specific Incorporation of Unnatural Amino Acids In vivo

While aaRSs have been evolved to strictly distinguish between natural amino acids, they do not have the same level of discrimination towards structurally similar unnatural amino acids. When a natural amino acid is depleted in the expression media, an unnatural analog may be activated by aaRS and be incorporated into proteins at positions normally occupied by its natural counterpart. Cellular uptake of nonproteinogenic building blocks as a response to the lack of natural ones is termed selective pressure incorporation. Because natural amino acids are invariably activated more rapidly than the analogs, the cellular concentration of the natural amino acid must be minimized to achieve high levels of incorporation. Low levels of the natural amino acid present during protein expression can lead to the enrichment of the amino acid in proteins. Auxotrophic strains are thus commonly used in residue-specific incorporation to cut off supply of the natural amino acid during protein expression. Natural amino acid added during cell growth are depleted via a medium shift procedure in which the cells are removed from the growth medium and resuspended in expression medium containing only the analog of interest.

A large collection of amino acid analogs have been incorporated into recombinant proteins using conventional *E. coli* hosts. A collection of the unnatural amino acids are shown in Figure 1.13. Most of these unnatural amino acids are isosteric to one of the twenty naturally occurring amino acids. For example, fluorinated analogs of proline (206), leucine (207), tryptophan (208, 209) and methionine (210-212) have been inserted into target proteins using the selective pressure method. Amino acids containing fluorinated side chains contain unique physical properties. Fluorines can be used as nonperturbing probes for protein structure and dynamics through ^{19}F NMR (213). The sensitivity of the ^{19}F signal is superior to that of ^{1}H and ^{13}C NMR due to the lack of fluorines in natural proteins. Collagen constructs containing fluoroprolines in place of hydroxyprolines are significantly more stable, mainly due the inductive effects associated with fluorine atoms (214, 215). In this thesis, we will also discuss the effects of fluorinated amino acids on the stabilities of proteins (216-219).

Amino acids containing heavy atoms have also been incorporated into proteins for phase determination in protein X-ray crystallography. Methionine analogs selenomethioine and telluromethioines (220, 221), tryptophan analog β-selenolo[3,2-b]pyrrolyl-alanine(222, 223) have all been incorporated in a

residue-specific manner. The electroactive amino acid 3-thienylalanine has been inserted into phenylalanine positions *in vivo* (224), opening the possibilities of biosynthesizing conducting polymers.

Our laboratory and others have studied the translational properties of methionine analogs extensively. Unnatural amino acids ethionine (225), norleucine (226), *S*-nitrohomocystein (227), homoallylglycine (228), homopropargylglycine (229), homoazidoalanine (230) and others are all active surrogates of MetRS in a methionine-depleted expression medium. The presence of unsaturated moieties in proteins allows metathesis (231, 232) and palladium-catalyzed coupling to be performed on proteins and biopolymers. The introduction of an azide functionality facilitates chemoselective modification of proteins through the mild Staudinger reaction (233). Kiick et al. have improved translational activities of poor methionine analogs such as *trans*-crotylglycine (234) and 2-butynyl-glycine (235) by elevating the cellular activities of MetRS. While these analogs are not incorporated using a conventional expression host, overexpression of MetRS during cell growth allowed these analogs to be activated during translation.

Relaxing the specificities of *E. coli* PheRS has allowed the incorporation of an array of analogs *in vivo*. While wild-type PheRS incorporated fPhe readily (236), phenylalanine analogs with larger *para* substituents are inactive during translation, possibly due to poor binding at the active site of PheRS. The terminal wall of PheRS is sealed off by two residues, Ala294 and Thr251 (39). Mutation of Ala294 to a glycine enlarged the binding site of the synthetase and accommodated larger phenylalanine analogs (201). Overexpression of the mutant synthetase facilitated incorporation of *p*-bromo-phenylalanine (237), *p*-iodo-phenylalanine and *p*-cyano-phenylalanine (238), etc. Mutation of both residues to glycines expanded the binding pocket further. Amino acids such as *p*-acetyl-phenylalanine and 2-naphthylalaine were readily incorporated into proteins when the A294G/T251G mutant was overexpressed (239).

This thesis describes our work on incorporating leucine analogs *in vivo*. Several unique features are associated with LeuRS/leucine. First, leucine is the most abundant amino acid in cellular proteins, occurring at a frequency of more than 9%. Second, leucine is one of the most hydrophobic amino acids, and is frequently found in the core of protein assemblies and plays an important role in protein folding. Third, the substrate specificities of LeuRS towards structurally similar amino acids, and the mechanism in which it proofreads are not well understood. The thesis is divided into four chapters. In the second chapter,

we will examined the effects of replacing leucine with trifluoroleucine, on the structure and stability of a leucine zipper protein GCN4-p1. The fluorinated zipper was synthesized chemically in this proof-of-concept experiment. We discovered that the thermal and chemical stabilities of fluorinated leucine zippers were significantly more enhanced over those of the wild type protein. We propose therefore that fluorinating protein cores may be a general method of stabilizing protein structures. Chapter 3 discusses the results of biosynthesizing fluorinated leucine zippers. We explored LeuRS substrate specificities toward trifluoroleucine and hexafluoroleucine. While trifluoroleucine supported protein biosynthesis under normal cellular conditions, the more sluggish analog hexafluoroleucine required elevation of intracellular concentration of LeuRS. Target proteins containing the fluorinated amino acids exhibited significant gain in chemical and thermal stabilities. The incorporation of hexafluoroleucine enhanced protein stabilities furthur. Successful demonstration of biosynthesis of fluorinated proteins increases the scope of fluorination to include larger, recombinant proteins. In the fourth chapter, we studied the proofreading mechanisms of LeuRS. Our objective is to relax the substrate specificities of LeuRS by disrupting its editing function. We generated mutant forms of LeuRS based on X-ray crystal structure and demonstrated in vitro that these mutants were defective in editing towards isoleucine and valine. When one of the mutants was overexpressed in vivo, previously translationally inert amino acids such as norvaline, norleucine, allylglycine were all readily incorporated into recombinant protein in vivo. The presence of the amino acids was verified using amino acid analysis and mass spectrometry. Our results suggest that disabling the synthetase's abilities to proofread is a powerful method of enlarging the pool of amino acids that can be used in protein engineering and biomaterials biosynthesis.

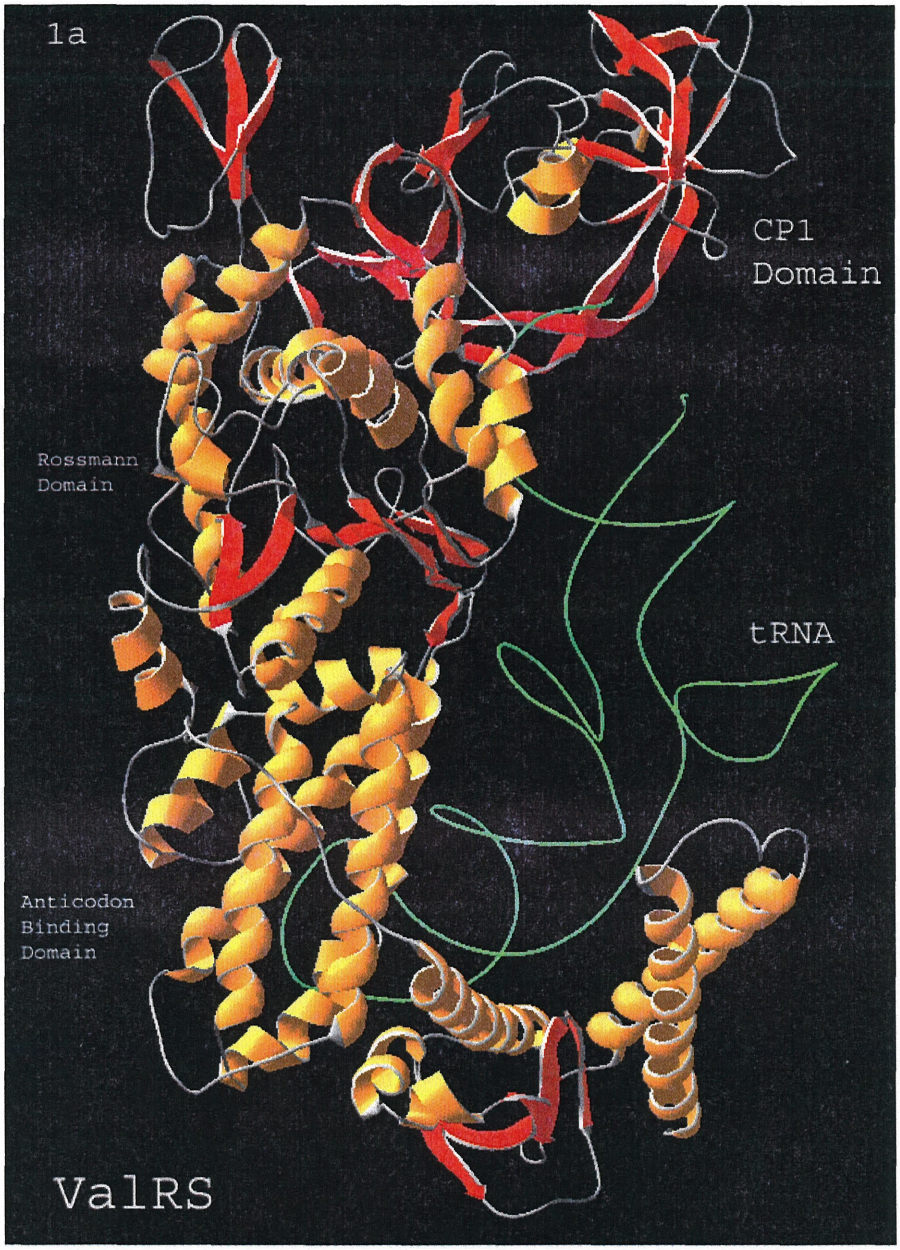


Figure 1.1: Crystal structure of *T. thermophilus* ValRS complexed with $tRNA^{Val}$ (37). The tRNA molecule is shown as a green line. The β-strands are shown in red and α-helices are shown in yellow. The Rossmann domain is positioned at the center of the molecule. The anticodon binding domain consists mainly of α-helices. The CP1 domain consists mainly of β-strands. In this representation, the acceptor stem of the tRNA extends into the editing site of the enzyme.

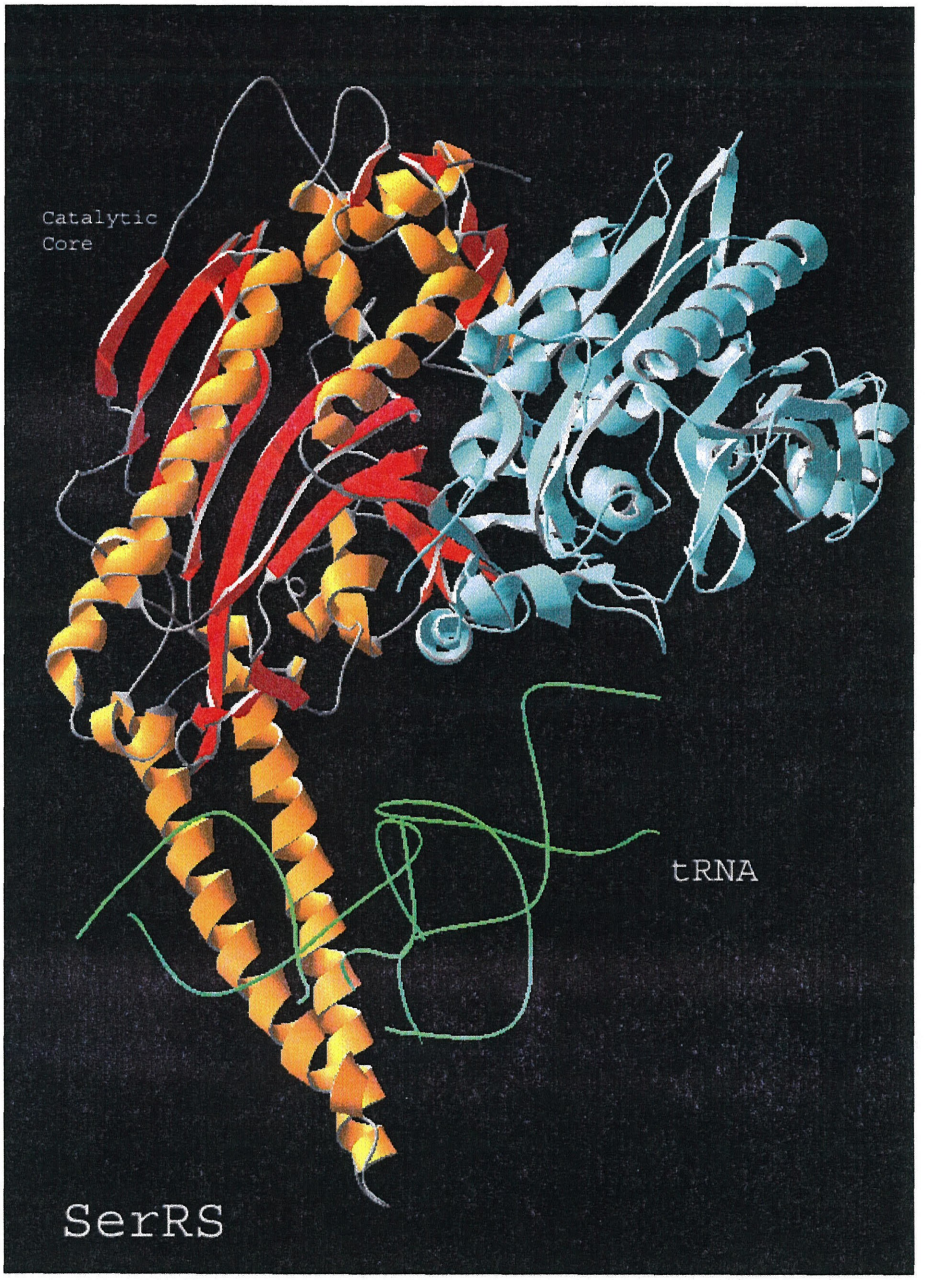
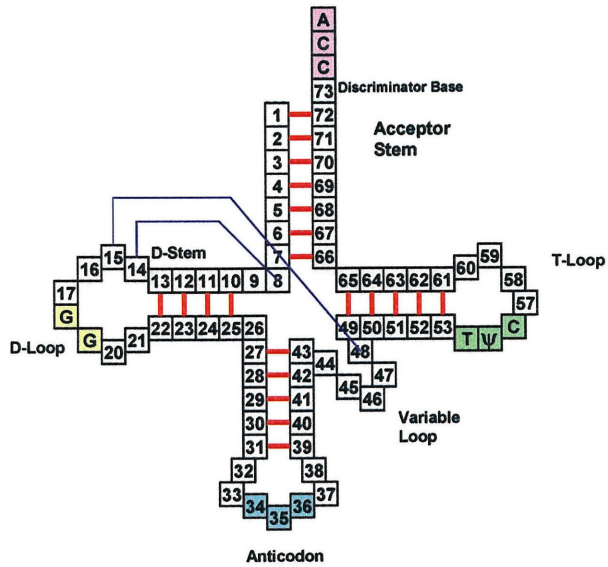
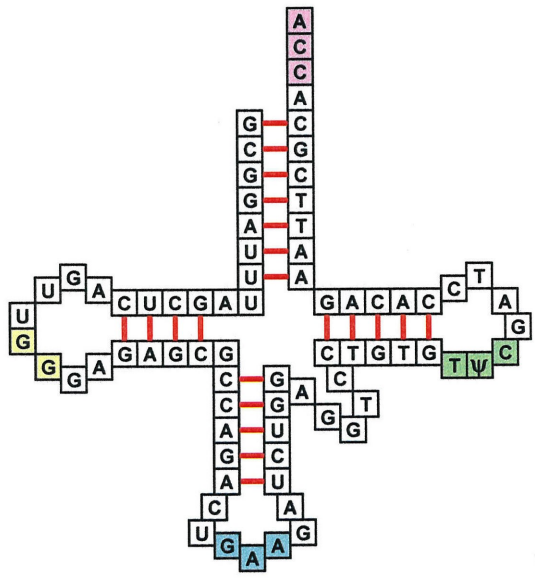


Figure 1.2. The crystal structure of *T. thermophilus* SerRS complexed with $tRNA^{Ser}$ (27, 31, 33). SerRS forms a dimer and one subunitis shown in cyan. The other subunit is shown in yellow and red. β-strands are represented in red and α-helices are shown in yellow. The seven-stranded β-sheet catalytic core is labeled. The *N*-terminal coiled-coil is shown only for one of the monomers. The long antiparallel helices interact with the anticodon loop and the variable arm of $tRNA^{Ser}$. Only part of the tRNA molecule is visible.





Yeast tRNA^{Phe}

Figure 1.3. Top: general features of a tRNA. Red lines indicate predicted and observed secondary base pairings. Blue lines represent tertiary base pairings. N8-N14 and N15-N48 are important for the L-shaped structure seen in Figure 4. D-loop nucleotides interact with T-loop nucleotides extensively. Conserved residues among all tRNAs are shown as letters. T: ribothymidine; ψ: pseudouridine. Bottom: yeast tRNA^{Phe} studied by Holbrook et al. (80) and Sampson et al. (92).

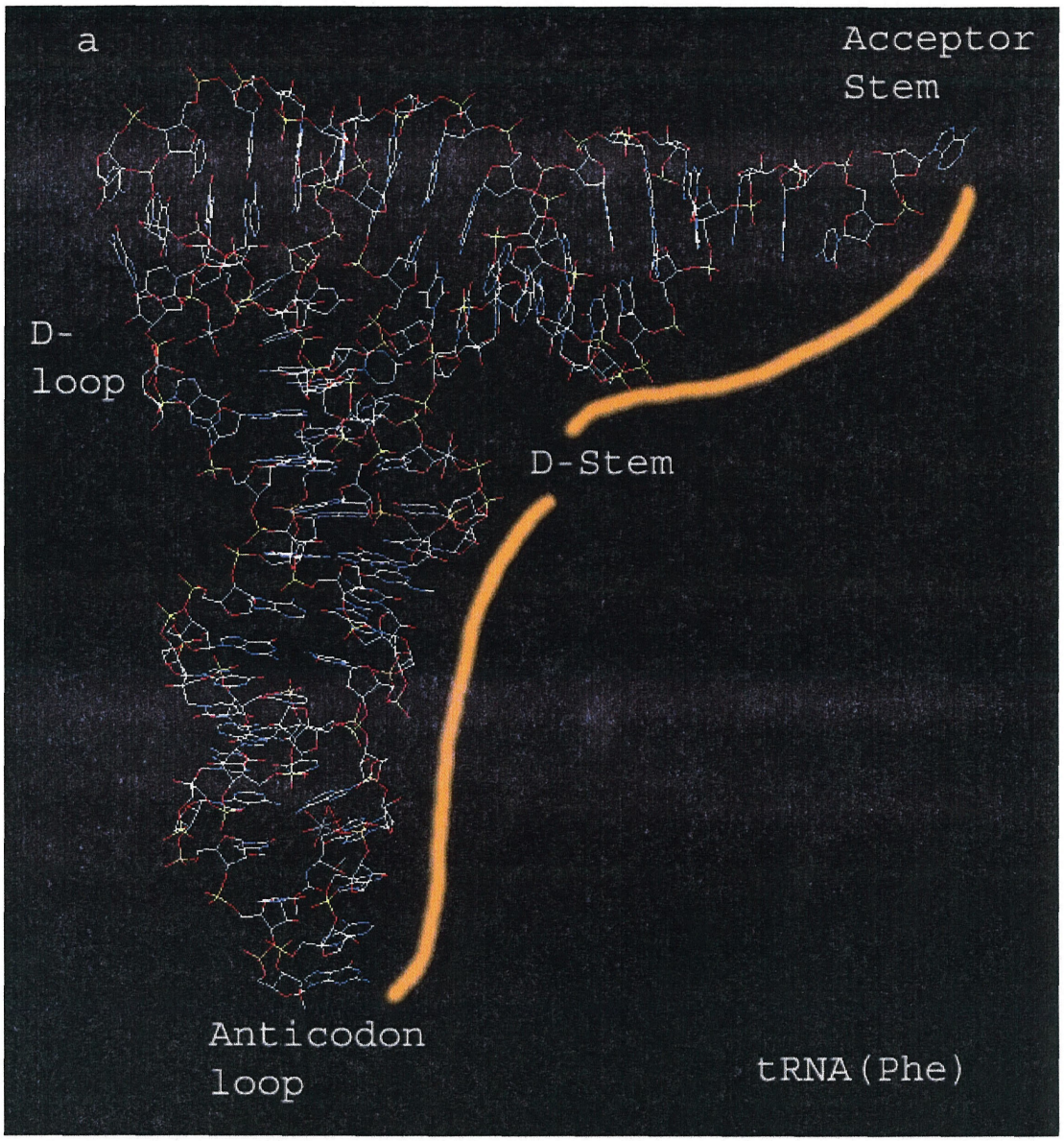


Figure 1.4a. The crystal structure of yeast $tRNA^{Phe}$ (80). The acceptor stem, D-stem, and anticodon loop lie on one surface of the L-shaped tRNA. The yellow line indicates where most of the tRNA/aaRS interactions are located (82). The L-shaped structure is stabilized by tertiary interactions between the D-loop and the T-loop and N15-N48. Carbon atoms are shown in white, nitrogens in blue, phosphorus in yellow, and oxygens in red.

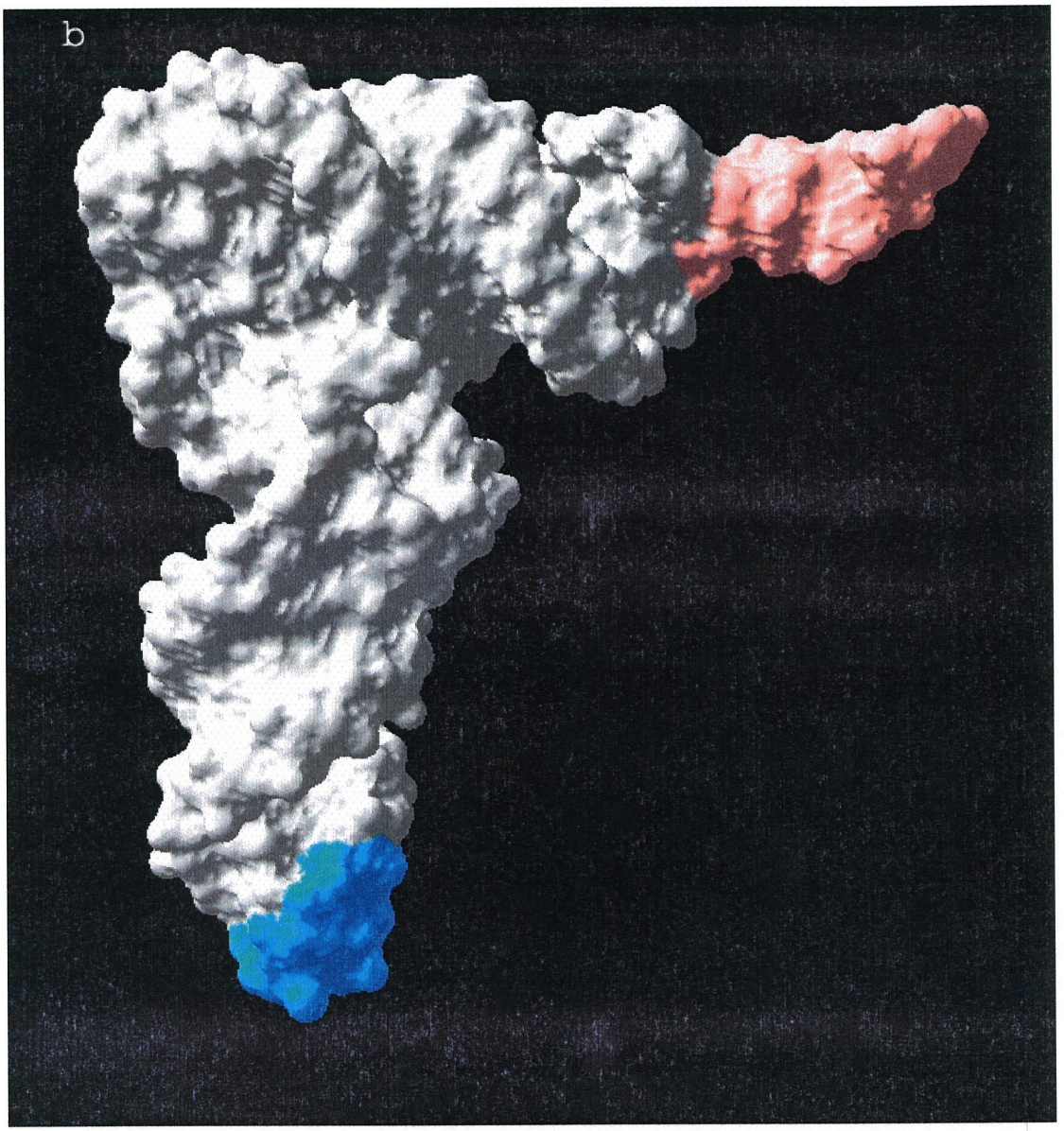


Figure 1.4b. Surface model of the tRNA molecule shown in Figure 4a. The anticodon loop is shown in light blue and the acceptor stem is shown in pink.

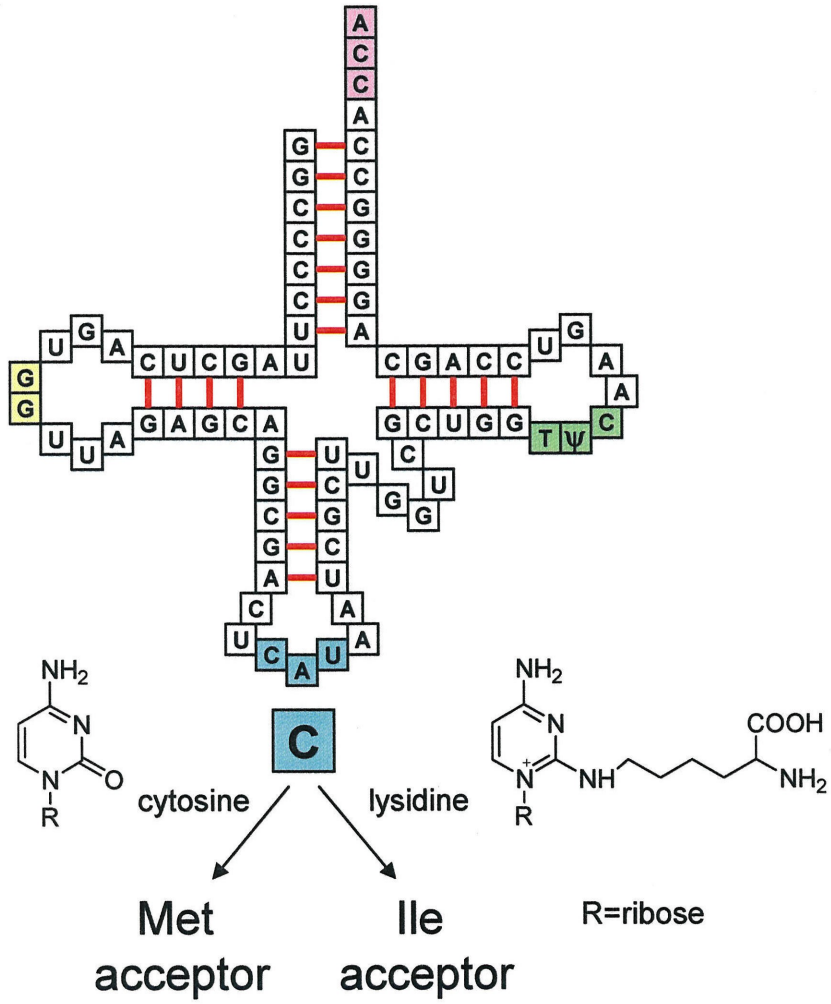


Figure 1.5. The effect of base modification on *E. coli* $tRNA^{Ile}_2$ identity (96, 97). The base of C34 is modified to a lysidine and the modified tRNA decodes the codon AUA with isoleucine. However, if the base is left as a cytosine *in vitro*, this tRNA becomes a methionine acceptor. Modification of C34 serves both as a positive identity element for IleRS and as a negative determinant against MetRS.

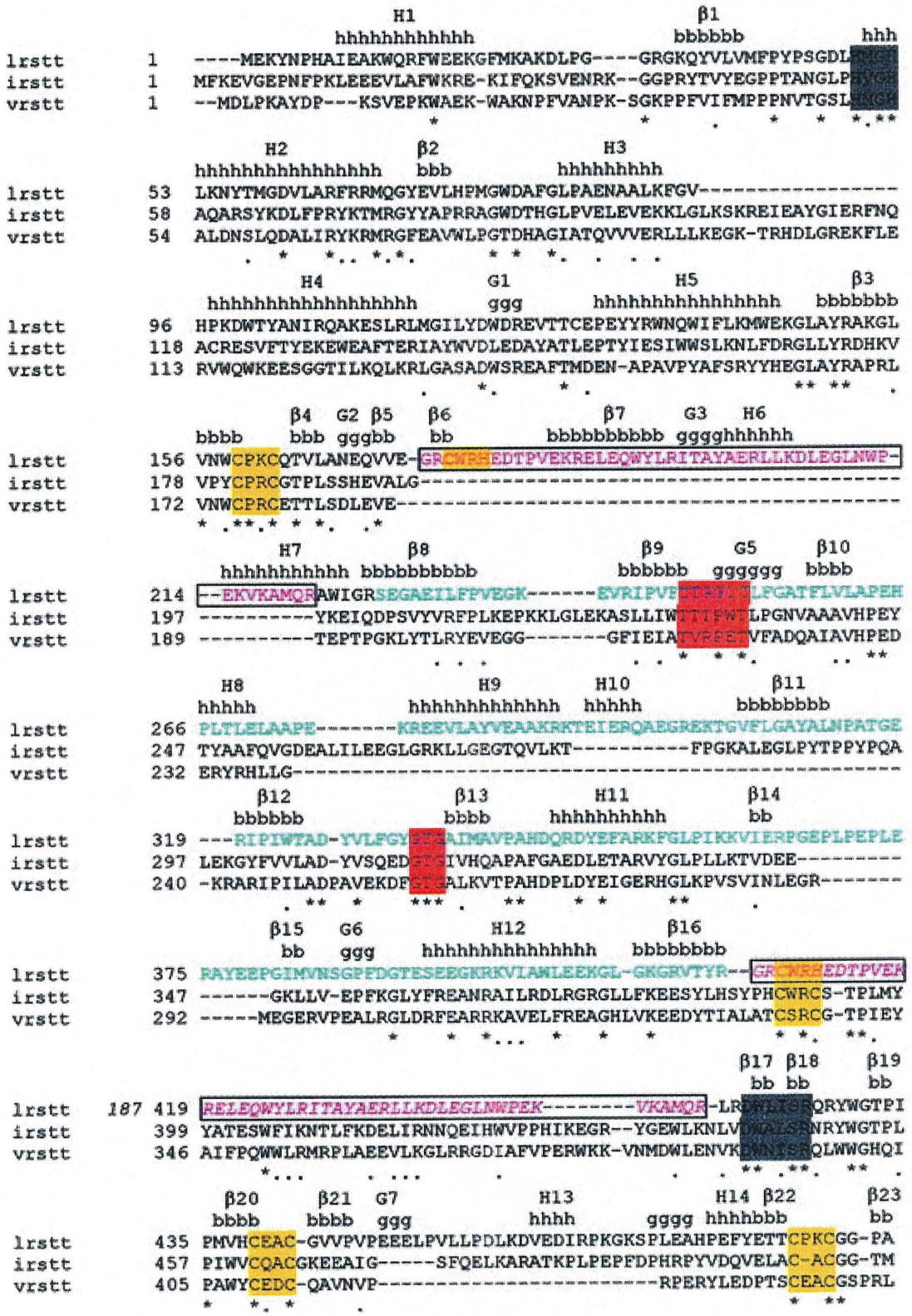


Figure 1.6. Continues on the next page

Figure 1.6. Sequence alignments of *T. thermophilus* LeuRS, IleRS and ValRS. Results are adopted from Cusack et al. (35). Secondary structures are labeled on top of the sequences (h: α-helix, b: β-sheet). Identical residues are indicated by an asterisk, conservative mutations are indicated with a dot. The leucyl-specific domain is shown in blue. HxGH and KMSKS motifs are shaded in dark blue. The editing domain (CP1) is shown in cyan. Conserved editing domain residues are shaded in red. Shown twice in pink is the 50-residue zinc binding domain. The first occurrence is at the observed location, which is before the CP1 insertion. The italicized, boxed region, is where similar sequences are found in IleRS and ValRS. Zinc binding ligands are shaded in yellow.

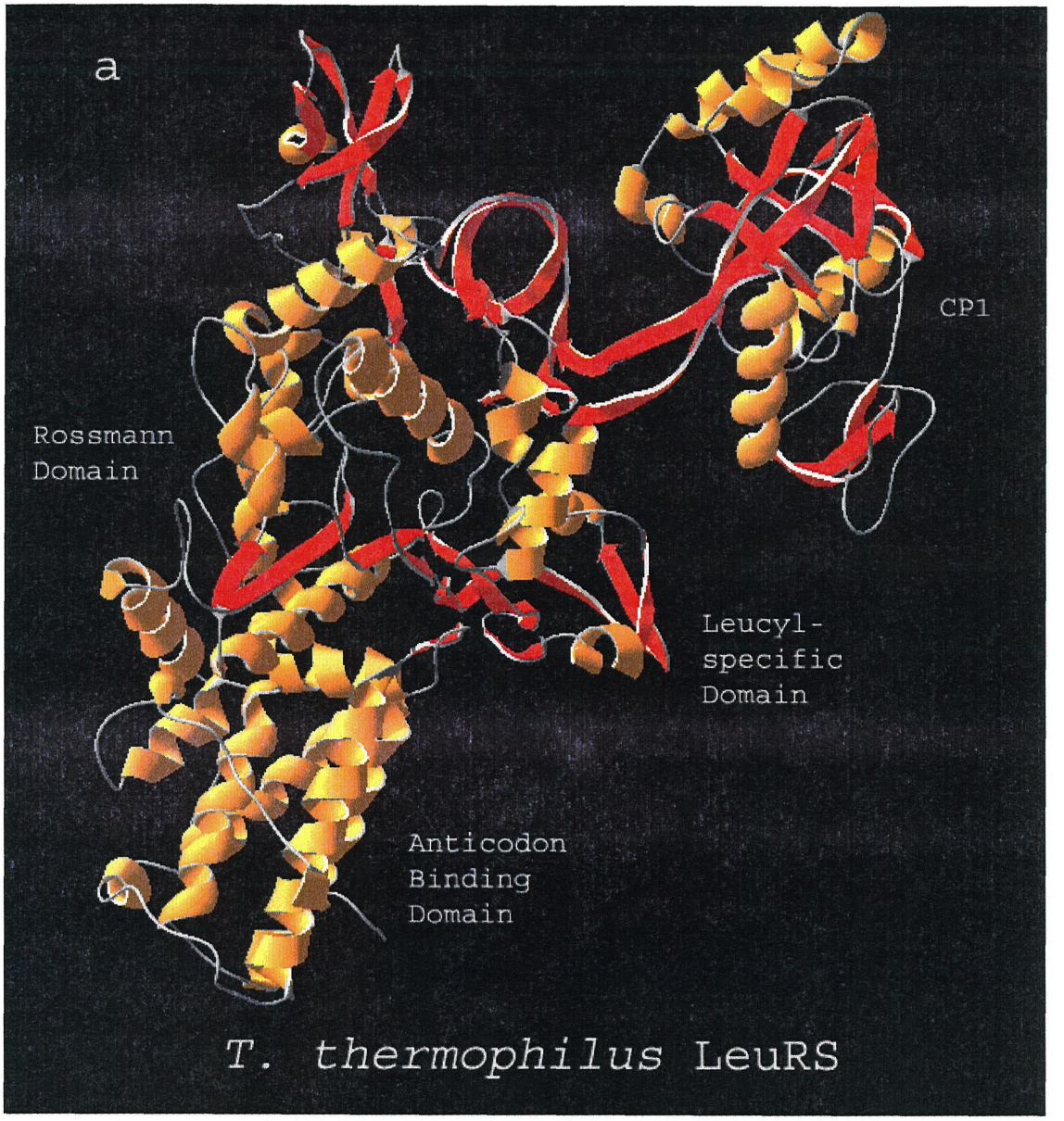


Figure 1.7a. Crystal structure of the *T. thermophilus* leucyl-tRNA synthetase (35). The different domains are labeled . β-strands are shown in red and α-helices are shown in yellow. The Rossmann domain is located at the center of the enzyme. The CP1 and leucyl-specific domains are connected to the Rossmann domain via flexible β-ribbons.

Figure 1.7b. The same structure as shown in Figure 7a, except the separate domains are shown in different colors for clarity. The synthetic active site is shown in grey. The leucyl-adenylate analog LeuAMS is shown in spacefilled representation. The C-terminal anticodon binding region is shown in pink. The leucyl-specific domain not found in ValRS and IleRS is shown in cyan. The editing domain is shown in yellow. The first zinc binding domain is shown in green. The editing domain inserts after the zinc binding domain in prokaryotic LeuRS. For IleRS and ValRS, the zinc binding domain follows the editing domain in the primary sequence.

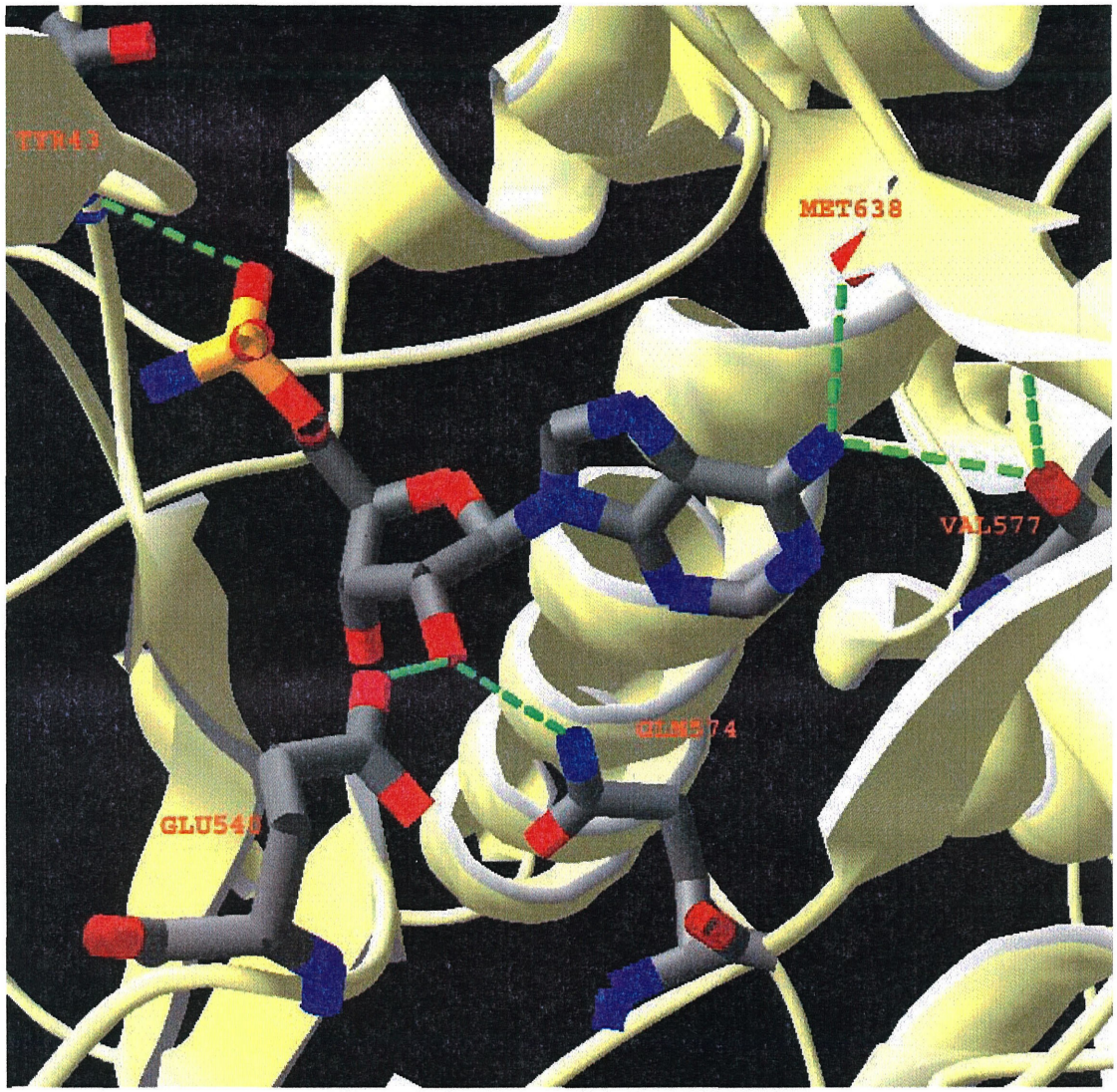


Figure 1.8. Active site of *T. thermophilus* LeuRS (AMP binding portion). The AMP molecule is shown in the middle. Carbons are shown in grey, oxygens in red, nitrogens in blue, and phosphorus in yellow. The adenine ring is fixed by hydrogen bonds from the backbones of Met638 (part of VMSKS), Val577 and other residues. The ribose ring is fixed by side chains of Gln574 and Glu543. Figure 9 shows the binding pocket for the leucine side chain.

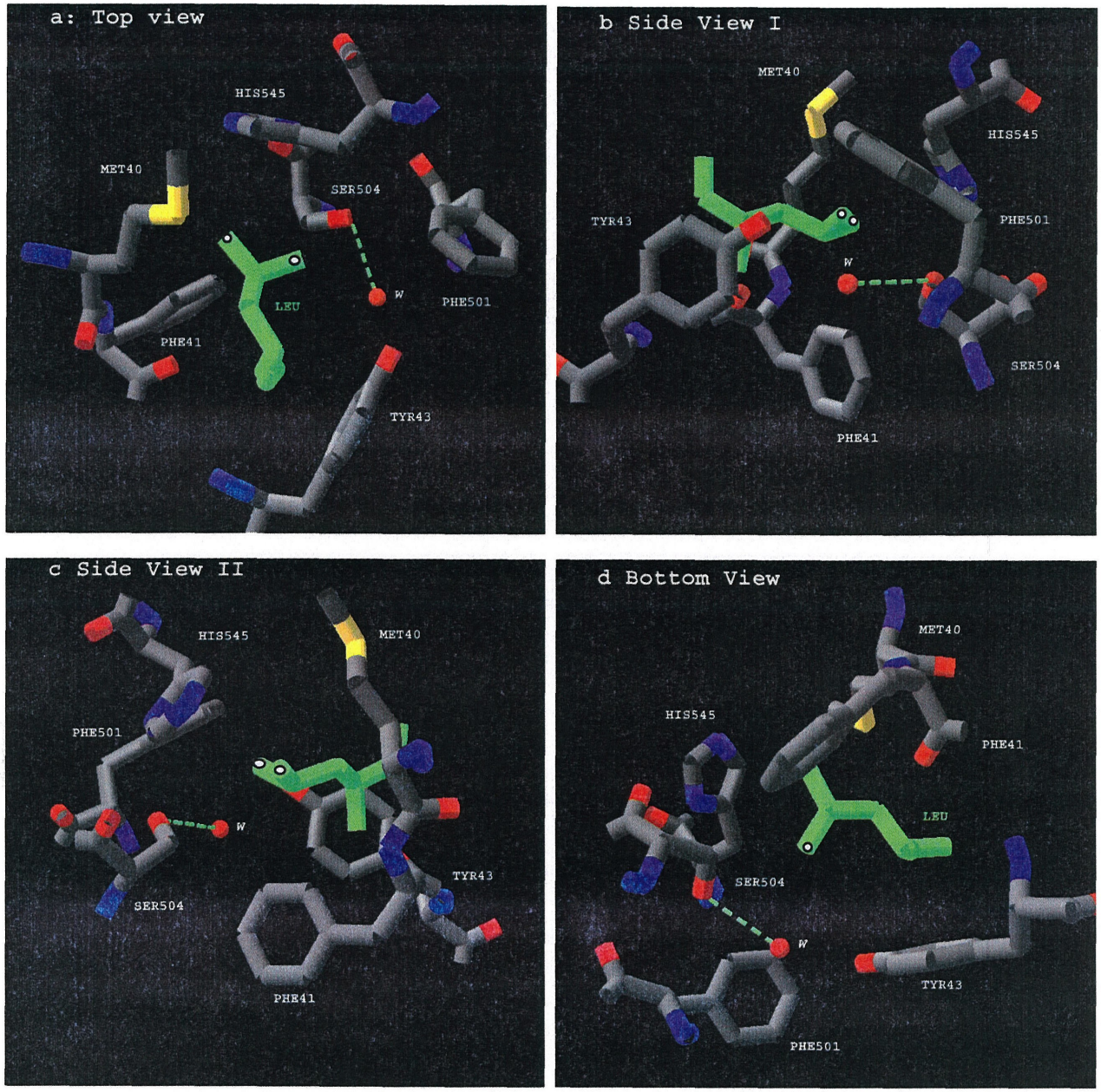


Figure 1.9. Active site of *T. thermophilus* LeuRS (leucine side chain binding region). Four different views are shown. Leucine is shown in green for clarity. The terminal carbon atoms of leucine are indicated with a white dot. Carbon atoms are in grey; oxygens in red, nitrogens in blue and sulfurs in yellow. One water molecule (W) is found in the active site, hydrogen bonded to Ser504 and Asp80 (not shown). The terminal wall of the active site is sealed off with His545. The phenolic –OH hydrogen bonds to the α-amino moiety of the amino acid.

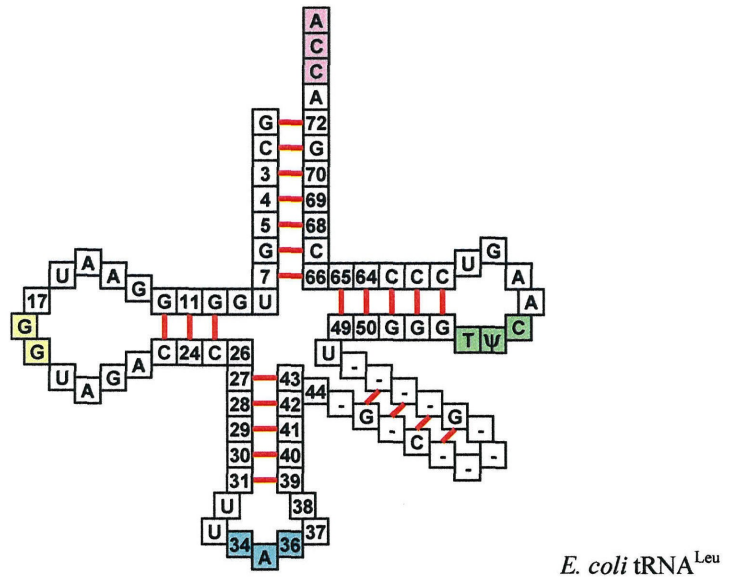


Figure 1.10. Comparison of the sequences and structures of *E. coli* $tRNA^{Leu}$ and $tRNA^{Ser}$. The conserved residues among all the isoacceptors are shown in letters. Nonconserved bases are shown as numbers. The major differences among the two sets of isoacceptors are 1) the discriminator nucleotide; 2) the position of the GG dinucleotide in the D-loop; 3) tertiary structure determinants N15-N48; and 4) the length of the variable loops. In *E. coli*, neither LeuRS nor SerRS relies on the anticodon loop for recognition (149). In yeast, LeuRS uses A35 as a major identity element and the bases at N15-N48 are less important to tRNA identity compared to *E. coli* LeuRS (151). Note that numbering in the D-loop is interrupted so that the rest of the bases in tRNA can be numbered consistently as in other tRNAs.

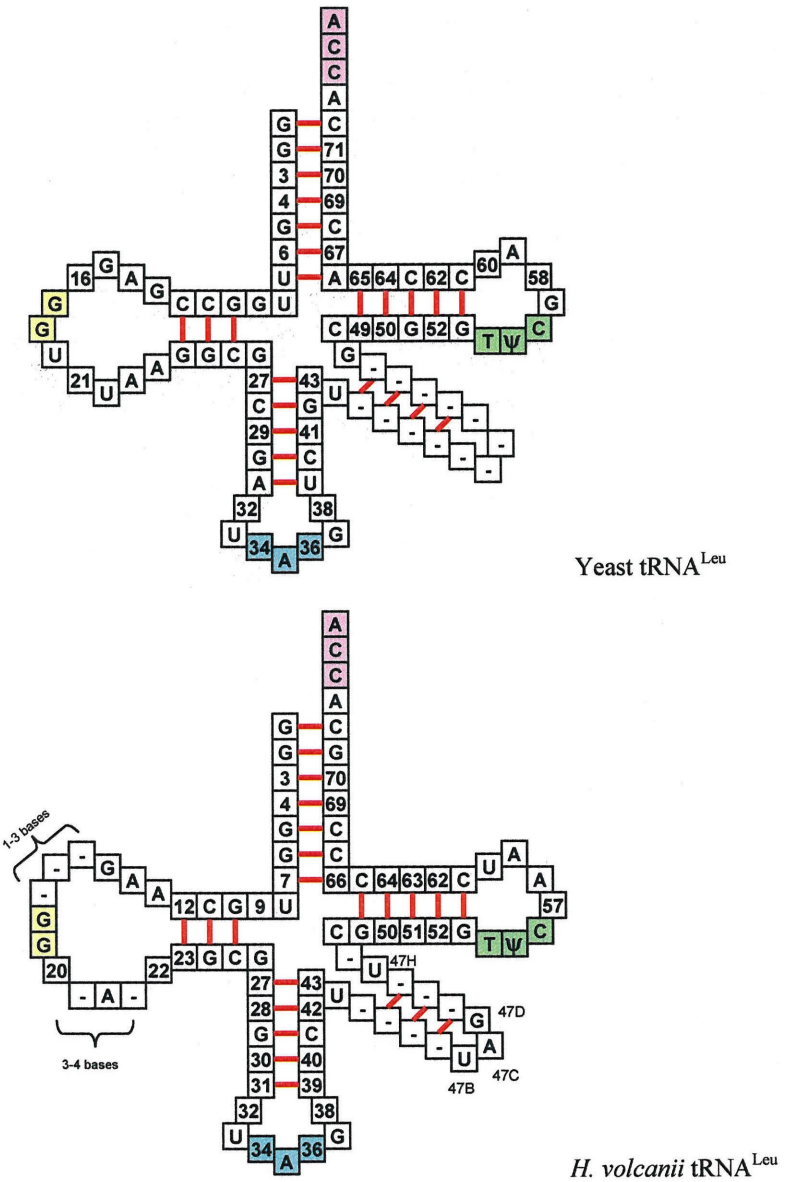


Figure 1.11. Conserved bases in yeast and *H. volcanii* tRNA^{Leu}. The structure for *E. coli* isoacceptors is shown in Figure 10. Yeast LeuRS uses A35 as a major identity element (151). *H. volcanii* LeuRS relies on conserved nucleotides at both the tip and base of the variable loop for recognition (153). All tRNAs contain A73 at the discriminator position. The D-loop of *H. volcanii* tRNA^{Leu} is more variable than those of *E. coli* and yeast.

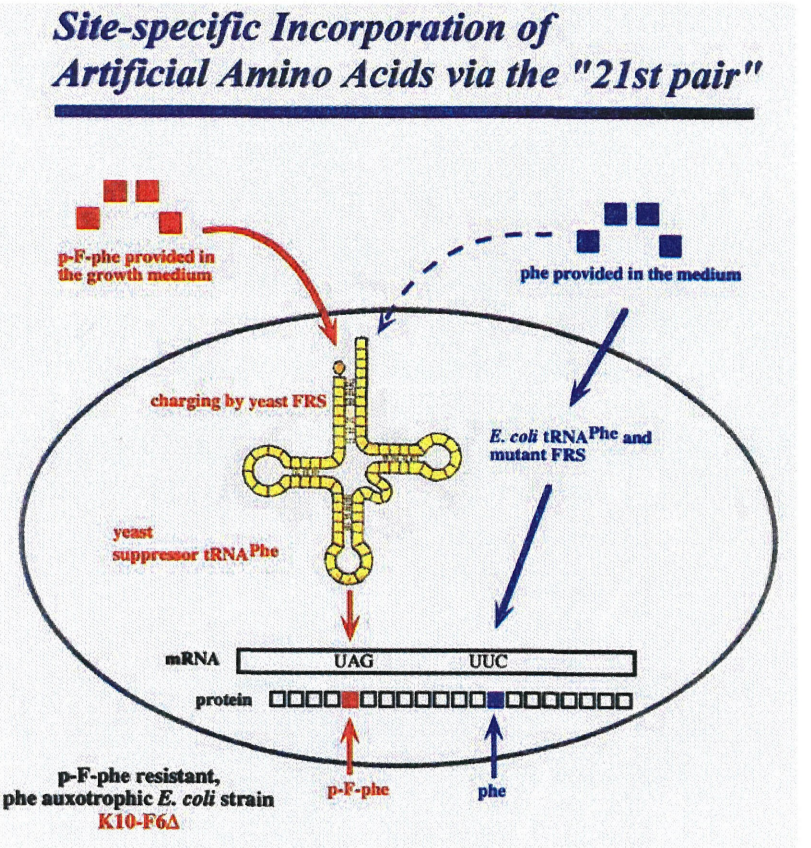


Figure 1.12. Site specific incorporation of p-F-phenylalanine *in vivo*. *E. coli* strain K10-F6Δ is used as the expression strain. This strain is resistant to fPhe because a chromosomal mutation is present within the PheRS gene. The mutation blocks binding of fPhe to host PheRS. Yeast PheRS and $tRNA^{Phe}_{CUA}$, which are orthogonal to the pair present in *E. coli*, are introduced into the host strain. The wild type yeast PheRS recognizes both Phe and fPhe. An amber suppressor codon UAG is inserted into the target gene. Phenylalanine codons UUC are unmodified. During translation, *E. coli* machinery incorporates phenylalanine at UUC positions as it normally would. The amber codon is translated by the suppressor $tRNA^{Phe}_{CUA}$. Because fPhe is supplemented at elevated concentrations compared to that of phenylalanine during protein translation, a majority (>75%) of amino acids appended to the suppressor tRNA by yeast PheRS are fPhe. Some of the tRNAs are unavoidably aminoacylated with phenylalanine. In a subsequent experiment, Wang and coworkers mutated the yeast PheRS so that it preferentially activated 2-naphthylalanine over phenylalanine. As a result, the incorporation of 2-naphthylalanine at amber codon position is nearly quantitative. Since *E. coli* PheRS cannot activate the bulkier amino acid, the phenylalanine positions were not substituted. Image adopted from Ref. 200.

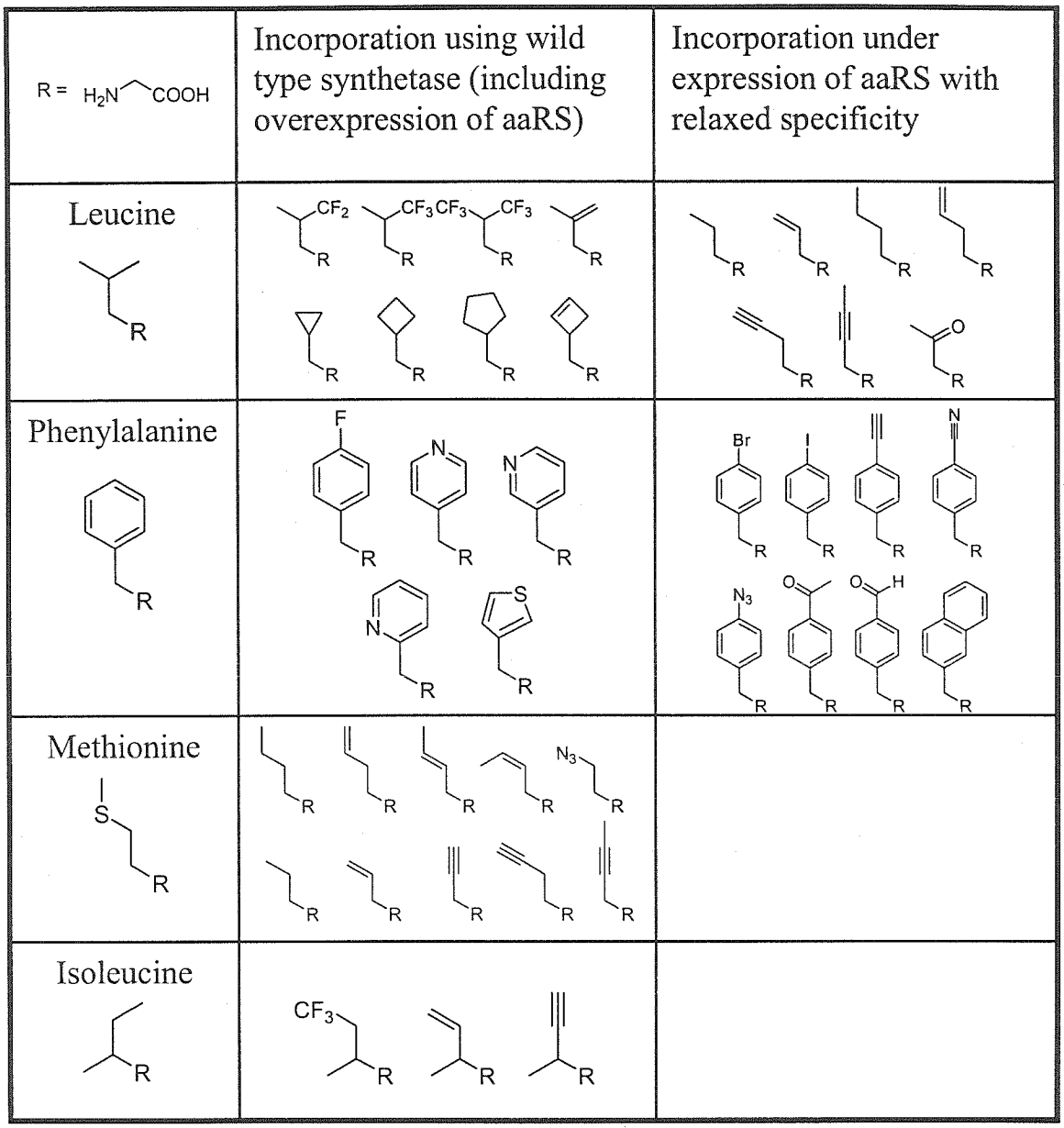


Figure 1.13. Analogs of natural amino acids investigated by our laboratory. Analogs that do not require overexpression of an aaRS mutant is shown in column 2. Analogs that require overexpression of aaRS mutant with relaxed substrate specificities are shown in column 3. This thesis will cover the leucine analogs shown above. Chapter three discusses the incorporation of trifluoroleucine and hexafluoroleucine. Chapter 4 discusses the use of an editing impaired LeuRS towards the incorporation of analogs shown in column 3.

1.6 References

1. Ibba, M., Curnow, A. W., and Soll, D. (1997) *Trends Biochem Sci 22*, 39-42.
2. Schimmel, P. R., and Soll, D. (1979) *Annu Rev Biochem 48*, 601-48.
3. Schimmel, P. (1987) *Annu Rev Biochem 56*, 125-58.
4. Carter, C. W., Jr. (1993) *Annu Rev Biochem 62*, 715-48.
5. Ibba, M., and Soll, D. (2000) *Annu Rev Biochem 69*, 617-50.
6. Schimmel, P., Tao, J., and Hill, J. (1998) *Faseb J 12*, 1599-609.
7. Wolfenden, R., Rammler, D. H., and Lipmann, F. (1964) *Biochemistry 3*, 329-338.
8. Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990) *Nature 347*, 203-6.
9. Arnez, J. G., and Moras, D. (1997) *Trends Biochem Sci 22*, 211-6.
10. Webster, T., Tsai, H., Kula, M., Mackie, G. A., and Schimmel, P. (1984) *Science 226*, 1315-7.
11. Webster, T. A., Lathrop, R. H., and Smith, T. F. (1987) *Biochemistry 26*, 6950-7.
12. First, E. A., and Fersht, A. R. (1995) *Biochemistry 34*, 5030-43.
13. Eriani, G., Prevost, G., Kern, D., Vincendon, P., Dirheimer, G., and Gangloff, J. (1991) *Eur J Biochem 200*, 337-43.
14. Fraser, T. H., and Rich, A. (1975) *Proc Natl Acad Sci USA 72*, 3044-8.
15. Sprinzl, M., and Cramer, F. (1975) *Proc Natl Acad Sci USA 72*, 3049-53.
16. Barstow, D. A., Clarke, A. R., Chia, W. N., Wigley, D., Sharman, A. F., Holbrook, J. J., Atkinson, T., and Minton, N. P. (1986) *Gene 46*, 47-55.
17. Hohmann, S., and Thevelein, J. M. (1992) *Gene 120*, 43-9.
18. Jasin, M., Regan, L., and Schimmel, P. (1983) *Nature 306*, 441-7.
19. Jasin, M., and Schimmel, P. (1984) *J Bacteriol 159*, 783-6.
20. Carter, P., Bedouelle, H., and Winter, G. (1986) *Proc Natl Acad Sci USA 83*, 1189-92.
21. Bedouelle, H., and Winter, G. (1986) *Nature 320*, 371-3.
22. Gaillard, C., and Bedouelle, H. (2001) *Biochemistry 40*, 7192-9.
23. Aberg, A., Yaremchuk, A., Tukalo, M., Rasmussen, B., and Cusack, S. (1997) *Biochemistry 36*, 3084-94.

24. Arnez, J. G., Augustine, J. G., Moras, D., and Francklyn, C. S. (1997) *Proc Natl Acad Sci U S A 94*, 7144-9.

25. Belrhali, H., Yaremchuk, A., Tukalo, M., Larsen, K., Berthet-Colominas, C., Leberman, R., Beijer, B., Sproat, B., Als-Nielsen, J., Grubel, G., and et al. (1994) *Science 263*, 1432-6.

26. Berthet-Colominas, C., Seignovert, L., Hartlein, M., Grotli, M., Cusack, S., and Leberman, R. (1998) *Embo J 17*, 2947-60.

27. Biou, V., Yaremchuk, A., Tukalo, M., and Cusack, S. (1994) *Science 263*, 1404-10.

28. Brick, P., Bhat, T. N., and Blow, D. M. (1989) *J Mol Biol 208*, 83-98.

29. Brunie, S., Zelwer, C., and Risler, J. L. (1990) *J Mol Biol 216*, 411-24.

30. Cavarelli, J., Delagoutte, B., Eriani, G., Gangloff, J., and Moras, D. (1998) *Embo J 17*, 5438-48.

31. Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N., and Leberman, R. (1990) *Nature 347*, 249-55.

32. Cusack, S., Yaremchuk, A., and Tukalo, M. (1996) *Embo J 15*, 6321-34.

33. Cusack, S., Yaremchuk, A., and Tukalo, M. (1996) *Embo J 15*, 2834-42.

34. Cusack, S., Yaremchuk, A., Krikliviy, I., and Tukalo, M. (1998) *Structure 6*, 101-8.

35. Cusack, S., Yaremchuk, A., and Tukalo, M. (2000) *Embo J 19*, 2351-61.

36. Doublie, S., Bricogne, G., Gilmore, C., and Carter, C. W., Jr. (1995) *Structure 3*, 17-31.

37. Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D. G., and Yokoyama, S. (2000) *Cell 103*, 793-803.

38. Logan, D. T., Mazauric, M. H., Kern, D., and Moras, D. (1995) *Embo J 14*, 4156-67.

39. Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M., and Safro, M. G. (1995) *Nat Struct Biol 2*, 537-47.

40. Nureki, O., Vassylyev, D. G., Katayanagi, K., Shimizu, T., Sekine, S., Kigawa, T., Miyazawa, T., Yokoyama, S., and Morikawa, K. (1995) *Science 267*, 1958-65.

41. Nureki, O., Vassylyev, D. G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T. L., Schimmel, P., and Yokoyama, S. (1998) *Science 280*, 578-82.

42. Onesti, S., Miller, A. D., and Brick, P. (1995) *Structure 3*, 163-76.

43. Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989) *Science 246*, 1135-42.

44. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., and Moras, D. (1991) *Science 252*, 1682-9.

45. Sankaranarayanan, R., Dock-Bregeon, A. C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B., and Moras, D. (1999) *Cell 97*, 371-81.

46. Silvian, L. F., Wang, J., and Steitz, T. A. (1999) *Science 285*, 1074-7.

47. Irwin, M. J., Nyborg, J., Reid, B. R., and Blow, D. M. (1976) *J Mol Biol 105*, 577-86.

48. Newberry, K. J., Kohn, J., Hou, Y. M., and Perona, J. J. (1999) *Acta Crystallogr D Biol Crystallogr 55*, 1046-7.

49. Rao, S. T., and Rossmann, M. G. (1973) *J Mol Biol 76*, 241-56.

50. Wells, T. N., Ho, C. K., and Fersht, A. R. (1986) *Biochemistry 25*, 6603-8.

51. Fersht, A. R., Leatherbarrow, R. J., and Wells, T. N. (1987) *Biochemistry 26*, 6030-8.

52. Wells, T. N., and Fersht, A. R. (1989) *Biochemistry 28*, 9201-9.

53. Wells, T. N., Knill-Jones, J. W., Gray, T. E., and Fersht, A. R. (1991) *Biochemistry 30*, 5151-6.

54. Hountondji, C., Beauvallet, C., Dessen, P., Hoang-Naudin, C., Schmitter, J. M., Pernollet, J. C., and Blanquet, S. (2000) *Eur J Biochem 267*, 4789-98.

55. Belrhali, H., Yaremchuk, A., Tukalo, M., Berthet-Colominas, C., Rasmussen, B., Bosecke, P., Diat, O., and Cusack, S. (1995) *Structure 3*, 341-52.

56. Fersht, A. R., and Dingwall, C. (1979) *Biochemistry 18*, 1245-9.

57. Cusack, S. (1993) *Biochimie 75*, 1077-81.

58. Cusack, S., Hartlein, M., and Leberman, R. (1991) *Nucleic Acids Res 19*, 3489-98.

59. Cavarelli, J., Rees, B., Ruff, M., Thierry, J. C., and Moras, D. (1993) *Nature 362*, 181-4.

60. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S., and Venter, J. C. (1996) *Science 273*, 1058-73.

61. Curnow, A. W., Ibba, M., and Soll, D. (1996) *Nature 382*, 589-90.

62. Becker, H. D., and Kern, D. (1998) *Proc Natl Acad Sci USA 95*, 12832-7.

63. Schon, A., Kannangara, C. G., Gough, S., and Soll, D. (1988) *Nature 331*, 187-90.

64. Ibba, M., Morgan, S., Curnow, A. W., Pridmore, D. R., Vothknecht, U. C., Gardner, W., Lin, W., Woese, C. R., and Soll, D. (1997) *Science 278*, 1119-22.

65. Ibba, M., Losey, H. C., Kawarabayasi, Y., Kikuchi, H., Bunjun, S., and Soll, D. (1999) *Proc Natl Acad Sci U S A 96*, 418-23.

66. Stathopoulos, C., Li, T., Longman, R., Vothknecht, U. C., Becker, H. D., Ibba, M., and Soll, D. (2000) *Science 287*, 479-82.

67. Lipman, R. S., Sowers, K. R., and Hou, Y. M. (2000) *Biochemistry 39*, 7792-8.

68. Martinis, S. A., Plateau, P., Cavarelli, J., and Florentz, C. (1999) *Embo J 18*, 4591-6.

69. Putney, S. D., and Schimmel, P. (1981) *Nature 291*, 632-5.

70. Lechler, A., and Kreutzer, R. (1998) *J Mol Biol 278*, 897-901.

71. Myers, C. A., Kuhla, B., Cusack, S., and Lambowitz, A. M. (2002) *Proc Natl Acad Sci U S A 19*, 19.

72. Wakasugi, K., and Schimmel, P. (1999) *Science 284*, 147-51.

73. Lund, E., and Dahlberg, J. E. (1998) *Science 282*, 2082-5.

74. Wakasugi, K., and Schimmel, P. (1999) *J Biol Chem 274*, 23155-9.

75. Normanly, J., and Abelson, J. (1989) *Annu Rev Biochem 58*, 1029-49.

76. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) *Nucleic Acids Res 26*, 148-53.

77. Giege, R., Sissler, M., and Florentz, C. (1998) *Nucleic Acids Res 26*, 5017-35.

78. Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R., and Rutter, W. J. (1978) *Proc Natl Acad Sci U S A 75*, 190-4.

79. Bull, P., Thorikay, M., Moenne, A., Wilkens, M., Sanchez, H., Valenzuela, P., and Venegas, A. (1987) *DNA 6*, 353-62.

80. Holbrook, S. R., Sussman, J. L., Warrant, R. W., and Kim, S. H. (1978) *J Mol Biol 123*, 631-60.

81. Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M., and Kim, S. H. (1978) *J Mol Biol 123*, 607-30.

82. Rich, A., and Schimmel, P. R. (1977) *Nucleic Acids Res 4*, 1649-65.

83. Schulman, L. H., and Pelka, H. (1977) *Biochemistry 16*, 4256-65.

84. Dock-Bregeon, A. C., Westhof, E., Giege, R., and Moras, D. (1989) *J Mol Biol 206*, 707-22.

85. Normanly, J., Ogden, R. C., Horvath, S. J., and Abelson, J. (1986) *Nature 321*, 213-9.

86. Sampson, J. R., and Uhlenbeck, O. C. (1988) *Proc Natl Acad Sci U S A 85*, 1033-7.

87. Hou, Y. M., and Schimmel, P. (1988) *Nature 333*, 140-5.

88. Behlen, L. S., Sampson, J. R., DiRenzo, A. B., and Uhlenbeck, O. C. (1990) *Biochemistry 29*, 2515-23.

89. Milligan, J. F., and Uhlenbeck, O. C. (1989) *Methods Enzymol 180*, 51-62.

90. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res 15*, 8783-98.

91. Hall, K. B., Sampson, J. R., Uhlenbeck, O. C., and Redfield, A. G. (1989) *Biochemistry 28*, 5794-801.

92. Sampson, J. R., DiRenzo, A. B., Behlen, L. S., and Uhlenbeck, O. C. (1989) *Science 243*, 1363-6.

93. Schulman, L. H., and Pelka, H. (1988) *Science 242*, 765-8.

94. Jahn, M., Rogers, M. J., and Soll, D. (1991) *Nature 352*, 258-60.

95. Perret, V., Garcia, A., Grosjean, H., Ebel, J. P., Florentz, C., and Giege, R. (1990) *Nature 344*, 787-9.

96. Muramatsu, T., Yokoyama, S., Horie, N., Matsuda, A., Ueda, T., Yamaizumi, Z., Kuchino, Y., Nishimura, S., and Miyazawa, T. (1988) *J Biol Chem 263*, 9261-7.

97. Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, T., and Yokoyama, S. (1988) *Nature 336*, 179-81.

98. Putz, J., Florentz, C., Benseler, F., and Giege, R. (1994) *Nat Struct Biol 1*, 580-2.

99. Brutlag, D., and Kornberg, A. (1972) *J Biol Chem 247*, 241-8.

100. Jakubowski, H., and Goldman, E. (1992) *Microbiol Rev 56*, 412-29.

101. Edelmann, P., and Gallant, J. (1977) *Cell 10*, 131-7.

102. Loftfield, R. B., and Vanderjagt, D. (1972) *Biochem J 128*, 1353-6.

103. Jakubowski, H., and Fersht, A. R. (1981) *Nucleic Acids Res 9*, 3105-17.

104. Lin, S. X., Baltzinger, M., and Remy, P. (1984) *Biochemistry 23*, 4109-16.

105. Tsui, W. C., and Fersht, A. R. (1981) *Nucleic Acids Res 9*, 4627-37.

106. Dock-Bregeon, A., Sankaranarayanan, R., Romby, P., Caillet, J., Springer, M., Rees, B., Francklyn, C. S., Ehresmann, C., and Moras, D. (2000) *Cell 103*, 877-84.

107. Fersht, A. R., and Dingwall, C. (1979) *Biochemistry 18*, 1238-45.

108. Eldred, E. W., and Schimmel, P. R. (1972) *J Biol Chem 247*, 2961-4.

109. Englisch, S., Englisch, U., von der Haar, F., and Cramer, F. (1986) *Nucleic Acids Res 14*, 7529-39.

110. Pauling, L. (1957) *Festschrift fur*, 597-602.

111. Fersht, A. R., Shindler, J. S., and Tsui, W. C. (1980) *Biochemistry 19*, 5520-4.

112. Hopfield, J. J., Yamane, T., Yue, V., and Coutts, S. M. (1976) *Proc Natl Acad Sci USA 73*, 1164-8.

113. Fersht, A. R., and Dingwall, C. (1979) *Biochemistry 18*, 2627-31.

114. Yamane, T., and Hopfield, J. J. (1977) *Proc Natl Acad Sci U S A 74*, 2246-50.

115. Baldwin, A. N., and Berg, P. (1966) *J Biol Chem 241*, 839-45.

116. Hopfield, J. J. (1974) *Proc Natl Acad Sci U S A 71*, 4135-9.

117. Hale, S. P., Auld, D. S., Schmidt, E., and Schimmel, P. (1997) *Science 276*, 1250-2.

118. Hale, S. P., and Schimmel, P. (1996) *Proc Natl Acad Sci USA 93*, 2755-8.

119. Farrow, M. A., and Schimmel, P. (2001) *Biochemistry 40*, 4478-83.

120. Hendrickson, T. L., Nomanbhoy, T. K., and Schimmel, P. (2000) *Biochemistry 39*, 8180-6.

121. Fersht, A. R. (1977) *Biochemistry 16*, 1025-30.

122. Jakubowski, H., and Goldman, E. (1984) *J Bacteriol 158*, 769-76.

123. Mulvey, R. S., and Fersht, A. R. (1977) *Biochemistry 16*, 4731-7.

124. Schmidt, E., and Schimmel, P. (1994) *Science 264*, 265-7.

125. Starzyk, R. M., Webster, T. A., and Schimmel, P. (1987) *Science 237*, 1614-8.

126. Shiba, K., and Schimmel, P. (1992) *Proc Natl Acad Sci USA 89*, 1880-4.

127. Lin, L., Hale, S. P., and Schimmel, P. (1996) *Nature 384*, 33-4.

128. Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., and Steitz, T. A. (1988) *Proc Natl Acad Sci USA 85*, 8924-8.

129. Beese, L. S., and Steitz, T. A. (1991) *Embo J 10*, 25-33.

130. Hendrickson, T. L., Nomanbhoy, T. K., de Crecy-Lagard, V., Fukai, S., Nureki, O., Yokoyama, S., and Schimmel, P. (2002) *Mol Cell 9*, 353-62.

131. Bishop, A. C., Nomanbhoy, T. K., and Schimmel, P. (2002) *Proc Natl Acad Sci U S A 99*, 585-90.

132. Nomanbhoy, T. K., Hendrickson, T. L., and Schimmel, P. (1999) *Mol Cell 4*, 519-28.

133. Lin, C. S., Irwin, R., and Chirikjian, J. G. (1979) *Nucleic Acids Res 6*, 3651-60.

134. Zagorski, W., Castaing, B., Herbert, C. J., Labouesse, M., Martin, R., and Slonimski, P. P. (1991) *J Biol Chem 266*, 2537-41.

135. Souciet, G., Dietrich, A., Colas, B., Razafimahatratra, P., and Weil, J. H. (1982) *J Biol Chem 257*, 9598-604.

136. Hartlein, M., and Madern, D. (1987) *Nucleic Acids Res 15*, 10199-210.

137. Vander Horn, P. B., and Zahler, S. A. (1992) *J Bacteriol 174*, 3928-35.

138. Bullard, J. M., Cai, Y. C., and Spremulli, L. L. (2000) *Biochim Biophys Acta 1490*, 245-58.

139. Williamson, R. M., and Oxender, D. L. (1990) *Proc Natl Acad Sci USA 87*, 4561-5.

140. Li, D., Wang, E. D., and Wang, Y. L. (1997) *Acta Biochimica et Biophysica Sinica 29*, 591-596.

141. Chen, J., Li, Y., Wang, E., and Wang, Y. (1999) *Protein Expr Purif 15*, 115-20.

142. Li, G. Y., Herbert, C. J., Labouesse, M., and Slonimski, P. P. (1992) *Curr Genet 22*, 69-74.

143. Nureki, O., Kohno, T., Sakamoto, K., Miyazawa, T., and Yokoyama, S. (1993) *J Biol Chem 268*, 15368-73.

144. Landro, J. A., Schmidt, E., Schimmel, P., Tierney, D. L., and Penner-Hahn, J. E. (1994) *Biochemistry 33*, 14213-20.

145. Labouesse, M., Dujardin, G., and Slonimski, P. P. (1985) *Cell 41*, 133-43.

146. Herbert, C. J., Labouesse, M., Dujardin, G., and Slonimski, P. P. (1988) *Embo J 7*, 473-83.

147. Houman, F., Rho, S. B., Zhang, J., Shen, X., Wang, C. C., Schimmel, P., and Martinis, S. A. (2000) *Proc Natl Acad Sci USA 97*, 13743-8.

148. Dietrich, A., Romby, P., Marechal-Drouard, L., Guillemaut, P., and Giege, R. (1990) *Nucleic Acids Res 18*, 2589-97.

149. Asahara, H., Himeno, H., Tamura, K., Hasegawa, T., Watanabe, K., and Shimizu, M. (1993) *J Mol Biol 231*, 219-29.

150. Tocchini-Valentini, G., Saks, M. E., and Abelson, J. (2000) *J Mol Biol 298*, 779-93.

151. Soma, A., Kumagai, R., Nishikawa, K., and Himeno, H. (1996) *J Mol Biol 263*, 707-14.

152. Breitschopf, K., Achsel, T., Busch, K., and Gross, H. J. (1995) *Nucleic Acids Res 23*, 3633-7.

153. Soma, A., Uchiyama, K., Sakamoto, T., Maeda, M., and Himeno, H. (1999) *J Mol Biol 293*, 1029-38.

154. Breitschopf, K., and Gross, H. J. (1994) *Embo J 13*, 3166-7.

155. Breitschopf, K., and Gross, H. J. (1996) *Nucleic Acids Res 24*, 405-10.

156. Duester, G., Campen, R. K., and Holmes, W. M. (1981) *Nucleic Acids Res 9*, 2121-39.

157. Hsu, L. M., Klee, H. J., Zagorski, J., and Fournier, M. J. (1984) *J Bacteriol 158*, 934-42.

158. Nakajima, N., Ozeki, H., and Shimura, Y. (1981) *Cell 23*, 239-49.

159. Yoshimura, M., Inokuchi, H., and Ozeki, H. (1984) *J Mol Biol 177*, 627-44.

160. Komine, Y., Adachi, T., Inokuchi, H., and Ozeki, H. (1990) *J Mol Biol 212*, 579-98.

161. Grosjean, H., and Fiers, W. (1982) *Gene 18*, 199-209.

162. Asahara, H., Himeno, H., Tamura, K., Nameki, N., Hasegawa, T., and Shimizu, M. (1994) *J Mol Biol 236*, 738-48.

163. Soma, A., and Himeno, H. (1998) *Nucleic Acids Res 26*, 4374-81.

164. Asahara, H., Nameki, N., and Hasegawa, T. (1998) *J Mol Biol 283*, 605-18.

165. Chen, J. F., Guo, N. N., Li, T., Wang, E. D., and Wang, Y. L. (2000) *Biochemistry 39*, 6726-31.

166. Tang, Y., and Tirrell, D. A, *unpublished results.*

167. Martinis, S. A., and Fox, G. E. (1997) *Nucleic Acids Symp Ser 36*, 125-8.

168. Apostol, I., Levine, J., Lippincott, J., Leach, J., Hess, E., Glascock, C. B., Weickert, M. J., and Blackmore, R. (1997) *J Biol Chem 272*, 28980-8.

169. Li, T., Li, Y., Guo, N., Wang, E., and Wang, Y. (1999) *Biochemistry 38*, 9084-8.

170. Li, T., Guo, N., Xia, X., Wang, E. D., and Wang, Y. L. (1999) *Biochemistry 38*, 13063-9.

171. Chen, J. F., Li, T., Wang, E. D., and Wang, Y. L. (2001) *Biochemistry 40*, 1144-9.

172. Mursinna, R. S., Lincecum, T. L., Jr., and Martinis, S. A. (2001) *Biochemistry 40*, 5376-81.

173. Dougherty, D. A. (2000) *Curr Opin Chem Biol 4*, 645-52.

174. Jackson, D. Y., Burnier, J., Quan, C., Stanley, M., Tom, J., and Wells, J. A. (1994) *Science 266*, 243-7.

175. Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) *Science 266*, 776-9.

176. Nilsson, B. L., Kiessling, L. L., and Raines, R. T. (2000) *Org Lett 2*, 1939-41.

177. Heckler, T. G., Chang, L. H., Zama, Y., Naka, T., Chorghade, M. S., and Hecht, S. M. (1984) *Biochemistry 23*, 1468-73.

178. Baldini, G., Martoglio, B., Schachenmann, A., Zugliani, C., and Brunner, J. (1988) *Biochemistry 27*, 7951-9.

179. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989) *Science 244*, 182-8.

180. Mendel, D., Cornish, V. W., and Schultz, P. G. (1995) *Annu Rev Biophys Biomol Struct 24*, 435-62.

181. Mendel, D., Ellman, J. A., Chang, Z., Veenstra, D. L., Kollman, P. A., and Schultz, P. G. (1992) *Science 256*, 1798-802.

182. Judice, J. K., Gamble, T. R., Murphy, E. C., de Vos, A. M., and Schultz, P. G. (1993) *Science 261*, 1578-81.

183. Wang, B., Brown, K. C., Lodder, M., Craik, C. S., and Hecht, S. M. (2002) *Biochemistry 41*, 2805-13.

184. Nowak, M. W., Gallivan, J. P., Silverman, S. K., Labarca, C. G., Dougherty, D. A., and Lester, H. A. (1998) *Methods Enzymol 293*, 504-29.

185. Saks, M. E., Sampson, J. R., Nowak, M. W., Kearney, P. C., Du, F., Abelson, J. N., Lester, H. A., and Dougherty, D. A. (1996) *J Biol Chem 271*, 23169-75.

186. Li, L., Zhong, W., Zacharias, N., Gibbs, C., Lester, H. A., and Dougherty, D. A. (2001) *Chem Biol 8*, 47-58.

187. Gallivan, J. P., Lester, H. A., and Dougherty, D. A. (1997) *Chem Biol 4*, 739-49.

188. Kearney, P. C., Zhang, H., Zhong, W., Dougherty, D. A., and Lester, H. A. (1996) *Neuron 17*, 1221-9.

189. Short III, G. F., Golovine, S. Y., and Hecht, S. M. (1999) *Biochemistry 38*, 8808-19.

190. Hohsaka, T., Ashizuka, Y., Murakami, H., and Sisido, M. (1996) *J Am Chem Soc 118*, 9778-9779.

191. Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., Nakayama, H., Takio, K., Yabuki, T., Kigawa, T., Kodama, K., Nishikawa, K., and Yokoyama, S. (2002) *Nat Biotechnol 20*, 177-82.

192. Hohsaka, T., Ashizuka, Y., Sasaki, H., Murakami, H., and Sisido, M. (1999) *J Am Chem Soc 121*, 12194-12195.

193. Hohsaka, T., Kajihara, D., Ashizuka, Y., Murakami, H., and Sisido, M. (1999) *J Am Chem Soc 121*, 34-40.

194. Murakami, H., Hohsaka, T., Ashizuka, Y., and Sisido, M. (1998) *J Am Chem Soc 120*, 7520-7529.

195. Bain, J. D., Switzer, C., Chamberlin, A. R., and Benner, S. A. (1992) *Nature 356*, 537-9.

196. van Hest, J. C., and Tirrell, D. A. (2001) *Chem Commun*, 1897-1904.

197. Liu, D. R., and Schultz, P. G. (1999) *Proc Natl Acad Sci USA 96*, 4780-5.

198. Pastrnak, M., Magliery, T. J., and Schultz, P. G. (2000) *Helv. Chim. Acta 83*, 2277-2286.

199. Kowal, A. K., Kohrer, C., and RajBhandary, U. L. (2001) *Proc Natl Acad Sci U S A 98*, 2268-73.

200. Furter, R. (1998) *Protein Sci 7*, 419-26.

201. Kast, P., and Hennecke, H. (1991) *J Mol Biol 222*, 99-124.

202. Wang, P., Son, S., Tang, Y., Furter, R., and Tirrell, D. A. (2002) *Manuscript in Preparation*.

203. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) *Science 292*, 498-500.

204. Wang, L., Brock, A., and Schultz, P. G. (2002) *J Am Chem Soc 124*, 1836-1837.

205. Wang, L., and Schultz, P. G. (2001) *Chem Biol 8*, 883-90.

206. Renner, C., Alefelder, S., Bae, J. H., Budisa, N., Huber, R., and Moroder, L. (2001) *Angew Chem Int Ed Engl 40*, 923-925.

207. Feeney, J., McCormick, J. E., Bauer, C. J., Birdsall, B., Moody, C. M., Starkmann, B. A., Young, D. W., Francis, P., Havlin, R. H., Arnold, W. D., and Oldfield, E. (1996) *J Am Chem Soc 118*, 8700-8706.

208. Mohammadi, F., Prentice, G. A., and Merrill, A. R. (2001) *Biochemistry 40*, 10273-83.

209. Parsons, J. F., Xiao, G., Gilliland, G. L., and Armstrong, R. N. (1998) *Biochemistry 37*, 6286-94.

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

211. Duewel, H. S., and Honek, J. F. (1998) *J Protein Chem 17*, 337-50.

212. Duewel, H. S., Daub, E., Robinson, V., and Honek, J. F. (2001) *Biochemistry 40*, 13167-76.

213. Danielson, M. A., and Falke, J. J. (1996) *Annu Rev Biophys Biomol Struct 25*, 163-95.

214. Holmgren, S. K., Bretscher, L. E., Taylor, K. M., and Raines, R. T. (1999) *Chem Biol 6*, 63-70.

215. Jenkins, C. L., and Raines, R. T. (2002) *Nat Prod Rep 19*, 49-59.

216. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F., and Tirrell, D. A. (2001) *Angew Chem Int Ed Engl 40*, 1494-1496.

217. Tang, Y., Ghirlanda, G., Vaidehi, N., Kua, J., Mainz, D. T., Goddard, I. W., DeGrado, W. F., and Tirrell, D. A. (2001) *Biochemistry 40*, 2790-6.

218. Tang, Y., and Tirrell, D. A. (2001) *J Am Chem Soc 123*, 11089-90.

219. Bilgicer, B., Fichera, A., and Kumar, K. (2001) *J Am Chem Soc 123*, 4393-9.

220. Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (1995) *Eur J Biochem 230*, 788-96.

221. Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neuefeind, T., Moroder, L., and Huber, R. (1997) *J Mol Biol 270*, 616-23.

222. Budisa, N., Alefelder, S., Bae, J. H., Golbik, R., Minks, C., Huber, R., and Moroder, L. (2001) *Protein Sci 10*, 1281-92.

223. Bae, J. H., Alefelder, S., Kaiser, J. T., Friedrich, R., Moroder, L., Huber, R., and Budisa, N. (2001) *J Mol Biol 309*, 925-36.

224. Kothakota, S., Mason, T. L., Tirrell, D. A., and Fournier, M. J. (1995) *J Am Chem Soc 117*, 536-537.

225. Yuan, T., Walsh, M. P., Sutherland, C., Fabian, H., and Vogel, H. J. (1999) *Biochemistry 38*, 1446-55.

226. Kiick, K. L., and Tirrell, D. A. (2000) *Tetrahedron 56*, 9487-9493.

227. Jakubowski, H. (2000) *J Biol Chem 275*, 21813-6.

228. van Hest, J. C., and Tirrell, D. A. (1998) *FEBS Lett 428*, 68-70.

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

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

231. Clark, T. D., Kobayashi, K., and Ghadiri, M. R. (1999) *Chem Eur J. 5*, 782-792.

232. Miller, S. J., Blackwell, H. E., and Grubbs, R. H. (1996) *J Am Chem Soc 118*, 9606-9614.

233. Saxon, E., and Bertozzi, C. R. (2000) *Science 287*, 2007-10.

234. Kiick, K. L., van Hest, J. C., and Tirrell, D. A. (2000) *Angew Chem Int Ed Engl 39*, 2148-2152.

235. Kiick, K. L., Weberskirch, R., and Tirrell, D. A. (2001) *FEBS Lett 502*, 25-30.

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

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

238. Kirshenbaum, K., Carrico, I. S., and Tirrell, D. A. (2002) *Chembiochem 3*, 235-7.

239. Datta, D., Wang, P., Carrico, I. S., Mayo, S., and Tirrell, D. A. (2002) *J Am Chem Soc in press*.

CHAPTER 2

Stabilization of Coiled-Coil Peptide Domains by Introduction of Trifluoroleucine

This chapter appeared as a full paper in Tang, Y., Ghirlanda, G., Vaidehi, N., Kua, J., Mainz, D. T., Goddard, W. A., DeGrado, W. F. and Tirrell, D. A. *Biochemistry* (**2001**) *40*, 2790-2796.

2.0 Abstract

Substitution of leucine residues by 5,5,5-trifluoroleucine at the *d*-positions of the leucine zipper peptide GCN4-p1d increases the thermal stability of the coiled-coil structure. The midpoint thermal unfolding temperature of the fluorinated peptide is elevated by 13°C at 30 μM peptide concentration. The modified peptide is more resistant to chaotropic denaturants, and the free energy of folding of the fluorinated peptide is 0.5 to 1.2 kcal/mol larger than that of the hydrogenated form. A similarly fluorinated form of the DNA-binding peptide GCN4-bZip binds to target DNA sequences with affinity and specificity identical to those of the hydrogenated form, while demonstrating enhanced thermal stability. Molecular dynamics simulation on the fluorinated GCN4-p1d peptide using the Surface Generalized Born implicit solvation model revealed that the coiled-coil binding energy is 55% more favorable upon fluorination. These results suggest that fluorination of hydrophobic substructures in peptides and proteins may provide new means of increasing protein stability, enhancing protein assembly, and strengthening receptor-ligand interactions.

2.1 Introduction and Background

Engineering of stable enzymes and robust therapeutic proteins is of central importance to the biotechnology and pharmaceutical industries. Although protein engineering provides powerful tools for the enhancement of enzymatic activity and protein stability (1-4), the scope of *in vivo* engineering methods is limited by the availability of just twenty naturally occurring proteinogenic amino acids (5). Increasing success in the incorporation of noncanonical amino acids into recombinant proteins *in vivo* has allowed the introduction of novel side-chain functionality into engineered proteins (6-10), and raises prospects of new approaches to the design of peptides and proteins of enhanced activity and/or stability.

Leucine zipper peptides are ideal models for the study of protein secondary and tertiary interactions (11-20, and references therein). Such peptides assemble into coiled-coil dimers, trimers and tetramers in order to exclude solvent at the hydrophobic interfaces between adjacent peptide helices (Figure 2.1), and the relations between sequence and stability have been carefully examined (21). In particular, the structure (13), dimerization kinetics (20), and thermodynamics (19) of the model peptide GCN4-p1 have been thoroughly described.

GCN4-p1 constitutes the dimerization domain of bZip, which is a 56-amino acid DNA binding segment (residues 226-281) of the eukaryotic transcription factor GCN4. The N-terminus of bZip contains a DNA recognition domain rich in basic residues such as lysine and arginine. The C-terminal subdomain of bZip contains the GCN4-p1 peptide segment and facilitates dimerization of the protein. While direct contact between DNA and the N-terminal subdomain is important to recognition, protein-protein interactions at the C-terminus also contribute to the specificity and affinity of peptide-DNA binding (22-24).

We present here a successful attempt to stabilize the coiled-coil forms of both GCN4-p1 (Fig. 2A) and bZip (Fig. 2B) by substitution of the core leucine residues of the peptide with 5,5,5-trifluoroleucine (Tfl, **1**) (Fig. 2C). Furthermore, the DNA binding behavior of the fluorinated form of bZip is shown to be identical to that of the wild-type peptide.

The choice of Tfl (Fig 2C) for these studies was based on three factors. First, we imagined that Tfl might behave as a hyper-hydrophobic analog of leucine, since many organofluorine compounds exhibit

lower solubility in water than their hydrocarbon equivalents (25, 26). Second, fluorinated amino acids are nearly isosteric to their natural counterparts, and are equally (or more) inert chemically (27). Third, Tfl is accepted as a substrate by the endogenous leucyl-tRNA synthetase (LeuRS) of microbial cells, and is readily incorporated into recombinant proteins *in vivo (*28, 29*)*. When leucine-depleted cultures of *Escherichia coli* are shifted to Tfl-supplemented medium, more than 95% of the leucine positions in a recombinant protein can be substituted by Tfl under appropriate conditions (30).

We also discuss in this paper the results of molecular dynamics simulation of the thermodynamic properties of the fluorinated leucine zipper peptide. Simulation with an implicit solvation model confirmed the experimental finding that the Tfl-substituted coiled-coils are indeed more stable than their wild-type counterparts.

2. 2. Material and Methods

2.2.1. Trifluoroleucine Synthesis

Fluorinated starting material (**2**) was purchased from Lancaster Synthesis (Windham, NH) or Oakwood Products Inc. (West Columbia, SC). All other reagents were obtained from Aldrich Chemical Company. Porcine kidney acylase I was purchased from Sigma. Unless otherwise stated, all reagents were used as received. Hydrogenation was performed with a high-pressure hydrogenation apparatus. ^1H spectra were measured with a General Electric NMR QE-300 Plus spectrometer.

Ethyl-Trifluoromethylbutyrate **(3).** Trifluoromethylcrotonic acid **2**, 20 g (130 mmol) was dissolved in 40 ml of absolute ethanol. 200 mg of 10% Pd/C was added to the solution. The mixture was hydrogenated under 60 psig of H_2 until hydrogen consumption had ceased. The solution was filtered under vacuum with celite to remove the catalyst. The filtrate was then diluted with 300 ml of absolute ethanol. 40 g of concentrated sulfuric acid was then added dropwise to the solution. The mixture was then fitted with a condenser and refluxed at 120°C for 24 hours. After cooling, the reaction mixture was poured into a separatory funnel and a large excess (600 ml) of water was added. Without shaking, the denser liquid **3** separated from the water-ethanol mixture. The phases were separated and the bottom phase was distilled

(bp=123°C) to give 18g (98 mmol, 74%) of 3 as a colorless liquid. ^{1}H NMR: ($CDCl_3$) δ 1.1 (d, J=6, 3H), 1.2 (t, J=3, 3H), 2.2-2.6 (m, 2H), 2.7 (m, 1H), 4.1 (q, 2H).

2-oxo-4-trifluoromethyl pentanoic acid **(5)**. NaH, 9 g (225 mmol, 60% dispersion in mineral oil) was added to 200 ml of diethyl ether and purged with argon. A 50% solution of 17.1 g (117 mmol) diethyl oxalate in diethyl ethyl was added to the slurry with stirring. 18 g (98 mmol) of 3 was added to the slurry and the solution was refluxed at 80°C for 36 hours. At the end of the 36 hours, reaction mixtures was quenched by pouring into a mixture of ice-water-sulfuric acid (10%, 15 g). The aqueous phase was then separated and discarded. The organic phase was then washed with 300ml of 5% sodium bicarbonate and dried over magnesium sulfate. (Note: although sodium bicarbonate wash was not necessary and might result in loss of some products, this extra step greatly cleans up the product and makes subsequent steps easier). After evaporation of ether and distillation of traces of **3** (60°C, 80 mmHg), a crude mixture of **4** in mineral oil was obtained, 20 g (71%). The crude was then added to 4 parts water and 2 parts concentrated hydrochloric acid. The mixture was refluxed at 130°C for 24 hours and was stopped when carbon dioxide evolution had apparently ceased. After cooling, the mixture was first made slightly alkaline with sodium hydroxide and washed with diethyl ether to remove mineral oil and traces of starting material. Then the aqueous solution was acidified to pH 1 with hydrochloric acid and extracted with ether. The organic phase was dried over magnesium sulfate followed by evaporation of ether. The yellow/orange oil residue was **5**. Although further purification can be performed, NMR showed satisfactory purity. 11.3 g (61 mmol, 87%). 1H NMR: ($CDCl_3$) δ 1.1 (d, J=6, 3H), 2.8 (m, 2H), 2.9-3.1 (m, 2H).

2-oximino-4-trifluoromethyl pentanoic acid **(6)**. 10.5 g (57 mmol) of **5** was added to 100 ml of 1:1 mixture of ethanol and water. 4.3 g (65 mmol) of hydroxylamine (50% in H2O) was added to mixture and the pH adjusted to between 4-5. The reaction mixture was stirred at room temperature for 10 hours during which time the solution became clear. The solution was then acidified to pH 1 with 1N HCl and extracted with ether. The ether phase was dried over magnesium sulfate and evaporated to dryness. 8g (40mmol, 70%) of yellow/white solid was recovered. The oxime can be recrystallized from chloroform and pentane to yield 7.1g (36 mmol, 62%) of **6** as while/yellow needles. ^{1}H NMR: (CH_3OD) δ 1.1 (d, J=6, 3H), 2.7 (m, 3H).

5,5,5-D,L-trifluoroleucine (**7**). 6.4 g (32 mmol) of **6** was dissolved in 21 ml of a 10:10:1 mixture of water:ethanol:ammonium hydroxide. After the addition of 0.3 g of PtO_2, the mixture was hydrogenated over 60 psig of hydrogen overnight. When the hydrogen consumption stopped, another 0.1 g of the fresh catalyst was added. When hydrogen pressure drop had ceased, the catalysts were filtered. The retained solids were washed with water to dissolve the trifluoroleucine. The filtrate was evaporated to dryness under reduced pressure. The solids were washed by cold ethanol to remove any unreacted oxime and filtered again. The retained crude trifluoroleucine was recrystallized from water and ethanol to yield 4.6 g (25 mmol, 78%) of pure **7**. $[\alpha]_D^{23}$ = 0.03 (c=1, 1N HCl), ^{1}H NMR: (D_2O) δ 1.1 (d, J=6, 3H), 1.7-2.2 (m, 2H), 2.4-2.6 (m, 1H), 3.7-3.8 (m, 1H).

5,5,5-trifluoro-*N*-acetyl-leucine (**8**). 4 g (21.6 mmol) of **7** was suspended in 50 ml of acetic acid and heated to above 100°C. **7** dissolves rapidly in acetic acid when heated. The heat sources was removed and when the temperature reached 100°C, 3.6 g (35 mmol) of acetic anhydride was added to the solution. When the temperature dropped down to 80°C, 20ml of water was added. The mixture was then dried under vacuum. Cyclohexane was added in 10ml portions near the end of the evaporation to form an azeotrope with water and acetic acid in order to dry the syrup. After complete evaporation, 4.8 g (21.1 mmol, 98%) of **8** was obtained as a white solid. No reaction with ninhydrin and completely soluble in acetone.

5,5,5-L-trifluoroleucine (**1**). 4.8 g (21.1 mmol) of **8** was suspended in 100ml of water. The pH of the solution was adjusted to 7.8 and the solution was filtered. 100 mg of porcine kidney acylase I was added to the solution and the mixture capped and stirred overnight. The solution was first acidified to pH 5. 100 mg of Norit was added and the solution heated to coagulate and denature the enzyme. The solution was then filtered and acidified again to pH 1. The solution was then washed with ethyl acetate. The organic phase was applied to a column of DOWEX ion-exchange resins previously equilibrated with water. The column was washed with water until the eluent became neutral. The amino acid was then eluted with 1N ammonium hydroxide solution. Evaporation of ammonia and water gave pure L amino acid **1**. Yield was 0.4 g (2.2 mmol, 20%). $[\alpha]_D^{23}$ = 14.5 (c=0.2, 5N HCl), ^{1}H NMR: (D_2O) δ 1.1 (d, J=6, 3H), 1.7-2.2 (m, 2H), 2.4-2.6 (m, 1H), 3.7-3.8 (m, 1H).

5,5,5-L-trifluoro-N-FMOC-leucine (**9**). 1 g (5.4 mmol) of **1** was suspended in 20.2 ml of 10% Na_2CO_3 and 10 ml of dioxane. 1.4 g of fluorenylchloroformate in 15 ml of dioxane was added to the slurry

in ice bath. The mixture was then stirred at room temperature for 3 hours. The reaction was quenched by pouring into 400 ml of water and extracted twice with either. The aqueous phase was then acidified to pH 1 resulting in white precipitation which was subsequently extracted with ethyl acetate. The ethyl acetate extract was then dried with sodium sulfate and concentrated. Cyclohexane is added in 10ml fractions at the end of evaporation to help completely remove ethyl acetate as an azeotrope. The solid is then recrystalized from ether/pentane to yield 1.65g (4mmol) of protected L-trifluoroleucine. $[\alpha]_D^{23} = -23$ (c=0.1, DMF)

2.2.2. Peptide Synthesis and Purification

Peptides were synthesized at the Biopolymer Synthesis Center at the California Institute of Technology (Pasadena, CA 91125). Automated, stepwise solid-phase synthesis was performed on an ABI 433A synthesizer employing Fmoc chemistry. The trifluoroleucine units were incorporated into the peptide with extended coupling cycles. After chain assembly was complete, the peptide was deprotected and removed from the resin support with trifluoroacetic acid in the presence of 1,2-ethanedithiol, thioanisole and water. Peptides were precipitated into cold methyl t-butyl ether and isolated by centrifugation. Peptide products were purified by preparative C_{18} reverse phase HPLC using a non-linear gradient of 0-80 % elution solution (0.1% TFA / 60% acetonitrile / 40% H_2O) in 120 min. The peptides examined in this work were *not* acylated at their N-termini, and are designated GCN4-p1d to distinguish them from the acylated variants reported in earlier papers (e.g., 11). The thermal melting temperature of Leu-GCN4-p1d is slightly lower than that reported for the acylated form GCN4-p1 (*11*).

2.2.3. Ultracentrifugation Analysis

Sedimentation equilibrium analysis was performed using a Beckman XLI analytical ultracentrifuge, recording interference data and radial absorbance at 236 and 280 nm at the same time. Initial peptide concentrations ranged between 100 and 300 μM; buffer was 0.01M sodium phosphate, pH 7.4, containing 0.1 M NaCl. The samples were centrifuged at 35000, 40000, 45000 rpm, until equilibrium was reached. Partial specific volumes were calculated by the residue-weighted average method of Cohn and Edsall (31). Solution densities were estimated by using solute concentration-dependent density tables in the *CRC Handbook of Chemistry and Physics*. The data were fit as single species to provide an estimate of the

aggregation states. Curve fitting of analytical ultracentrifuge data was done using Igor Pro (Wavemetrics Inc., Oswego WS) with procedures adapted from Ref 32 by Dr. James D. Lear.

2.2.4. Spectroscopic Analysis

CD spectra were recorded on an Aviv 62DS spectropolarimeter (Lakewood, NJ) in PBS buffer, pH 7.0. All peptide concentrations were determined by amino acid analysis of a stock solution (2 mg/ml). Experiments were performed in a rectangular cell with pathlength of 1 mm. Spectra were scanned from 260 nm to 194 nm with points taken every 1 nm. The temperature of the solution was maintained by a thermostatically controlled cuvette holder (HP model 89101A). Temperature scans were performed from 0°C to 100°C in 1°C steps. Five scans were performed on a single sample and averaged. Each data point was collected after 30 seconds of thermal equilibration at the desired temperature.

The thermodynamic data were calculated by fitting the thermal denaturation curves to a monomer-dimer equilibrium according to published procedures (16). The data were fit using a standard state of 1 M. Mean residue ellipticity at 222 nm was fit as a function of temperature and total peptide concentration using MLAB software as stated in the reference. The data in the text were obtained from global fits of the thermal denaturation data at peptide concentrations of 3 μM and 85 μM. All thermodynamic quantities reported in the text are per mole of monomers.

2.2.5. Gel Retardation Assays

Oligonucleotides containing the AP-1 binding site (AP-1, 5'-GTGGAGATGACTCATCTCCGG-3'), the CREB binding site (CREB, 5'-TGGAGATGACGTCATCTCCT-3') and the nonspecific sequence (NON, 5'-GATCCCAACACGTGTTGGGATC-3') were synthesized at the Caltech DNA synthesis facility. The oligonucleotides were labeled with γ-^{32}P-ATP (> 6,000 Ci/mmol), annealed and purified. Labeled probes (5,000 cpm) were incubated with protein for 30 min at 4°C in 40 μl of binding buffer (20 mM Tris-HCl, pH 7.0, 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 10% glycerol and 40 μg/ml poly(dl-dC)•poly(dl-dC)). An aliquot of the reaction volume (10 μl) was analyzed by electrophoresis on 5% polyacrylamide gel. Free and protein-bound DNA were visualized by autoradiography. The relative intensities of the bands were measured by densitometry.

2.2.6. Molecular Dynamics Simulation

Molecular dynamics (MD) simulation of Leu-GCN4-p1d and Tfl-GCN4-p1d were performed using the MPSim program (33). The MPSim MD program, which includes the Cell Multipole Method (34), provides fast and accurate calculations of non-bond interactions. MPSim describes the energies and forces due to polarization of the continuum solvent using either the Poisson-Boltzmann (PBF) (35) or Surface Generalized Born (SGB) (36) models. For the dynamics we used the SGB approximation to the PB continuum solvent model to calculate the forces on the Leu-GCN4-p1d dimer due to solvent polarization; the final energies were then calculated using the PBF method, which leads to more accurate energies.

The DREIDING force field (FF) (*37*) was used in all calculations. The recommended van der Waals potentials in DREIDING (having the exponential-six form) are used here. The default in the MSI software (PolyGraf and Cerius^2) is Lennard-Jones 12-6 which we have found to be significantly less accurate. Also we use a dielectric constant of $\varepsilon = 1$ (not the distance-scaled ε of PolyGraf and Cerius^2). The charges on Tfl-GCN4-p1d and Hfl-GCN4-p1d were derived from first principles quantum mechanics (QM): (LMP2 method using the 6-31G** basis set [Jaguar 3.5 (38) from Schrodinger Inc., Portland, Oregon. Jaguar 3.5 is distributed by Schrodinger Inc, Portland Ore.]. These charges were calculated for the tripeptide, Gly-Leu (nF)-Gly. The charges for the natural residues in the protein were taken from the CHARMM FF (39).

2.3 Results and Discussion

2.3.1 Amino acid preparation and peptide synthesis

The synthesis of 5,5,5-L-trifluoroleucine is shown in Figure 2.3. The overall reaction proceeded with 18% overall yield. The resolution of the L-isomer is accomplished with porcine kidney acylase I. The purity of the resolution is checked with modified Mosher's acid (data not shown) and was shown to be > 99% e.e. at the α position. The γ position was left as a mixture of isomers.

We prepared four peptides to study the effects of fluorination on coiled-coil stability and DNA binding affinity and specificity (Figure 2.2). We prepared the wild-type and fluorinated forms for both the

GCN4-p1d (Leu-GCN4-p1d and Tfl-GCN4-p1d) and bZip (Leu-bZip and Tfl-bZip) peptides. Solid phase peptide synthesis was used in this initial study to replace quantitatively all four ***d***-position leucines in GCN4-p1d (Fig. 2.4). For Tfl-bZip, the corresponding ***d***-position leucines in the GCN4-p1d dimerization domain were substituted with Tfl. *N*-Fmoc-5,5,5-trifluoro-L-leucine (**9**) was used as an equimolar mixture of the 2S,4S- and the 2S,4R- isomers to prepare the fluorinated peptides. None of the four peptides was acylated at the *N*-terminus. The peptides were purified by HPLC and the molar masses of the purified peptides were verified by mass spectrometry (Figure 2.5-2.8).

2.3.2. Spectroscopic Characterization of Leu-GCN4-p1d and Tfl-GCN4-p1d

Circular dichroism (CD) spectra of Leu-GCN4-p1d and Tfl-GCN4-p1d indicated that both peptides are highly helical as evidenced by intense minima at 222 nm and 208 nm (Fig. 2.9). The spectra of the wild type and fluorinated peptides are essentially coincident, suggesting nearly identical secondary structures; both peptides are highly helical at 0°C. This observation confirmed our conjecture that replacement of leucine with Tfl would not disrupt interhelical packing and interfere with folding of the coiled-coil structure. Ultracentrifugation indicates that Tfl-GCN4-p1d is predominantly dimeric at the concentrations of interest in this work. Data for Tfl-GCN4-p1d were fit to a monomer-dimer-trimer equilibrium, giving values of K_d of the order of 10^{-8} M and 10^{-14} M^2, respectively, for the monomer-to-dimer and monomer-to-trimer equilibria. In the concentration range of approximately 10 μM – 40 μM, the peptide is ca. 85% dimeric.

The thermal stabilities of the coiled-coil dimers of Leu-GCN4-p1d and Tfl-GCN4-p1d were examined by CD spectroscopy (Fig. 2.10). A significant increase in the thermal stability of Tfl-GCN4-p1d as compared to Leu-GCN4-p1d is reflected in an elevation of the thermal denaturation temperature from 48°C to 61°C at a peptide concentration of 30 μM. The 13°C increment in T_m is remarkable in view of the fact that no increase in the thermal stability of GCN4-p1d has been reported based solely on substitution of the leucine residues at the ***d***- positions. Mutations at the ***d***-positions to other natural amino acids have all resulted in losses in helix stability due to decreases in packing efficiency since such changes are usually of the "large to small" type (21). As expected for a monomer-dimer equilibrium, the denaturation curves depend on the peptide concentrations, and their midpoints shift to higher temperature as the concentrations

of peptides are increased (Fig. 2.11). The thermodynamic changes associated with the transition (folded dimer to unfolded monomers) can be calculated from the melting curves by fitting the data to a monomer-dimer equilibrium. The thermodynamic analysis of Tfl-GCN4-p1d is complicated by its heterogeneity as a result of the presence of different Tfl stereoisomers. A more accurate thermodynamic analysis of the Tfl-GCN4-p1d can be obtained when pure stereoisomers of Tfl become available. (During HPLC purification, an additional peptide fraction of the same molecular weight as Tfl-GCN4-p1d was recovered. This fraction has a thermal denaturation temperature slightly lower than that of Tfl-GCN4-p1d. We attribute this result to stereochemical heterogeneity at the γ-position of L-Tfl, consistent with the prediction of the molecular dynamics simulation reported here.)

Global analysis of the thermal unfolding curves at two different concentrations (approximately 85 μM and 3 μM) yielded ΔH° = 60.2±1.1 kcal mol^{-1}, T_m = 385.4± 0.4 K and ΔC_p = 530± 40 cal mol^{-1} K^{-1} for Tfl-GCN4-p1d (1 M standard state); the corresponding values for Leu-GCN4-p1d are 70.3±1.4 kcal mol^{-1}, 365.6±0.4 K and 740±50 cal mol^{-1} K^{-1}, respectively (all thermodynamic data reported are for per mole monomer). Under all conditions where a direct experimental comparison was possible, Tfl-GCN4-p1d is 0.5 ~ 1.2 kcal mol^{-1} more stable than Leu-GCN4-p1d; for example, at 50°C, K_d for dimerization is 67.8 μM for Leu-GCN4-p1d and 9.8 μM for Tfl-GCN4-p1d.

The stability of Tfl-GCN4-p1d toward denaturation by chaotropic reagents was demonstrated through guanidine hydrochloride (GuHCl) titration experiments (Fig. 2.12). At each temperature examined, the fluorinated peptide displayed significantly lower susceptibility toward denaturation by GuHCl; in each case, the concentration of GuHCl needed to unfold 50% of the peptide was higher for Tfl-GCN4-p1d than for the wild-type peptide. The free energies of folding at 25°C can be obtained by globally fitting the GuHCl denaturation curves at two different concentrations for each peptide to a monomer-dimer equilibrium, resulting in extrapolated ΔG° (in absence of GuHCl) of 8.0±0.1 kcal mol^{-1} for Leu-GCN4-p1d and 8.6±0.1 kcal mol^{-1} for Tfl-GCN4-p1d, corresponding to values of K_d of 1.2 μM and 0.51 μM, respectively.

2.3.3 DNA Binding Studies.

To determine whether protein function can be retained upon fluorination, we compared the affinities and specificities of DNA binding by Leu-bZip and Tfl-bZip. The secondary structures of the two proteins were identical as indicated by CD spectroscopy and the thermal melting temperature of Tfl-bZip was elevated by 8°C compared to Leu-bZip at 10 μM peptide concentration (data not shown).

The DNA-binding domain of Leu-bZip is conformationally disordered in the absence of specific DNA sequences, while the dimerization domain forms a two-stranded coiled coil through the leucine zipper motif at concentrations above the monomer-to-dimer equilibrium (40). Upon presentation of target DNA sequences, the DNA binding region folds into a α-helical structure and the peptide binds to the DNA in a "chopstick" model (41-43). CD analysis of Tfl-bZip revealed that the fluorinated peptide behaves in the same manner as Leu-bZip (Fig. 2.13). Before addition of oligonucleotides containing the target (CREB) binding site, Tfl-bZip is approximately 70% helical, as expected on the basis of the length of its helical dimerization domain. After addition of target DNA, Tfl-bZip is essentially 100% helical, indicating a transition of the DNA-binding region from coil to helix. The similar changes in secondary structure observed for Leu-bZip and Tfl-bZip suggest that fluorination of the zipper domain does not cause qualitative changes in the nature of the association between the peptide and DNA.

The affinity and specificity of binding of the fluorinated peptide were shown to be essentially identical to those of the wild type peptide on the basis of gel-retardation assays (44, 45) (Fig. 2.14). Leu-bZip binds to the AP-1 and CREB binding sites with nearly equal affinities even though the spacing between the half-sites of these DNAs are different. Densitometric analysis of mobility shift assays revealed that Tfl-bZip binds to both sequences with specificities and affinities (K_d =12.5 ± 0.7 nM for AP-1 and 5.4 ± 0.6 nM for CREB) nearly identical to those of Leu-bZip (K_d = 12.8 ± 1 nM for AP-1 and 4.8 ± 0.5 nM) for CREB. Neither Leu-bZip nor Tfl-bZip recognizes nonspecific (NON) sequences, as shown by the lack of detectable peptide-bound DNA in assays with NON oligonucleotides (Fig. 2.14, bottom).

2.3.4 Molecular Dynamics Simulation.

To determine the origins of the stabilizing effect of side-chain fluorination, we carried out molecular dynamics (MD) calculations on Leu-GCN4-p1d and Tfl-GCN4-p1d using the Poisson-

Boltzmann (PB) continuum description of the solvent. The PB description of solvation implicitly includes entropic changes in the solvent; thus the calculations lead directly to binding free energies (ΔG^{BE}). The MPSIM MD program and the DREIDING Force Field (FF) (37) were used for all calculations. The starting structure for the Leu-GCN4-p1d dimer was that of GCN4-p1 as reported in the RCSB Protein Data Bank; those of the fluorinated dimers were derived from the native dimer structure by replacement of the appropriate methyl hydrogens with fluorines, followed by re-optimization of the structure. Because the γ-carbon of Tfl is asymmetric (Figure 2C), multiple arrangements of adjacent diastereotopic trifluoromethyl groups must be considered (Figure 2.15). When both Tfl residues at a given ***d***-position are of the (2S,4S) configuration, the two trifluoromethyl groups are relatively close to one another; the fluorinated carbon centers are separated by ca. 6 Å. On the other hand, when two (2S,4R) isomers are juxtaposed, the corresponding carbon-carbon distance increases to about 8 Å. In the remaining configurations (where the two strands carry different isomers), the trifluoromethyl groups are separated by intermediate distances. We performed simulations on all configurations to determine how side-chain stereochemistry affects dimer stability. For simulation of strands containing different stereoisomers of trifluoroleucine, we considered only those cases in which all four trifluoroleucines on one strand have the same stereoconfiguration. For each dimer we carried out 1 ns of constant temperature (300K) Nose-Hoover MD with the SGB description of the water solvent. We used counter-ions for neutralizing the charges on the side chains of Arg, Lys, Asp and Glu. For the native dimer the 1 ns MD simulation leads to a structure in good agreement with the experimental structure of GCN4-p1, with an RMS deviation in coordinates of the main chain atoms of 2.15Å.

From the 1 ns trajectory, we calculated the average properties over 800 ps after equilibration. ΔG^{BE} was calculated as the difference in energies of the solvated dimer and the corresponding solvated monomers. The monomers were described with short (50 ps) SGB MD simulations starting with the random coil structure. The dynamics structures were minimized and an average energy of the minimized structures was considered for the calculation (ΔG^{BE}). For the final (ΔG^{BE}), solvation was calculated using PBF (which is more accurate than SGB). Table 2.1 reports the average values of ΔG^{BE} (per monomer) for the native and fluorinated forms of GCN4-p1d.

The Tfl-GCN4-p1d dimers are predicted to exhibit ΔG^{BE} ca. 55% larger than that of the Leu-GCN4-p1d (calculated relative to the respective random coil monomers). The various stereochemical arrangements lead to increases in binding energies ranging from 44% to 71%, indicating that side-chain configuration may have some differential effect on dimer stability. Similar calculations for the hexafluoroleucine (Hfl) dimer lead to the prediction that such dimers (which were not prepared experimentally in this work) should be significantly less stable than the Tfl dimers but marginally more stable (19%) than the wild type. To investigate the source of stability of the fluorinated dimers, we analyzed the components of the binding energy for each peptide (Table 2.2). We find that the primary driving forces for stabilizing the Tfl-GCN4-p1d dimers arise from van der Waals (vdW) and hydrogen bonding interactions. The predicted structures of the Hfl and wild type monomers are globular, while the Tfl-GCN4-p1d monomer is more extended, with local "kinks" arising from favorable electrostatic interactions between Tfl residues (e.g., between Tfl_5 and Tfl_{12}). The globular structures of Leu-GCN4-p1d and Hfl-GCN4-p1d form more nonlocal hydrogen bonds and more favorable vdW contacts than the more extended Tfl-GCN4-p1d monomer. Hence, the gain in H-bond and vdW energies in forming a dimer is greater for Tfl-GCN4-p1d than for the wild-type or Hfl peptides because the latter monomers must pay the energy cost of extension as a prerequisite for dimerization.

Furthermore, consideration of electrostatic (intra- and inter-peptide coulomb forces) and solvation interactions suggests that the hydrophobic preference in the dimer for burial of CF_3 is greater than for CH_3. Considering just coulomb and solvation interactions, the driving force for dimerization is predicted to decrease in the order Hfl > Tfl > Leu. It is the balance of desolvation, electrostatics, H-bonding and vdW forces that leads to the prediction that the Tfl dimers are more stable than the Hfl dimer which in turn is more stable than the native leucine dimer. The average helicity of the dimers is predicted to be 90.8% for Leu-GCN4-p1d, 83.5% for Tfl-GCN4-p1d, and 78.5% for Hfl-GCN4-p1d.

2.4 Conclusion

The experimental and computational results described here demonstrate that the subtle change from four leucine methyl groups to four trifluoromethyl groups results in a substantial gain in stability of

the folded structure of a dimeric coiled-coil peptide. It is remarkable that for a peptide of the size of GCN4-p1d, fluorination results in a coiled-coil structure that is highly resistant to both thermal and denaturant unfolding as compared to the wild-type peptide. Although these studies used solid phase peptide synthesis to prepare GCN4-p1d, we have also demonstrated that fluorinated peptides produced *in vivo* exhibit substantial elevation in stability compared to their wild type analogues (30) (next chapter).

Given the ease with which trifluoroleucine can be incorporated *in vivo*, a wide range of proteins can be prepared in fluorinated form as a means to improve stability. The method may prove to be quite general, as leucine is the most abundant of the amino acids in cellular proteins (9%) (46), and is especially important in determining the structure and stability of hydrophobic protein subdomains. A potential advantage of using fluorination to stabilize proteins is that it is complementary to other existing methods of protein stabilization: Fluorination might therefore serve as a "final push" toward higher stability after other methods, such as rational design (47) and directed evolution (48), have achieved some initial stabilization. Peptides that rely on hydrophobic side chains to form channels in membranes may also exhibit increased membrane (or inter-peptide) association upon fluorination (49). *In vivo* methods for incorporation of Tfl should allow fluorination of enzymes, signaling molecules, protein ligands, etc., and may prove to be of broad utility in the engineering of robust macromolecular assemblies.

Table 2.1 Binding free energies (ΔG^{BE}, kcal/mol) of Leu-GCN4-p1d and fluorinated dimers. ΔG^{BE} is the difference in energy (averaged over 800 ps of MD after equilibration) of solvated monomers and the solvated dimer each from separate SGB MD calculations (final solvation energies with PBF). ΔG^{BE} is quoted per mole of the monomer. % increase is the increase in ΔG^{BE} compared to the Leu-GCN4-p1d structure. Also shown is the % helicity of each peptide.

Structure	ΔG^{BE}	% increase	% helicity[b]
Leu-GCN4-p1d	**65.08**	**0**	**90.8**
[a]Close (4S, 4S)	93.75	44	84.3
Far (4R, 4R)	98.14	51	79.4
Mixed (4S, 4R)	99.20	52	81.1
Mixed (4R, 4S)	111.15	71	89.3
Tfl- average	**100.56**	**55**	**83.5**
Hfl-GCN4-p1d	**77.21**	**19**	**78.5**

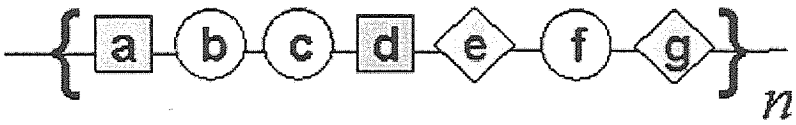
[a]Close, Far, Mixed: Configuration of the pair of trifluoromethyl groups as illustrated in Fig 4. **Tfl- average**: The averaged ΔG^{BE} of the four configurations. [b]Helicity quoted here has been calculated as the ratio of the residues with torsion angles ϕ and ψ in the helical region of the Ramachandran plot to the total number of residues in the protein.

Table 2.2 Components of ΔG^{BE} (kcal/mol) for Leu-GCN4-p1d and fluorinated dimers (quoted for one mole of the monomer)

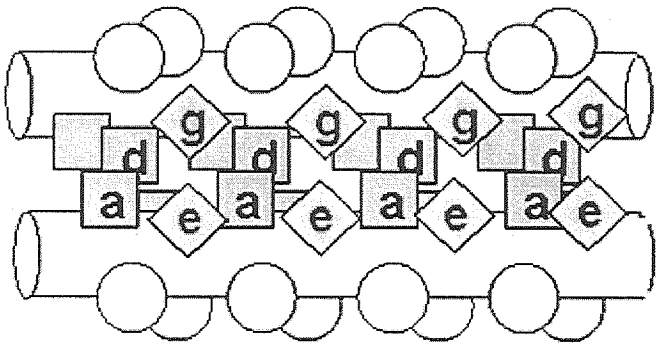
Structure	$\Delta G^{valence}$	$\Delta G^{coulomb+solvation}$	ΔG^{vdW}	ΔG^{Hbond}
Leu-GCN4-p1d	**-16.12**	**-16.66**	**41.82**	**56.05**
[a]Close (4S, 4S)	-8.46	-16.64	59.56	59.29
Far (4R, 4R)	-9.54	+1.73	42.16	63.80
Mixed1 (4S, 4R)	-5.36	-10.55	65.17	49.94
Mixed2 (4R, 4S)	-27.96	-1.79	48.29	92.61
Tfl- average	**-12.83**	**-6.82**	**53.80**	**66.41**
Hfl-GCN4-p1d	**-23.06**	**7.24**	**36.51**	**56.19**

[a]Close, Far, Mixed: Configuration of the pair of trifluoromethyl groups as illustrated in Fig 2.15. **Tfl- average**: The averaged ΔG of the four configurations.

A.



B.



C.

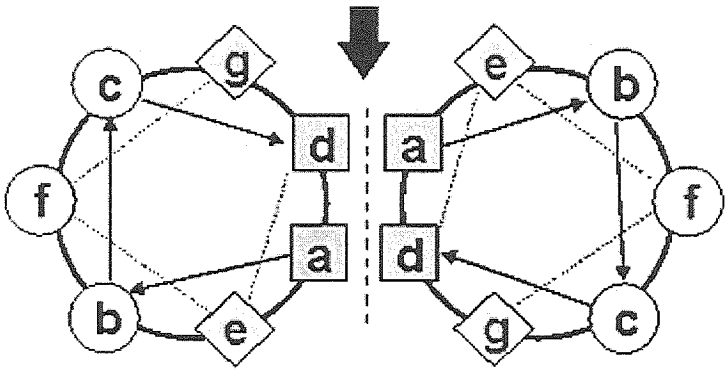


Figure 2.1. Three representations of a leucine zipper. **A.** The linear sequence of leucine zippers are characterized by repeats of heptads. The ***a*** and ***d*** positions are usually occupied by hydrophobic residues such as leucine, valine or isoleucine. The ***e*** and ***g*** positions are occupied by charged residues. GCN4-p1 contains four heptad repeats. **B.** A cartoon depicting the side view of a dimeric leucine zipper. **C.** The top view of a dimer leucine zipper. The hydrophobic surface is indicated with an arrow.

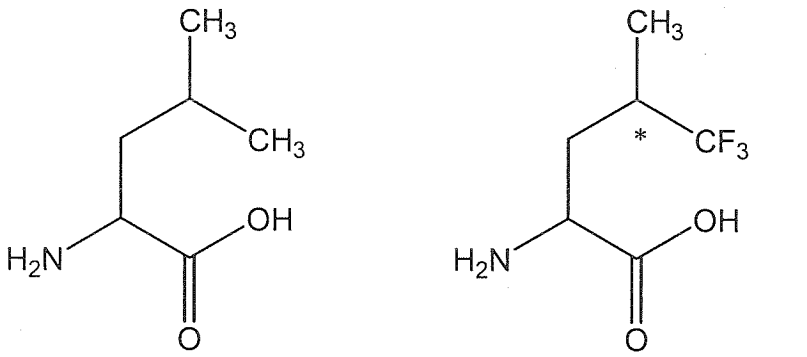
A.

R MKQ**L**EDK VEE**L**LSK NYH**L**ENE VAR**L**KKL VGER

B.

DPAALKRARNTEAARRSRARKLQ-*GCN4-p1*

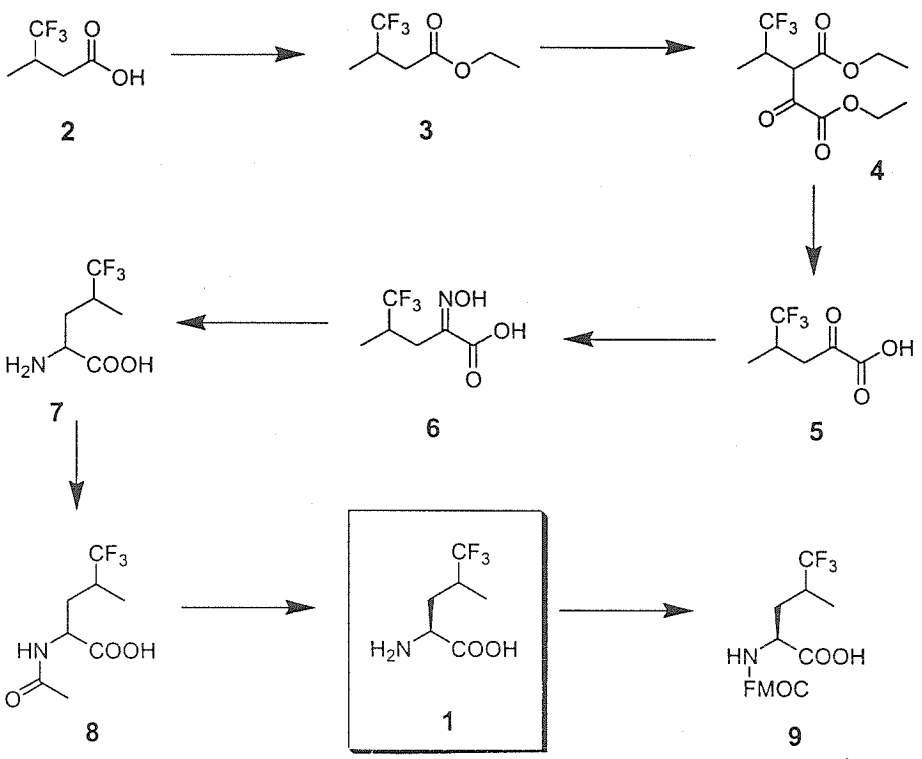
C.



Leucine **Trifluoroleucine**

Figure 2.2 (A) Amino acid sequence of GCN4-p1d. The amino acids in the leucine-zipper motif are represented by the heptad *abcdefg*. The leucine residues that were replaced by trifluoroleucine are at the ***d***-positions in the heptad repeat (shown in bold). The peptides prepared in this work were *not* acylated at the amino terminus. Amino acid abbreviations G: Glycine; A: Alanine; V: Valine; L: Leucine; I: Isoleucine; S: Serine; T: Threonine; P: Proline; D: Aspartic Acid; N: Asparagine; E: Glutamic Acid; Q: Glutamine; K: Lysine; R: Arginine; C: Cystein; M: Methionine; H: Histidine; W: Tryptophan; F: Phenylalanine; Y: Tyrosine. **(B)** Amino acid sequence of bZip. The GCN4-p1 domain is at the C-terminus of the protein. **(C)** Comparison between trifluoroleucine (right) and leucine. Trifluoroleucine was used as a mixture of the (2S, 4S) and (2S, 4R) isomers.

A.



B.

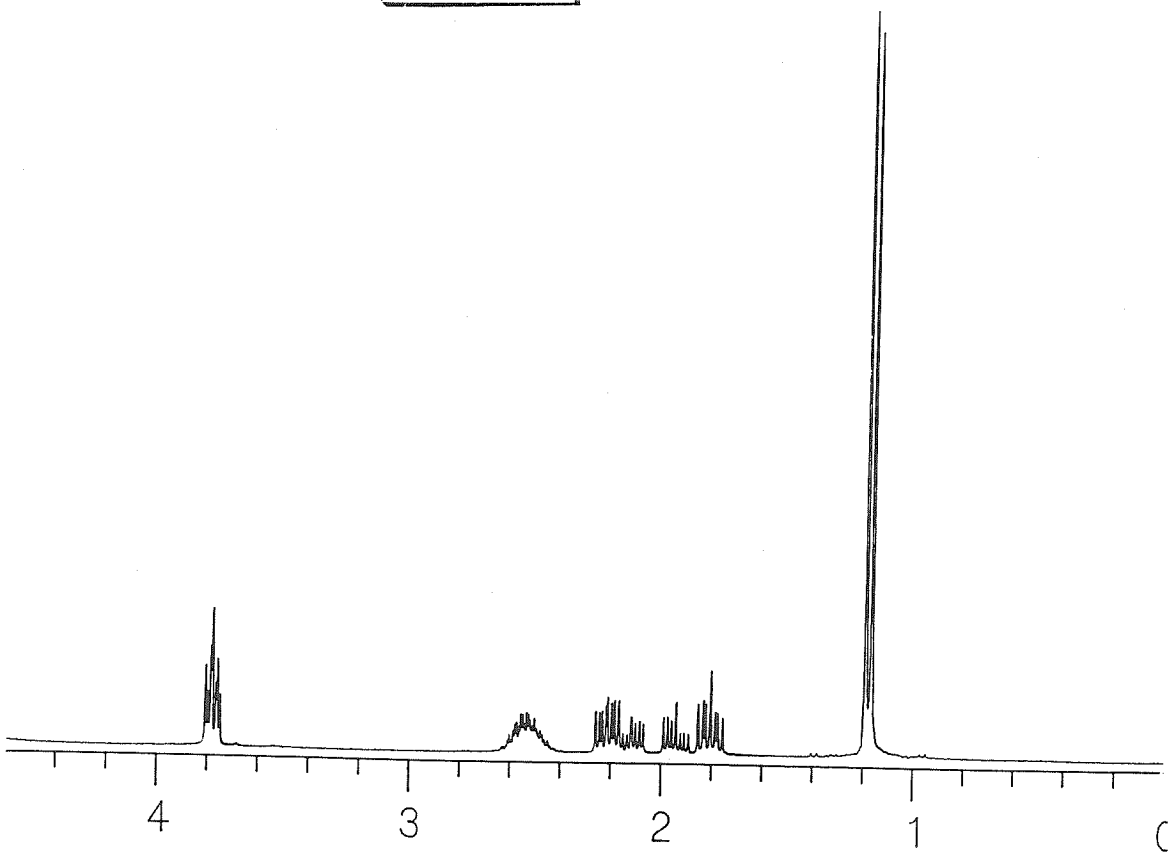


Figure 2.3 (A) Synthesis and Resolution of L-trifluoroleucine, **1**. **(B)** ^{1}H NMR of **1**.

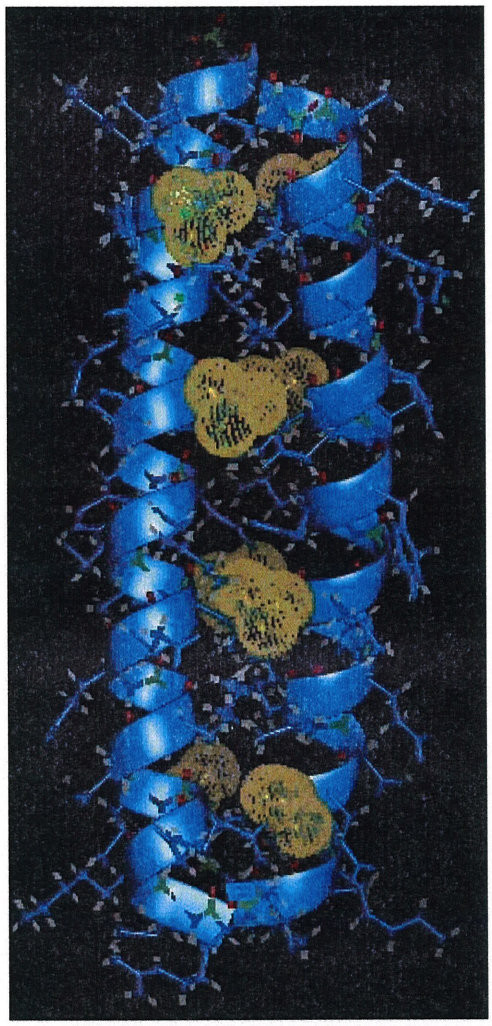


Figure 2.4 Three-dimensional representation of the dimeric form of GCN4-p1d substituted with trifluoroleucine at the four ***d***-positions in the helix. The van der Waals radii of the fluorine atoms are shown as yellow spheres.

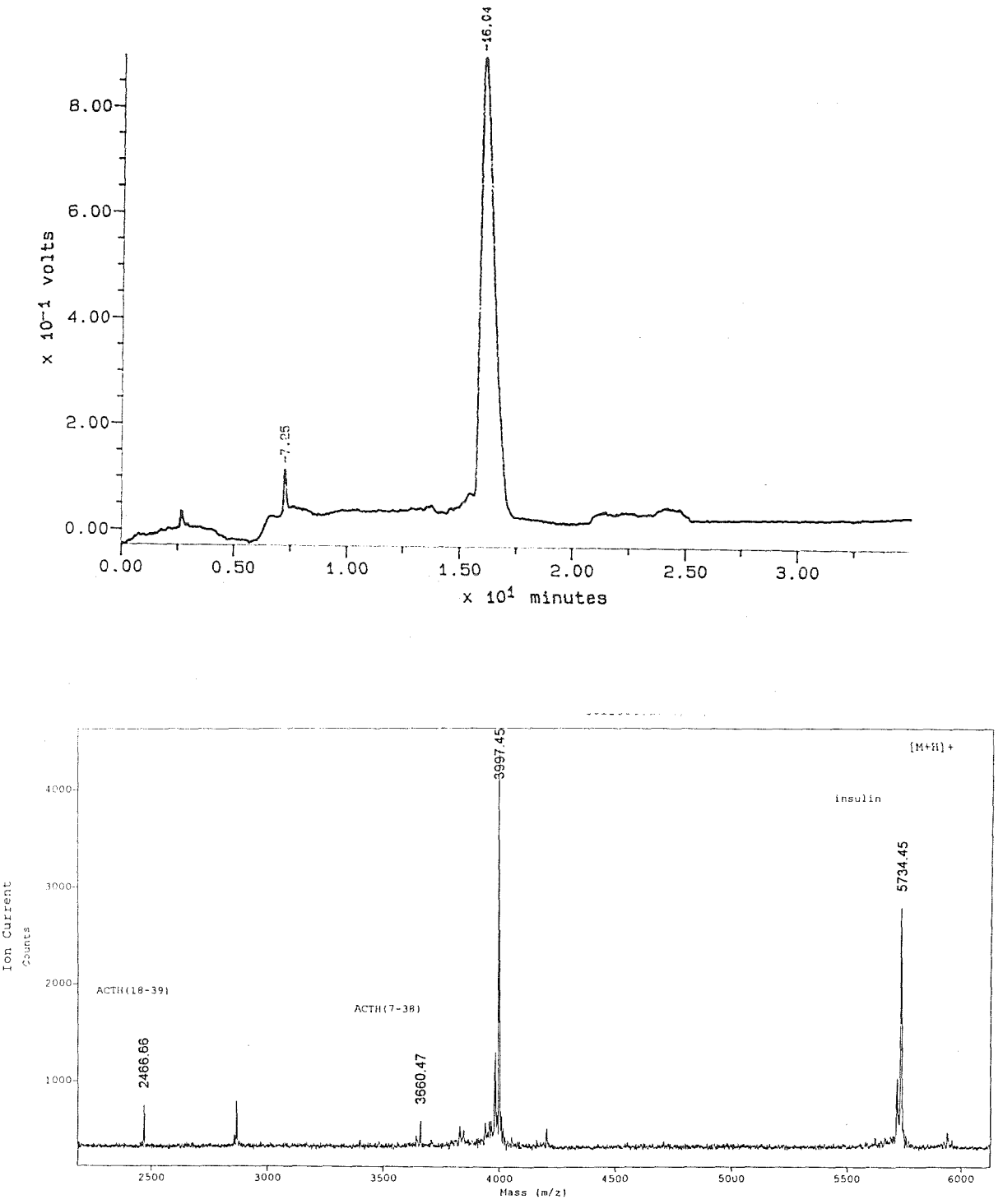


Figure 2.5. HPLC trace (upper) and MS analysis (bottom) of wild-type GCN4.

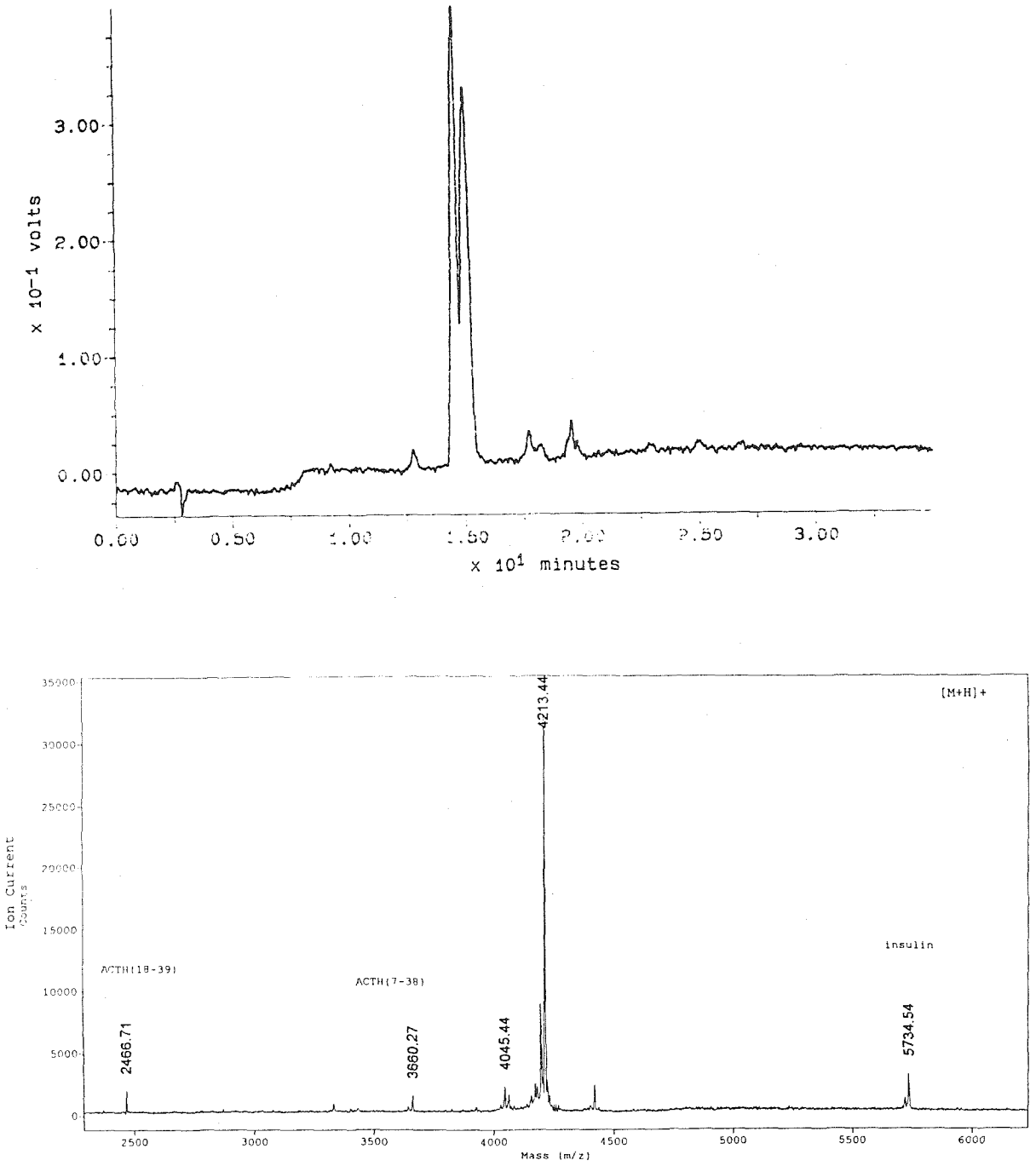
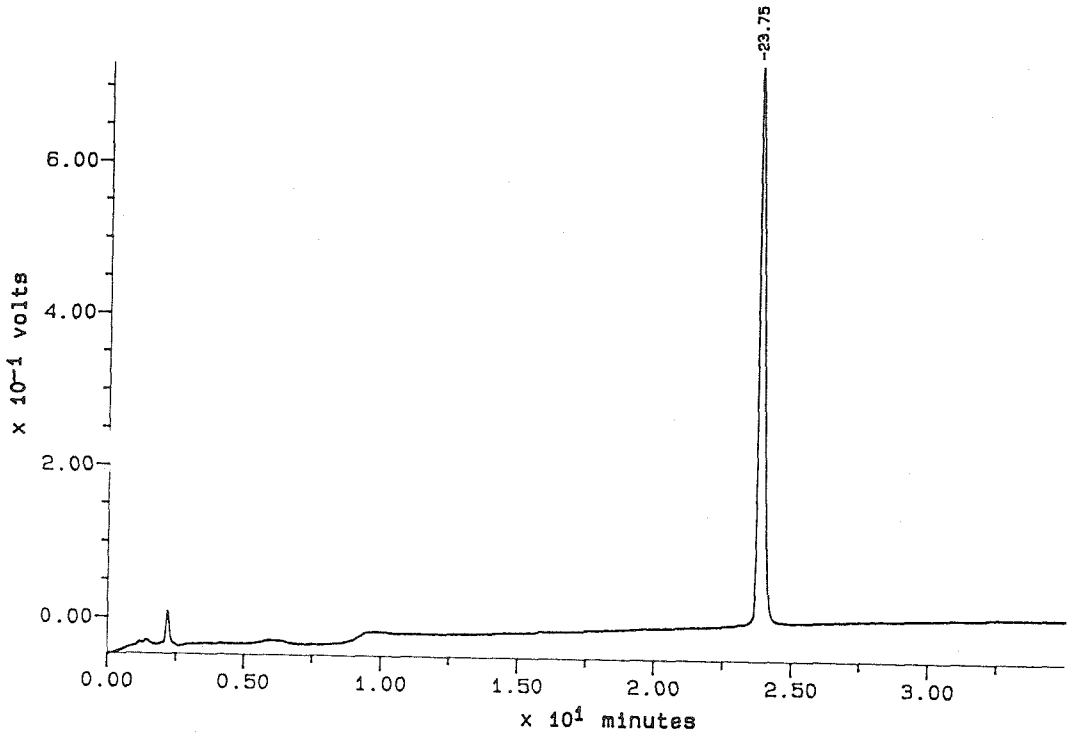


Figure 2.6. HPLC trace (upper) and MS analysis (bottom) of fluorinated GCN4. The two peaks are result of Tfl composition heterogeneity. The mass of the protein species correspond to each of the peaks are identical. Protein purified from the second peak is used in the structural and thermodynamic studies described in this work.



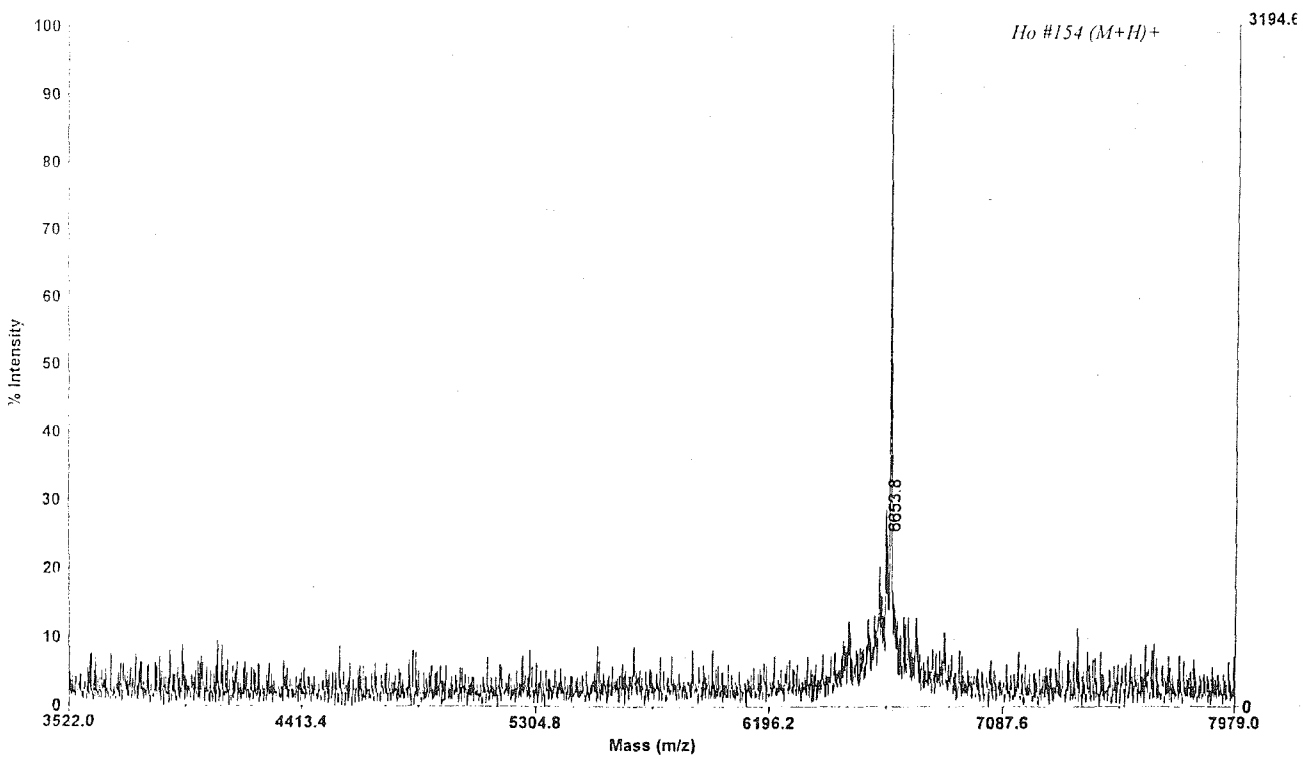


Figure 2.7 HPLC trace (upper) and MS analysis (bottom) of wild type bZip.

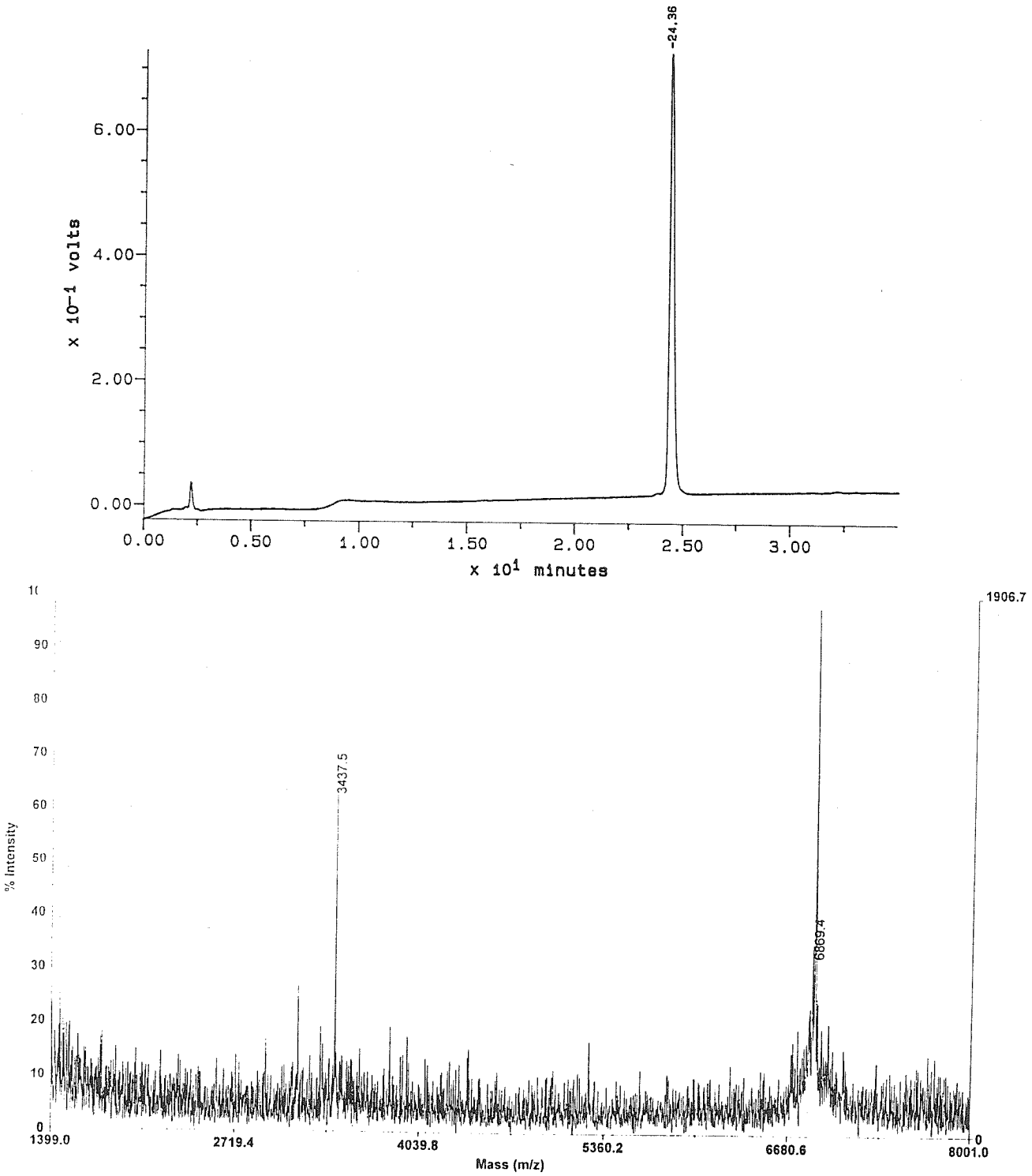


Figure 2.8 HPLC trace (upper) and MS analysis (bottom) of fluorinated bZip. Only one peak was observed during HPLC purification, possibly due to the diminishing effects of Tfl hetereogeneity on a larger protein.

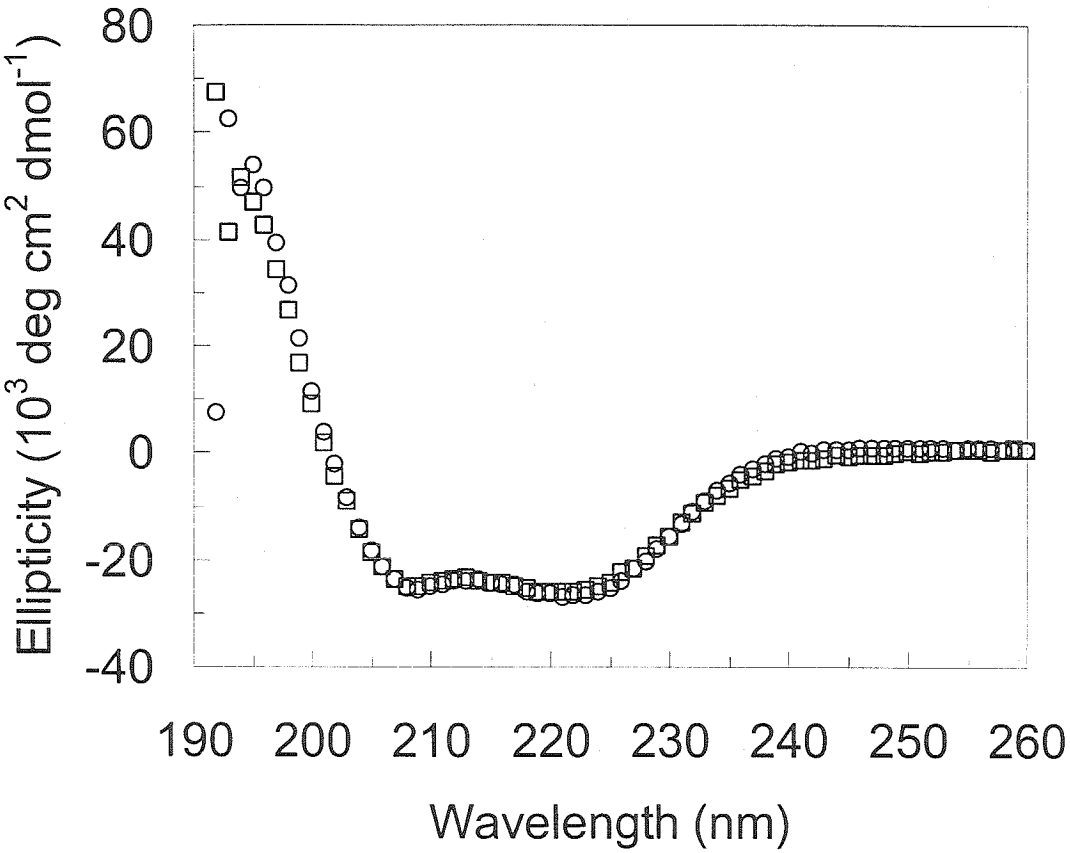


Figure 2.9 CD spectra of Leu-GCN4-p1d (□) and Tfl-GCN4-p1d (○) at 0°C and 30 μM. The nearly identical spectra and the magnitude of the intensity at 222 nm indicate that the peptides have similar secondary structures and are highly helical.

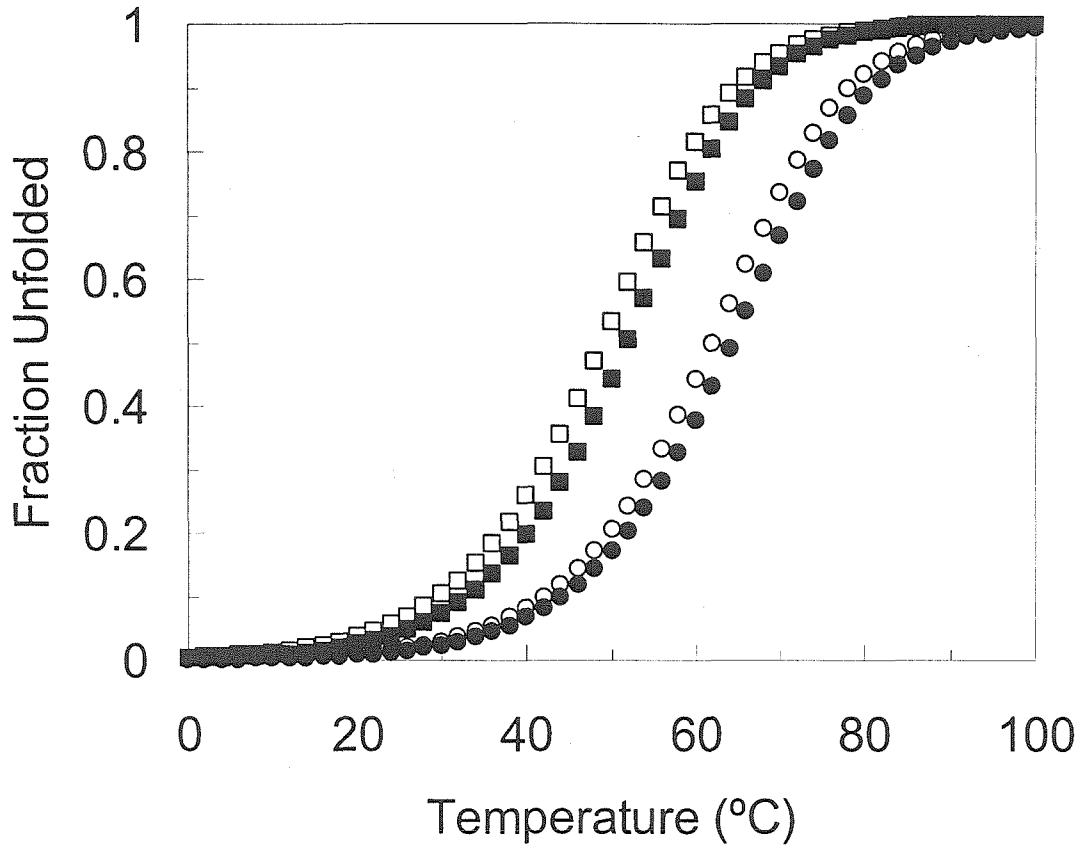


Figure 2.10 Thermal unfolding profiles for Leu-GCN4-p1d (squares) and Tfl-GCN4-p1d (circles) at 30 μM (open symbols) and 85 μM (closed symbols). The thermal melting temperature, T_m, is defined as the temperature at which 50% of the peptide is unfolded. At 30 μM, the values of T_m for Leu-GCN4-p1d and Tfl-GCN4-p1d are 49°C and 62°C, respectively. The thermal melting profiles were fitted globally to yield the thermodynamic quantities (see text and Ref. 20).

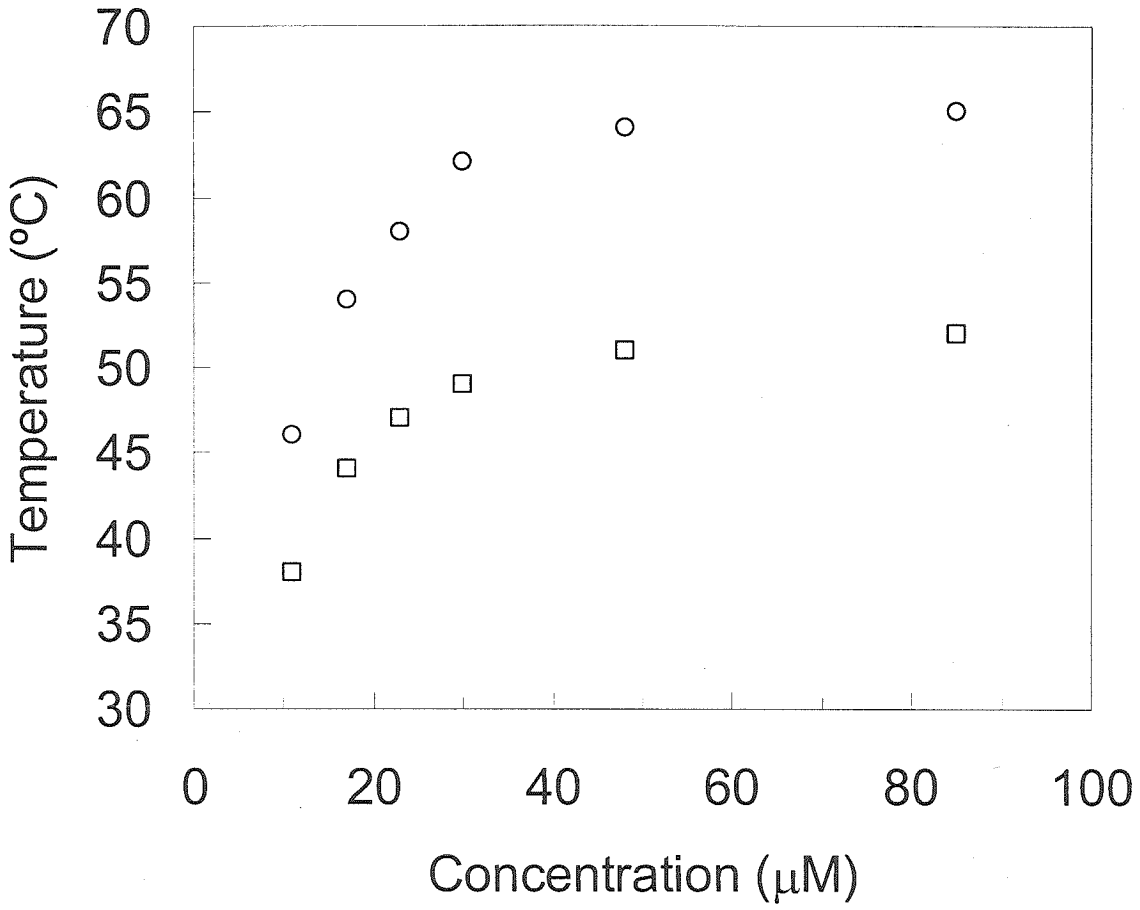


Figure 2.11 Concentration dependence of the thermal melting temperatures of Leu-GCN4-p1d (□) and Tfl-GCN4-p1d (○). The decrease in T_m at decreasing concentration is evidence of self-association.

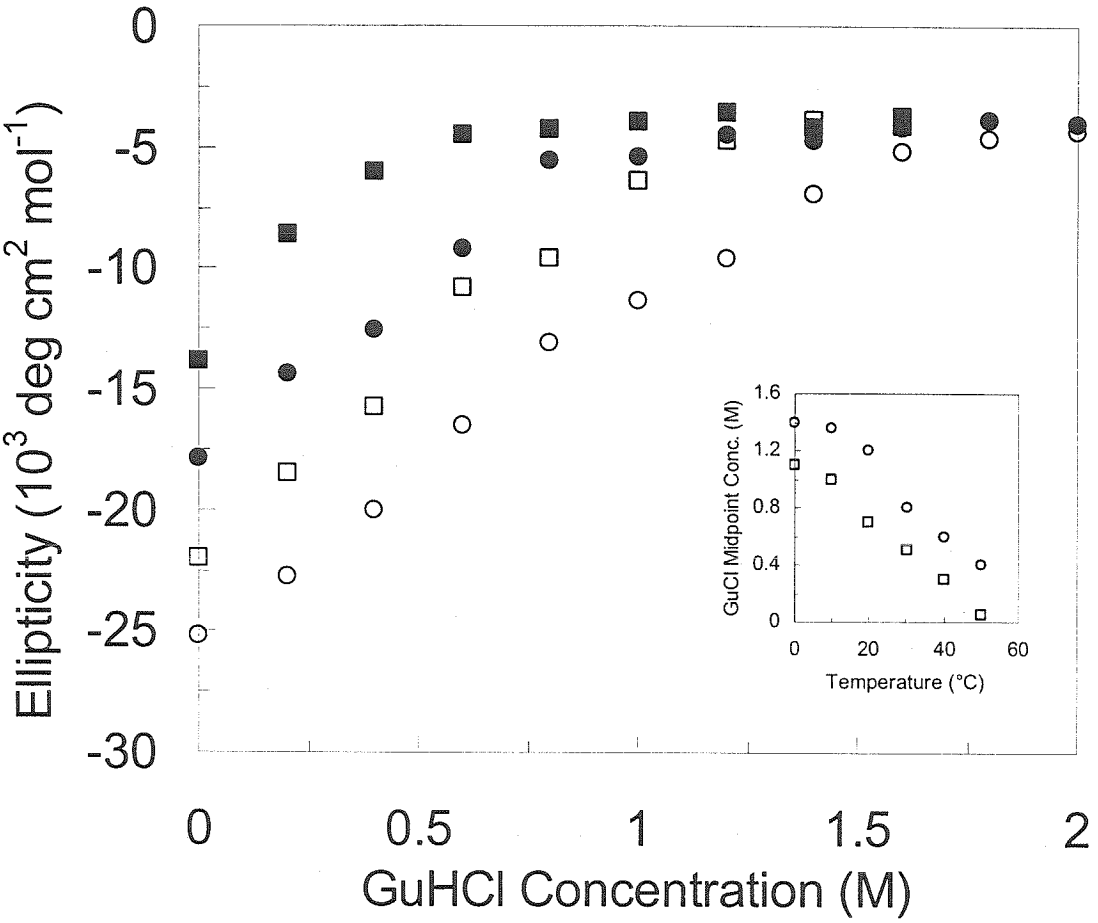


Figure 2.12 Guanidinium hydrochloride (GuHCl) titration of Leu-GCN4-p1d (squares) and Tfl-GCN4-p1d (circles). Data were recorded at 30°C (open symbols) and 50°C (closed symbols). The ellipticity was monitored at 222 nm at a peptide concentration of 30 μM. At each temperature, the fluorinated peptide is less susceptible to GuHCl denaturation than the hydrogenated form.

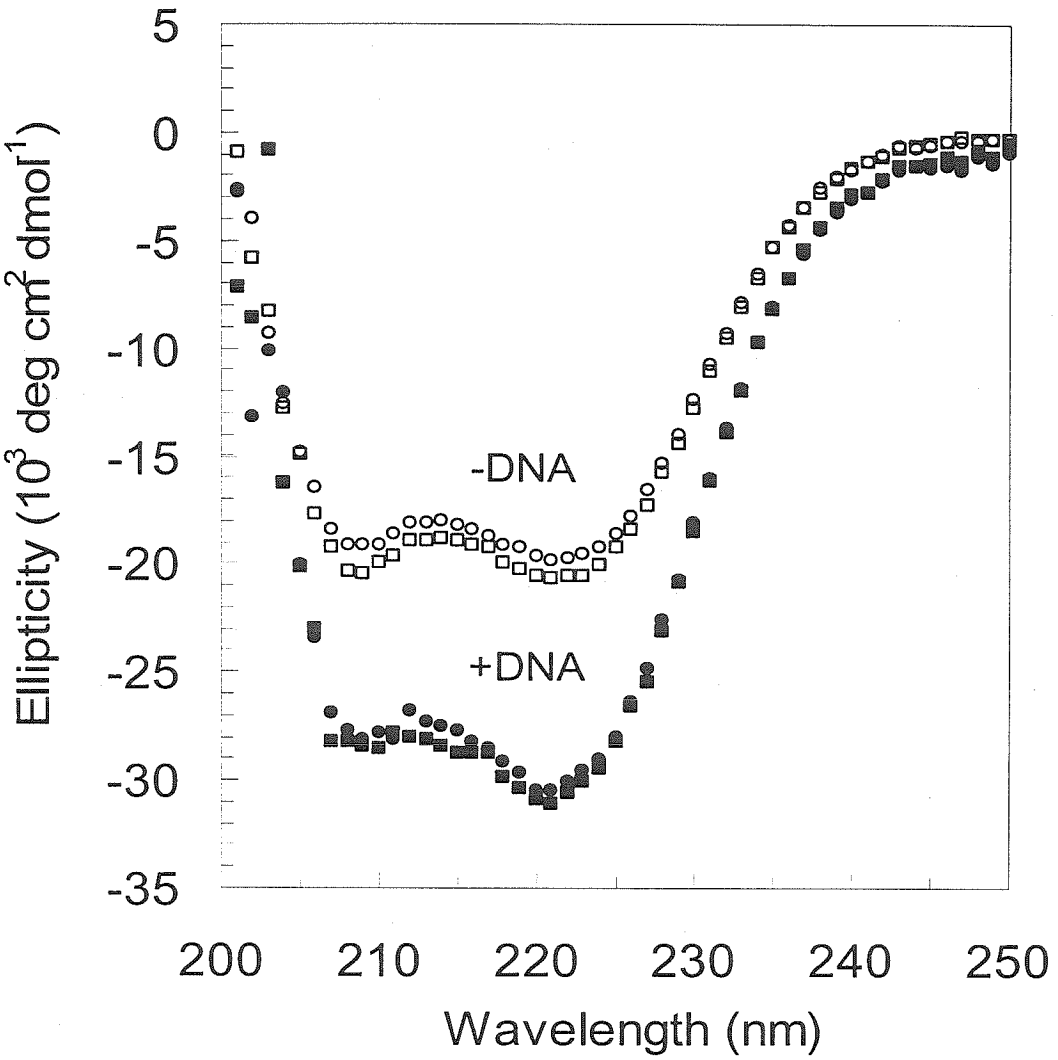


Figure 2.13 CD spectra for Leu-bZip (squares) and Tfl-bZip (circles) with (closed symbols) and without (open symbols) CREB at 0°C. The peptide helicity, as estimated from the molar ellipticity at 222 nm, is increased from approximately 70% (~ -20,000 10^3 deg cm^2 $dmol^{-1}$) to nearly 90% (~ -31,000 deg cm^2 $dmol^{-1}$). The DNA concentration is 5 μM and the protein concentration is 10 μM (PBS buffer, pH 7.4).

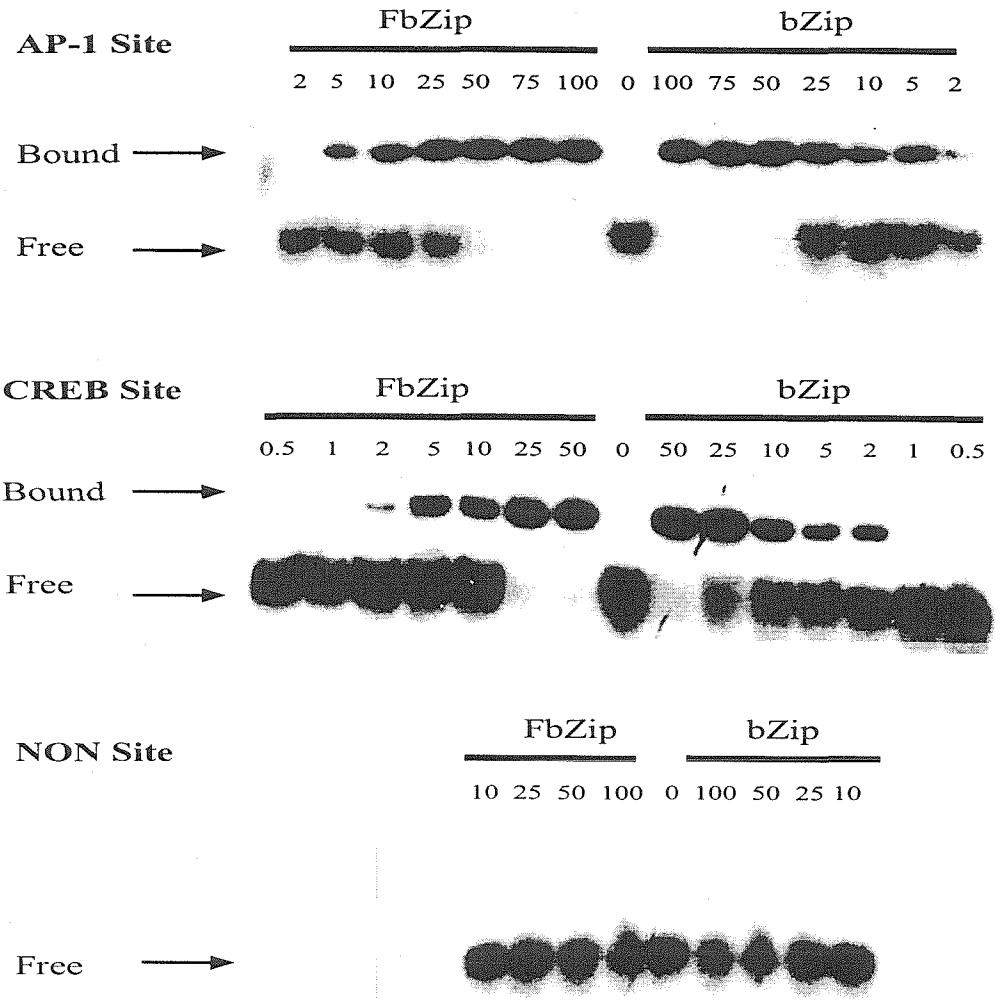


Figure 2.14 Mobility shift assay of Leu-bZip and Tfl-bZip binding to oligonucleotides containing the AP-1 binding site (5'-GTGGAGATGACTCATCTCCGG-3', top), the CREB binding site (5'-TGGAGATGA-CGTCATCTCCT-3', middle) and the nonspecific sequence (NON, 5'-GATCCCAACA-CGTGTTGGGATC-3', bottom). The protein concentrations (in nM) are indicated above each lane. Similar binding affinities were observed for Leu-bZip and Tfl-bZip. No binding to the nonspecific sequence could be detected.

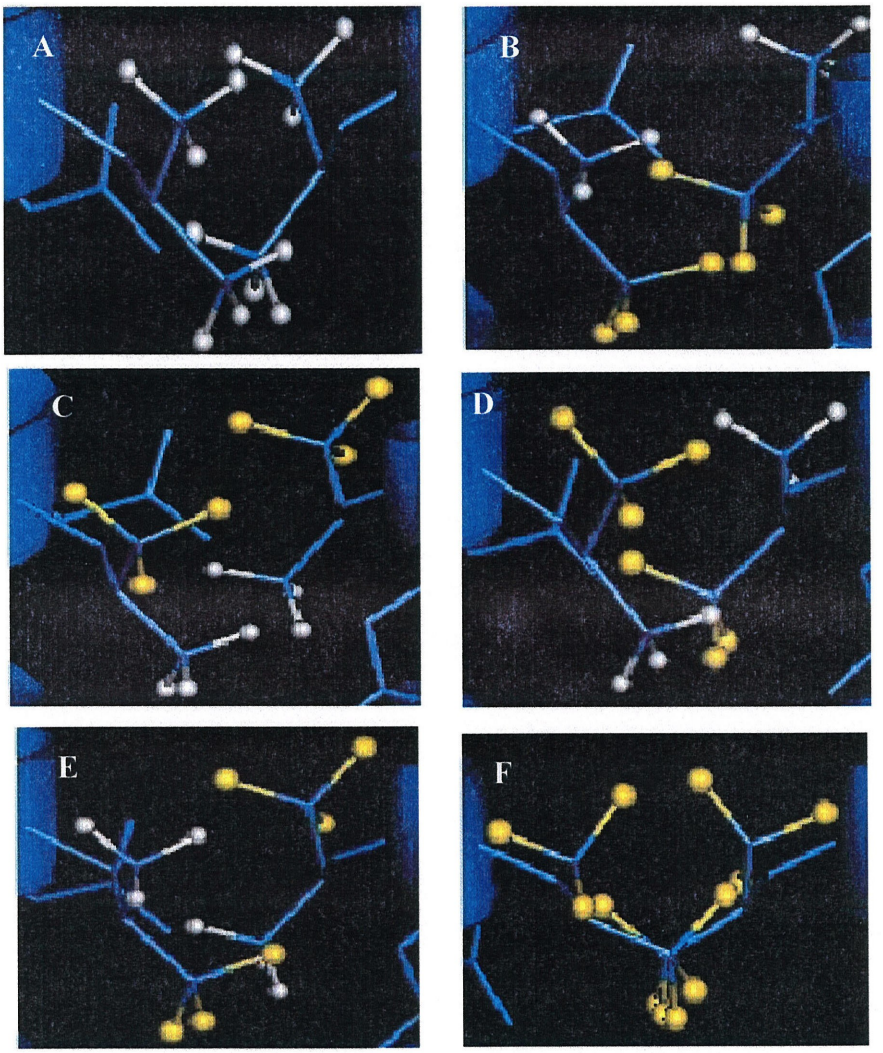


Figure 2.15 Possible configurations of the leucine/trifluoroleucine packing at each *d*- position in the heptad. Shown in the figure are residues Leu_{22} from both stands. Several combinations of the 4R- and 4S-isomers of trifluoroleucine are possible due to the side-chain stereocenter. Shown here are the various packing possibilities considered in the molecular dynamics simulations. **A**: Leu/Leu; **B**: 4S/4S, "close"; **C**: 4R/4R, "far"; **D**: 4R/4S; **E**: 4S/4R; and **F**: Hfl/Hfl, where both leucines are replaced by hexafluoroleucine.

2.5 References

1. Cleland, J. L., and Craik, C. S. (1996) *Protein Engineering: Principles and Practice*, Wiley-Liss, New York, NY.
2. Mendel, D., Ellman, J. A., Chang, Z. Y., Veenstra, D. L., and Kollman, P. A. (1992) *Science 256*, 1798.
3. Matthews, B. W. (1995) *Adv. Protein Chem. 46*, 249.
4. Fersht, A. R., and Serrano, L. (1995) *Curr. Opin. Struct. Biol. 3*, 75.
5. Cornish, V. W., Mendel, D., and Schultz, P. G. (1995) *Angew. Chem. Int. Ed. Engl. 34*, 621.
6. van Hest, J. C. M., and Tirrell, D. A. (1998) *FEBS Lett. 428*, 68.
7. Duewel, H., Daub, E., Robinson, V., and Honek, J. F. (1997) *Biochemistry 36*, 3404.
8. van Hest, J. C. M., Kiick, K. L., and Tirrell, D. A. (2000) *J. Am. Chem. Soc. 122*, 1282.
9. Kiick, K. L., van Hest, J. C. M., and Tirrell, D. A. (2000) *Angew. Chem. Int. Ed. 39*, 2148.
10. Sharma, N., Furter, R., and Tirrell, D. A. (2000) *FEBS Lett. 467*, 37.
11. a. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989) *Science 243*, 538. b. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) *Science 254*, 539. c. Lumb, K. J., and Kim, P. S. (1995) *Science 268*, 436. d. O'Shea, E. K., Lumb, K. H., and Kim, P. S. (1993) *Curr. Biol. 3*, 658
12. a. Hodges, R. S. (1996) Biochem. Cell. Biol. 74 133-54. b. Yu, Y., Monera, O. D., Hodges, R. S., and Privalov, P. L. (1996) *J. Mol. Biol. 255*, 367
13. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) *Science 262*, 1401.
14. Wendt, H., Berger, C., Baici, A., Thomas, R. M., and Bosshard, H. R. (1995) *Biochemistry 34*, 4097.
15. Krylov, D., Mikhailenko, I., and Vinson, C. (1994) *EMBO J. 13*, 2849.
16. Schneider, J. P., Lear, J. D., and DeGrado, W. F. (1997) *J. Am. Chem. Soc. 119*, 5742.
17. Moran, L. B., Schneider, J. P., Kentsis, A. Reddy, G. A. and Sosnick, T. R. (1999) *Proc. Natl. Acad. Sci. USA 96*, 10699.
18. Myers, J. K., and Oas, T. G. (1999) *J. Mol. Biol. 289,* 205.
19. Thompson, K. S., Vinson, C. R., and Freire, E. (1993) *Biochemistry 32*, 5491.

20. a. Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. E., and Matthews, C. R. (1995) *Biochemistry 34*, 12812. b. Zitzewitz, J. A., Ibarra-Molero, B., Fishel, D. R., Terry, K. L., and Matthews, C. R. (2000) *J Mol Biol 296*, 1105.

21. a. Moitra, J., Szilak, L., Krylov, D., and Vinson, C. (1997) *Biochemistry 36*, 12567. b. Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T., Hodges, R. S. (2000) *J. Mol. Biol. 300,* 377.

22. Arndt, K., and Fink, G. R. (1986) *Proc. Natl. Acad. Sci. USA 83*, 8516.

23. Aizawa, Y., Sugiura, Y., Ueno, M., Mori, Y., Imoto, K., Makino, K., and Morii, T. (1999) *Biochemistry 38*, 4008.

24. Hockings, S. C., Kahn, J. D., and Crothers, D. M. (1998) *Proc. Natl. Acad. Sci. USA 95*, 1410.

25. Gough, C. A., Pearlman, D. A., and Kollman, P. (1993) *J. Chem. Phys. 99*, 9103.

26. Hine, J., and Mookerjee, P. K. (1975) *J. Org. Chem. 40*, 292.

27. Kukhar, V. P., and Soloshonok, V. A. (1995) *Fluorine Containing Amino Acids - Synthesis and Properties*, John Wiley & Sons, Chichester.

28. Rennert, O. M., and Anker, H. S. (1963) *Biochemistry 2*, 471.

29. Kothakota, S., Dougherty, M. J., Fournier, M. J., Mason, T. L., and Tirrell, D. A. (1995) *Macromol. Symp. 98*, 573.

30. Tang, Y., Petka, W., and Tirrell, D. A. (2000) *J. Am. Chem. Soc. Manuscript Submitted.*

31. Harding, S. E., Rowe, A. J., and Harton, J. C. (1992) *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, The Royal Society of Chemistry, Cambridge.

32. Brooks, I. S., Soneson, K. K., and Hensley, P. (1993) *Biophys. J. 64*, 244.

33. Lim, K. T., Brunett, S., Iotov, M., McClurg, B., Vaidehi, N., Dasgupta, S., Taylor, S., and Goddard III, W. A. (1997) *J. Comp. Chem. 18*, 501.

34. Ding, H. Q., Karasawa, N., and Goddard III, W. A. (1992) *J. Chem. Phys. 97*, 4309.

35. Tannor, D. J., Marten, B., Murphy, R., Friesner, R. A., Sitkoff, D., Nicholls, A., Ringnalda, M., and Goddard III, W. A. (1994) *J. Am. Chem. Soc. 116*, 11875.

36. Ghosh, A., Rapp, C. S., and Friesner, R. A. (1998) *J. Phys. Chem. B. 102*, 10983.

37. Mayo, S. L., Olafson, B. D., and Goddard III, W. A. (1990) *J. Phys. Chem. 94*, 8897.

38. Greeley, B. H., Russo, T. V., Mainz, D. T., Friesner, R. A., Langlois, J.-M., Goddard III, W. A., Donnelly, R. E., and Ringnalda, M. N. (1994) *J. Chem. Phys. 101*, 4028.

39. Mackerell, A. D., Wiorkiewicz-Kuczera, J., and Karplus, M. J. (1996) *J. Am. Chem. Soc. 117*, 11946.

40. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990) *Nature 347*, 575.

41. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell 71*, 1223.

42. Konig, P., and Richmond, T. J. (1993) *J. Mol. Biol. 233*, 139.

43. Keller, W., Konig, P., and Richmond, T. J. (1995) *J. Mol. Biol. 234*, 657.

44. Metallo, S. J., and Schepartz, A. (1994) *Chem. Biol. 1*, 143.

45. Paolella, D. N., Palmer, C. R., and Schepartz, A. (1994) *Science 264*, 1130.

46. Creighton, T. (1997) *Proteins: Structure and Molecular Properties*, W. H. Freeman and Company, New York.

47. DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F., and Lombardi, A. (1999) *Annu. Rev. Biochem. 68*, 779.

48. Giver, L., Gershenson, A., Freskgard, P. O., and Arnold, F. H. (1995) *Proc. Natl. Acad. Sci. USA 95*, 12809.

49. Roux, M., Nezil, F., Monck, M., and Bloom, M. (1994) *Biochemistry 33*, 307.

CHAPTER 3

Biosynthesis of Proteins Containing Trifluoroleucine and Hexafluoroleucine

This chapter appeared as communications in 1) Tang, Y., Ghirlanda, G., W. A. Petka, Nakajima, T., DeGrado, W. F. and Tirrell, D. A. *Angew. Chem. Intl. Ed.* (**2001**) *40*, 1494-1497; and 2) Tang, Y., Tirrell, D. A., *J. Am. Chem. Soc.* (**2001**) *123*, 11089-11090.

3.0 Abstract

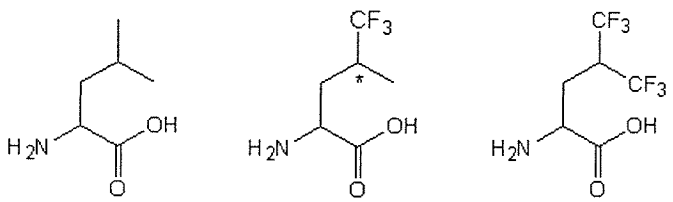
To make fluorination a general method of stabilizing protein structures, we studied *in vivo* incorporation of trifluoroleucine (Tfl, **2**) and hexafluoroleucine (Hfl, **3**) in place of leucine using leucine auxotrophic *E. coli* strains. The target protein is A1, which is a leucine zipper protein that has 74 residues and eight leucines. The leucine residues are buried at the dimer interface and stabilize the protein complex. Tfl supported protein synthesis efficiently and replaced up to 92% of leucines in the protein under normal expression conditions. The yield of fluorinated protein was reduced from 40 mg/L to 20 mg/L. We were able to tune the level of fluorination by altering the concentration of competing leucine in culture media. Tfl-A1 adopted the identical helical secondary structure and dimeric aggregation order. T_m of Tfl-A1 was elevated to 67°C, a 13°C increase over A1. The concentration of urea needed to denature 50% (C_m) of protein was elevated from 2.7 M to 7 M. In contrast to Tfl incorporation, the more hydrophobic amino acid Hfl did not support protein synthesis under similar conditions. From *in vitro* characterization of leucyl-tRNA synthetase (LeuRS) substrate specificity, Hfl was shown to be activated 4100 times slower than leucine (compared to the 240-fold rate attenuation of Tfl). The decreased rate of $tRNA^{Leu}$ aminoacylation by Hfl resulted in insufficient amounts of Hfl-$tRNA^{Leu}$ during protein synthesis. We raised the cellular LeuRS activity eightfold at the time of protein induction by overexpressing LeuRS under a constitutive promoter during cell growth. Under these conditions, Hfl was effectively incorporated into A1 at ~80% substitution rate. The presence of the nearly perfluorinated side chains in the protein core enhanced protein stability even further. T_m was increased to 76°C and ΔG_f decreased by 3.6 kcal/mol. More remarkably, C_m of Hfl-A1 was not observed within the urea solubility limit. Our results suggest that fluorination could be extended to stabilizing recombinant proteins.

3.1 Introduction and Background

In the previous chapter, we demonstrated that fluorination is an attractive approach to strengthening the hydrophobic driving force of protein self-assembly (1). We constructed fluorinated GCN4-p1 peptides with optically resolved 5,5,5-L-trifluoroleucine using solid phase peptide synthesis. To extend and generalize the fluorination approach to target larger proteins, including enzymes, growth factors and other industrially and pharmaceutically pertinent proteins, we need to develop techniques that allow the expression of fluorinated proteins *in vivo*. To this end, we will attempt to optimize *E. coli* expression conditions so that fluorinated amino acids can be utilized as alternative building blocks to the twenty proteinogenic amino acids. Incorporation of noncanonical amino acids provides a route to proteins and peptides with unique structural and chemical features (2-7). Functionality orthogonal to that of the naturally occurring amino acids, including alkenes (8), alkynes (9), aryl halides (10), and electroactive side chains (11) has been incorporated efficiently into proteins prepared in bacterial cultures.

In this chapter, we will discuss methods that allow the incorporation of trifluoroleucine (**2**) and hexafluoroleucine (**3**) into recombinant proteins in place of the natural amino acid leucine (**1**). Our objectives for these studies are threefold. First, we want to examine the recognition of the gatekeeping enzyme, leucyl-tRNA synthetase (LeuRS) towards these abiotic substrates *in vivo*, and the efficiencies of the ribosomal machinery in processing $tRNA^{Leu}$ appended with these amino acids. Second, for sluggish substrates of LeuRS, we will examine methods to improve their translational efficiencies. Third, we are interested in demonstrating the effects of fluorination on a larger, recombinant protein and the impact of inserting the more hydrophobic amino acid, **3** on the structure and stability of the protein.

Scheme 3.1



1 **2** **3**

Translational incorporation of an amino acid analogue *in vivo* requires that the analogue serve as a substrate for one of the aminoacyl-tRNA synthetases (aaRS) of the host. Several strategies have been employed to address this requirement, including directed evolution (12) and rational design (13) of aaRS that exhibit enhanced activity toward nonnatural substrates. Fluorinated amino acids are ideal candidates for incorporation because the small size of the fluorine atom renders such analogues nearly isosteric to their hydrocarbon counterparts (14). A number of fluorinated amino acids have been introduced into proteins in conventional *E. coli* expression strains (15-17). Trifluoroleucine has been shown to be active towards the *E. coli* biosynthetic machinery. It was reported over 30 years ago that **2** was able to support bacteria cell growth and was incorporated into cellular proteins in the absence of **1** (18). Our laboratory has been interested in **2** because of its unique side-chain physical properties as well as its enhanced hydrophobicity. Kothakota and coworkers have demonstrated that **2** can be efficiently incorporated into a recombinant, artificial polypeptide that adopts an ordered lamella-like fold (19). The trifluoromethyl groups emerged at the surface of the material, and significantly decreased the surface energy of the polymer film. The wetting properties of the fluoropolymer were similar to that of Teflon.

Unlike **2**, the *in vivo* properties of the highly fluorinated amino acid **3** has not been studied. The nearly perfluorinated **3** contains several attractive features over **2**. Most importantly, its side chain does not contain a stereocenter at the γ position, thus eliminating the composition heterogeneities associated with proteins substituted with **2**. Furthermore, the more hydrophobic side chain of **3** may enhance protein-folding driving force further and render substituted proteins more robust. We anticipated before the *in vivo* studies that **3** is also an efficient substrate of LeuRS, because both the 4S and 4R stereoisomers of **2** were incorporated at comparable levels in the artificial fluoropolymer, as studied by ^{19}F NMR spectroscopy (19). However, as experiments in this study show, **3** is a much poorer substrate of LeuRS and its incorporation during translation requires the elevation of intracellular LeuRS activities.

The target protein used in this study is an artificial leucine zipper protein. Leucine-zipper domains form coiled-coil structures comprising generic heptad repeats designated *abcdefg,* where the ***d*** positions are occupied predominantly by leucine residues (20-27). The protein is named A1 and its sequence and helical wheel representation are shown in Figure 3.1. The design of this protein has been discussed elsewhere (28, 29). A1 was initially used as the "Velcro" domain in a reversible hydrogel material because of its self-

assembling properties. A1 contains 74 residues and 8 leucines. A 6xHis leader sequence has been engineered at the N-terminus of the protein to facilitate purification. Seven of the eight leucine residues are located at the "*d*" positions of heptad repeats and contribute strongly to the stability of the dimeric complex. We will replace **1** with either **2** or **3** at these positions and observe the consequent changes in protein structure and stability.

Because we are interested in understanding the effects of multisite substitution on our target protein, we will use residue-specific techniques to introduce the unnatural amino acids. We will adopt the "media-shift" procedure to deplete the medium of the natural substrate **1**, and coerce the cells to use either **2** or **3** in place of **1** during protein translation. The procedure, which has been used in the incorporation of selenomethioine for X-ray crystallography studies (30, 31), is shown in Figure 3.2. A leucine auxotrophic strain is used as the expression host to minimize the intracellular concentration of leucine during protein expression. Cells are first grown in supplemented minimal medium for three to four hours before induction. Prior to IPTG addition, the cells are collected by centrifugation and are washed with dilute saline solutions. The washing steps are repeated three times to remove residual leucine. Cell pellets are then resuspended in fresh M9 medium without leucine and the culture is distributed to different aliquots. Each aliquot is supplemented with either **1**, **2** or **3** and is allowed to express A1 for three hours. The accumulation of target protein can then be visualized with SDS-PAGE. Unnatural amino acids compatible with the biosynthetic apparatus of *E. coli* should support detectable amounts of protein expression. The composition of protein products can be examined with amino acid analysis or mass spectrometry to verify the presence of the new building block.

3.2. Material and Methods

3.2.1. Synthesis of Hexafluoroleucine

The synthetic scheme of **3** (Figure 3.3) is a modified version of one reported by Zhang et al (32). An alternative procedure was reported by Xing et al (33). Procedures for steps a, b and e are adopted from Ref. 32. The procedure for α-oxime ester reduction (step d) is adopted from Hoffman et al (34). The overall yield of synthesis is 12%. The hydrochloride salt of **3** is used directly in both *in vivo* incorporation

studies and *in vitro* activation assays. ^{1}H-NMR of **3** hydrochloride (CD_3OD): 4.1 (dd, 1H), 3.9 (m, 1H), 2.4 (m, 1H), 2.25 (dt, 1H). ^{19}F of **3** (10% D_2O, referenced to CF_3COOH): 7.96 (q, J=10), 7.72 (q, J=10).

3.2.2. E. coli Strains and Expression Plasmids

Cloning was performed with XL-1 ultracompetent *E. coli* cells (Stratagene). Biosynthesis of the target coiled-coil protein was completed in a leucine auxotrophic variant of *E. coli* strain mc1000 (F^- lacΔx74anaD139/(Ana Abioc-leu)Δ7679 gal U gal K rspL), designated LAM1000. LAM1000 was provided by Prof. M. J. Fournier of the University of Massachusetts, Amherst. The initial plasmid for this study, pQEA1 was constructed by Wendy Petka (29). The plasmid encodes the gene of A1 in its polylinker region in frame with the start codon present in the parent plasmid, pQE9 (Qiagen). The LeuRS overexpression plasmid discussed in section 3.2.8 is renamed as pA1EL. The repressor plasmid pREP4 was cotransformed into the expression strain with either pQEA1 or pA1EL to ensure minimal expression before IPTG induction. The plasmid encoding *E. coli* LeuRS (pTrc-LeuS2) was generously provided by Prof. E. Wang at Shanghai Institute of Biochemistry (35).

3.2.3. Small-Scale Analog Incorporation Assay

3.2.3.1. Protein Expression: LAM1000 cells are transformed with pQEA1 and pREP4 to yield the expression strain LAM1000/pQEA1/pREP4. Analogs **2** and **3** were assayed for translational efficiency using a medium shift technique. M9 medium (40 mL) supplemented with all 20 amino acids (40 mg/L), 1 mM $MgSO_4$, 1 mM $CaCl_2$, 0.4 wt% glucose, 1 mg/L thiamine and the antibiotics ampicillin (200 mg/L) and kanamycin (25 mg/L) was inoculated with 100 μl of fresh overnight culture (M9) of the expression strain. After the culture had grown to OD_{600} of 1.0, the cells were pelleted and washed with cold 0.9% NaCl three times. The cell pellet was then resuspended in 40 mL M9 medium (as described above but without leucine). IPTG (1 mM) was added to the culture to induce A1 expression. After ten minutes of incubation at 37°C, the culture was distributed into four 10 mL aliquots. The first aliquot was left unsupplemented and serves as a negative control. The remaining three aliquots were supplemented with either **1**: 150 μM; **2**: 220 μM; **3**: 170 μM. Cells were collected after 3 hr by centrifugation (15000 g, 2 min, 4°C). The pellets

were resuspended in 600 μl of Buffer A (8 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris, pH 8.0) and subjected to two cycles of freezing and thawing. Proteins were detected by western blotting of the whole cell lysate with an antibody specific for the N-terminal His-Tag of A1. The whole cell lysate was then subjected to purification by Ni-NTA spin columns (Qiagen) according to the manufacturer's instructions.

3.2.3.2. Composition Analysis of Intact Protein: The eluent from the spin column was dialyzed against water extensively and lyophilized to a fluffy powder. The powder was sent directly for MALDI measurement and amino acid analyses (Cornell University Biopolymer Analysis Facility).

3.2.3.3. MALDI of Tryptic Fragments: The eluent from spin column purification was in Buffer A, pH 4.5. 450 μL of 50 mM $(NH_4)_2CO_3$ was added to 50 μL of eluent to bring the pH to optimal trypsin working pH (8.0). 5 μL of trypsin stock solution (20 μg/200 μL) was added and the sample was incubated at room temperature overnight. The reaction was quenched by addition of 2 μL of TFA. The reaction mixture was subjected to C18 ZipTip (Millipore) purification and eluted with 3 μL of 0.1 TFA, 50% CH_3CN. 1 μL of eluent was used for the MALDI analysis (α-cyano-β-hydroxycinnamic acid, 10 mg/mL in 1:1 H_2O/CH_3CN).

3.2.4. LeuRS Purification

Competent HB101 cells were transformed with pTrc-LeuS2 (35), which encodes *E. coli* LeuRS under a *lac* promoter, to yield the LeuRS expression strain HB-LEU. A 1 L culture of HB-LEU was grown in LB medium at 37°C to OD_{600} of 0.5, 0.4 mM of IPTG was added at this time to induce protein expression. The culture was grown for an additional 18 hours. Cells were harvested by centrifugation (5000 g, 4°C, 15 min) and resuspended in buffer B (5 mL per 1 g wet cell mass, 10 mM KH_2PO_4, pH 6.8, 10 mM β-mercaptoethanol, 2 mM PMSF, 10% glycerol). Cells were lysed with ultrasonification and cell debris was removed with ultracentrifugation (120,000 g, 4°C, 90 min). The supernatant was applied to a DEAE-sepharose CL-6B column equilibrated with buffer B. LeuRS was partially purified using a linear gradient of the same buffer containing up to 500 mM KH_2PO_4, pH 6.8. The LeuRS fractions were collected, desalted on a PD-10 column and concentrated. The concentrated solution was applied to a HA-

Ultragel column equilibrated with buffer B. A linear gradient of buffer B (20 mM to 500 mM KH_2PO_4) was applied to the column. LeuRS fractions were pooled, buffer exchanged to storage buffer (10 mM KH_2PO_4, 2 mM β-mercaptoethanol, 50% glycerol), and flash frozen with liquid nitrogen in 200 μL aliquots. LeuRS concentration was calculated with an extinction coefficient of 1698333 A_{280}/M-cm.

3.2.5. LeuRS Kinetic Assays

The LeuRS substrate kinetic assay was performed according to literature procedures (36). The assay buffer contained 50 mM HEPES, pH 7.6, 20 mM $MgCl_2$, 1mM DTT, 2 mM ATP and 2 mM [^{32}P]-PP_i (0.5 TBq/mol). The concentration of enzyme was 75 nM. The amino acid concentration ranges varied depending on the activities of the substrates (**1**: 10-500 μM; **2**: 50-5000 μM; **3**: 200-8000 μM). Fifteen μl aliquots were quenched in 500 μL buffer consisted of 200 mM PP_i, 7% w/v $HClO_4$ and 3% w/v activated charcoal. The charcoal was washed twice with 10 mM PP_i, 0.5% $HClO_4$ and counted. The reported data were averages of triplicate experiments.

3.2.6. Expression Plasmid Construction

E. coli genomic DNA was prepared from an overnight culture using DNEasy kit from Qiagen (Chatsworth, CA). The following primers were used to clone the *E. coli* leucyl-tRNA synthetase under the control of its endogenous promoter: LeuS53: ACGTTCTTTGTTGCTAGCTTTGCTAATACG and LeuS35: TATCACGCAGATGCTAGCCACACCCGGCC. The introduced flanking *NheI* restriction sites are underlined. PCR conditions were 50 ng/100 μl of template DNA, 250 ng/100 μl of each primer; 55°C annealing temperature (2 minutes) and 72°C extension temperature (4 minutes). The resulting 2850 based pair product was purified and digested overnight with *NheI*. The target plasmid pQEA1 was also digested with *NheI* and ligated with the digested PCR product. The orientation of the insertion was checked with *SalI* digestion. The preferred insertion orientation of the PCR product places the ORFs of A1 and *leuS* on the same coding strand. This orientation allows the transcription terminator sequence contained on pQE vectors to be immediately downstream of *leuS* and prevents unnecessary read-through during transcription. The ligated plasmid was renamed to pA1EL. LAM1000 was transformed with both pA1EL and pREP4 to yield the LeuRS overexpression strain LAM1000/pA1EL/pREP4. Overexpression of LeuRS was checked

by SDS-PAGE of overnight cultures of strains bearing pA1EL. The accumulation of LeuRS (86kDa) is visible only in strains bearing the overexpression vector.

3.2.7. Large Scale Protein Biosynthesis

M9AA medium (1 L) supplemented with 1 mM MgSO4, 1 mM $CaCl_2$, 0.4 wt% glucose, 1 mg/L thiamine and the antibiotics ampicillin (200 mg/L) and kanamycin (25 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, the medium shift procedure was applied. The cell pellet was then resuspended in 1 L M9 medium supplemented with **2** or **3** and **1** in appropriate concentrations. IPTG (1 mM) was added after either 10 min (for **2**) or 2 hr (for **3**) to induce protein expression. Cells were collected after 3 hr by centrifugation (5000 g, 15 min, 4°C). The pellets were resuspended in 20 ml of Buffer A and stored at -80°C overnight. The cells were thawed rapidly at 37°C, cell debris was sedimented (22,500 g, 50 min, 4°C), and the supernatant was applied to a Ni-NTA column (1 cm x 5 cm). The column was washed with 25 ml portions of Buffer A at pH 8.0, 6.5 and 5.9, sequentially. The target protein was eluted at pH 4.5, dialyzed against water extensively and lyophilized to a fluffy powder.

3.2.8. Protein Characterization

Ultracentrifugation and circular dichroism analyses were performed as described previously in Chapter 2. Analysis of CD thermal melting data was performed according to a previously described procedure using a two-state model (37). For A1-WT, the concentrations used for curve fitting were 10 and 100 μM, while concentrations of 2 and 100 μM were used to fit data for FA1-92 and HA1. The thermodynamic quantities T_m, ΔH_m, ΔCp and K_d were parameters of the fitting procedure and are reported in the 1 M standard state.

3.3 Results and Discussion

3.3.1 Incorporation Studies with a Conventional Host

The synthesis of **2** was shown in the previous chapter. Hexafluoroleucine is synthesized in eight steps starting from hexafluoroacetone, according to a modified procedure described by Zhang et al (32). The scheme is shown in figure 3.3. **3** is fed as the hydrochloride salt directly. Both **2** and **3** are used as mixtures of D and L isomers. We expect only the L-isomer to be utilized by bacteria.

The *E. coli* strain used in the study of fluorinated amino acids is LAM1000, which is a leucine auxotroph. The strain is transformed with pQEA1, which contains the A1 gene, and with pREP4, which contains the repressor gene *lacI*. The capacity of either fluorinated amino acid to support protein synthesis was determined by induction of gene expression in leucine-free culture media supplemented either with **2** or with **3**. The concentration of **2** supplemented is 220 μM and that for **3** is 170 μM (the mass concentration for both amino acids are the same at 40 mg/L of the L form, which is considered to be 1X). Accumulation of the target protein was visualized by western blotting as shown in Figure 3.4. As expected from previous results (18, 19), **2** (lane 4) clearly supports protein synthesis, although the yield of protein is reduced in comparison to that obtained from leucine-supplemented medium (lane 3). In contrast, **3** (lane 5) does not support measurable protein synthesis under the conditions examined in this experiment as shown by comparison with negative controls (lanes 1 and 2). The increased electrophoretic mobility of the fluorinated protein (cf. lanes 3 and 4) has been observed previously for other fluorinated proteins produced by similar methods (19).

To control the level of substitution of **1** by **2**, we varied the concentration of **1** in the medium while keeping the concentration of **2** (as the DL mixture) at 550 μM. The concentration of **1** ranged from 0 to 330 μM (as the L-amino acid). The extent of replacement was determined via amino acid analysis of the purified proteins. We also performed MALDI mass spectrometry on the protein products; in each case, the most intense signal in the mass spectrum corresponds closely to the substitution level calculated from amino acid analysis.

The results are shown in Table 3.1. The extent of **1** replacement decreased from 92% in medium supplemented only with **2**, to levels below the limit of detection in medium supplemented with 330 μM **1**.

We were not able to reach 100% incorporation of **2**, even in the absence of added **1**, possibly due to trace amounts of **1** liberated via cellular protein degradation. Protein yield decreased as the concentration of **1** in the culture medium was reduced. The yield of A1 obtained from leucine-enriched medium was 40 mg/L, twice that from medium supplemented only with 550 μM of **2**.

3.3.2. Characterization of Tfl-Containing A1

The secondary structures of the wild-type protein (A1-WT) and the variant containing 92% **2** (FA1-92) were examined by circular dichroism (CD) spectroscopy (Figure 3.5). Both proteins are more than 90% helical as determined from the molar ellipticity at 222 nm, and the extent of overlap of the spectra suggests essentially identical secondary structures. Equilibrium sedimentation indicated that both peptides formed stable dimers with dissociation constants ≤ 10 μM. At high μM concentrations the peptides further associated to form higher-order species. The data were well described by a dimer-to-tetramer equilibrium, yielding dimer-to-tetramer equilibrium constants of 80.5 μM and 380.2 μM for A1-WT and FA1-92, respectively.

Previous examples of replacement of ***d***-position leucine residues in leucine-zipper peptides by other natural amino acids have all resulted in reduction of coiled-coil stability (38-40). The reduction in stability observed in such experiments has been attributed to the fact that these substitutions are usually of the "large" to "small" type and cause losses in hydrophobic packing efficiency (41-43). In marked contrast, thermal and chemical unfolding studies show that fluorinated variants of A1 are highly resistant to denaturation. The "melting temperature" (T_m) of FA1-92 is elevated by 13°C compared to A1-WT (Figure 3.6), while the midpoint concentration (C_m) for urea denaturation for FA1-92 is increased to 7 M (vs. 2.8 M for A1-WT) (Figure 3.7). Surprisingly, proteins with low levels of fluorination (FA1-17 and FA1-29) exhibit pronounced increases in stability. In FA1-17, for example, the single substitution of a side-chain methyl group by a trifluoromethyl group results in an increase in T_m by 6°C and in C_m by ca. 2 M.

Global thermodynamic fitting was used to obtain the thermodynamic quantities associated with the monomer-to-dimer transitions of A1-WT and FA1-92 (Table 3.2) (37). The intermediate samples (FA1-17 and FA1-29) were not analyzed in this way because the compositional heterogeneity of these samples precludes application of a two-state transition model. It must be emphasized, however, that FA1-

92 is also heterogeneous, since **2** is incorporated into bacterial proteins as an equimolar mixture of the (2S, 4S) and (2S, 4R) diastereomers. The free energy of folding for FA1-92 is 2.4 kcal/mol (of monomer) more negative than A1-WT at 37°C, which corresponds to 0.4 kcal/mol of stabilization for each **2** site localized at the dimer interface. Consideration of the large numbers of leucine residues packed in the hydrophobic cores of many proteins leads to the conclusion that the additive stabilizing effects of **2** might in some cases be quite substantial.

The results reported so far show that one can incorporate **2** efficiently into proteins produced *in vivo*, control the level of incorporation, maintain secondary and higher-order protein structure, and elevate protein stability with respect to thermal and chemical denaturation. These conclusions led to the speculation whether introducing the nearly perfluorinated **3** might further enhance the stabilities of the leucine-zipper protein.

3.3.3 Incorporation of Hexafluoroleucine

Using a conventional bacterial host discussed above, the more hydrophobic amino acid **3** is completely inactive during protein translation. The failure of **3** to support A1 synthesis is surprising, especially considering that both (2S, 4S) and (2S, 4R) diastereomers of **2** are incorporated into proteins with equal efficiencies (19). We did not anticipate editing by the synthetase because the near identity of the geometries of **1** and **3** should preclude entry of the latter into the editing site of the enzyme (44, 45). We attributed the difference in translation activities between **2** and **3** to three possible reasons. First, the increased hydrophobicity of **3** results in poor transport of the amino acid through cell membrane; second, Hfl-tRNA^{Leu} cannot be processed correctly by the ribosomal machinery and may thus be rejected by accessory proteins such as EF-Tu; third and most likely, **3** is recognized poorly by LeuRS, and the catalytic velocity of Hfl-tRNA^{Leu} synthesis is too slow to support protein translation.

The recognition of LeuRS towards **3** is reflected in the rate of the first step of tRNA aminoacylation, the adenylation step (see Chapter 1). The second step, in which the bound amino acid is transferred to the cognate tRNA, is unlikely to be affected by the side chain functionality. To quantitatively understand the rate at which LeuRS activates its substrates during aminoacylation, we measured the rates of the adenylation reactions with purified LeuRS in the presence of **1**, **2** or **3** using *in*

vitro ATP-PP_i exchange assay. The synthetase was purified in two chromatographic steps from an overexpressing strain to > 95% homogeneity. The assay employs radiolabeled ^{32}PPi and measures the rate of accumulation of α-^{32}P-ATP. The results are shown in Table 3.3. Compared to **1**, **2** is activated ~240 times slower, but is still an efficient substrate during protein biosynthesis under normal conditions. In contrast, the rate of **3** activation is 4100 times worse compared to the cognate substrate **1**. The additional twenty-fold attenuation in the activation rate of **3** upon fluorinating both terminal methyl groups renders the amino acid inactive during translation.

The overall rate of activation can be described by the following equation:

$$v_{activation} = \frac{k_{cat}}{K_M}[HFL][LeuRS]$$

To compensate for the decrease in catalytic efficiency and accelerate the velocity of **3** activation, one can increase the concentrations of both the amino acid and the enzyme. Elevating the cellular concentration of the amino acid can be accomplished by increasing the feed concentration of **3** in the medium, assuming it can be efficiently transported across cell membranes. The concentration of **3** in the medium is limited by its solubility.

Increasing the concentration of LeuRS requires the cellular expression levels of the enzyme to be elevated beyond endogenous conditions. Previous work with methionine analogues has shown that elevation of the MetRS activity of the host facilitates incorporation of analogues that are poor substrates for the synthetase (46). We cloned a copy of the LeuRS gene (47) on the plasmid pQEA1. The plasmid is present in cells in multiple copies (copy number ~ 30), and can thus increase the number of intracellular LeuRS transcripts. The gene of LeuRS, along with its endogenous promoter, was amplified from *E. coli* genomic DNA and ligated into pQEA1 to yield pA1EL (Figure 3.8). The promoter is constitutive, which allows the continuous expression and accumulation of LeuRS during cell growth. We attempted to clone the LeuRS gene under the control of a stronger promoter, such as the constitutive *tac* promoter. When transformed into the host strain, the strong promoter resulted in very slow cell growth and rapid cell death after A1 induction. The cloned *leuS* gene is inserted between the A1 gene and a transcription terminator at a unique *NheI* restriction site. Overexpression of LeuRS by LAM1000 strain transformed with pA1EL can

be visualized by analyzing the whole cell lysate of overnight culture with SDS-PAGE (figure 3.8). Accumulation of LeuRS (86kDa) is visible only in strains bearing the overexpression vector.

The intracellular LeuRS activity of LAM1000 carrying pA1EL is compared to that carrying pQEA1 using the ATP-PP_i exchange assay. Cells from each strain were grown in minimal medium culture to OD_{600} of 1.0 and were collected, washed and lysed in reaction buffer. The whole cell lysate is used directly in the activation assay in place of purified enzymes. LAM1000/pA1EL exhibited an eight-fold increase in LeuRS activity compared to LAM1000/pQEA1.

Hexafluoroleucine is assayed for translation activity using the expression strain LAM1000/pA1EL/pREP4. The results are shown in figure 3.9. Lanes 1-5 correspond to the expression profile of the overexpression strain, while lanes 6-9 are that for the conventional strain. As observed before, no A1 expression is evident with the conventional strain, even when **3** is supplied at 2.7 mM (640 mg/L). However, when the concentrations of both the amino acid and the synthetase are elevated, accumulation of **3**-containing A1 becomes clearly evident from SDS-PAGE. The increases in electrophoretic mobility observed in lanes 4 and 5 are consistent with a similar behavior observed with FA1 (48), confirming the presence of fluorinated amino acids in the target protein.

3.3.4 Composition Analysis of Hfl-containing A1

During the media shift procedure, a ten minute incubation period is included between the resuspension of washed cell pellets and IPTG addition. This period is included as a means for cells to utilize residual intracellular leucine before the synthesis of target protein. Figure 3.10 shows the MALDI spectrum of A1 produced under these conditions in the presence of 1.4 mM of **3**. The mass of the wild-type A1 is 8307 and each substitution of **1** with **3** leads to a mass increase of 108 units. The multiple peaks in the spectrum suggest that A1 species with multiple levels of **3** substitutions are present. Replacement of all eight sites in the protein is observed in the spectrum (~ 9171), with the highest peak corresponding to five replacements (~ 8847). Unexpectedly, in addition to the peaks separated by 108 mass units, intermediate peaks separated by 54 mass units are also present. These less intense peaks indicate the presence of **2** in protein in addition to **3**. This result is especially surprising considering that **2** is not added

to the expression medium, and no cellular pathways of **2** biosynthesis is available. Amino acid analysis confirmed that the ~ 15% of leucine positions in the protein is replaced with **2**.

A more detailed analysis of the composition of **3** used in protein expression reveals that **2** was present as a trace contaminant. Figure 3.11 shows the ^{19}F NMR spectrum of **3**. The chemical shifts are referenced with respect to trifluoroacetic acid. The signals at 7.8 ppm correspond to the fluorines of **3**. Expansion of the region where fluorines of **2** are expected (~2.2 ppm) shows the characteristic doublet of doublets associated with **2**. The contamination is likely to be present throughout the synthesis (We believed it originated from the starting material hexafluoroacetone, which is of 97% purity, contaminants unknown). Even though **2** is present at less than 0.2% in the sample of **3** used to supplement the medium, its twenty-fold faster rate of activation by LeuRS allows it to be incorporated into proteins much more readily. The result is enrichment of **2** in the target protein. We postulated that extending the preinduction incubation time could allow the limited amount of **2** to be utilized in the synthesis of other cellular proteins. Subsequently at the time of target protein expression, the amount of **2** is diminished. We performed experiments with incubation times extended to 30, 60, 90, and 120 minutes. We did not extend the time beyond two hours because cell death was observed (shown by a decrease of OD_{600}). The proteins produced under these conditions were analyzed by MALDI-MS. The results are shown in Figures 3.12-14. As we anticipated, the amount of **2** present significantly decreased as we increased the incubation time. When the cells were incubated in M9 supplemented with **3** for two hours prior to induction, the amount of **2** in the protein product was reduced to below 3% as compared to 15% for 10 minute incubation, and the amount of **2** in A1 was elevated to 80% as compared to 50%. The increase in the levels of **3** could also be due to the depletion of residual **1** as well. Figure 3.15 is the MALDI spectrum of A1 obtained in a different experiment in which the preinduction incubation period was maintained at two hours and the level of **3** supplementation was increased to 3 mM. The highest peak corresponds to seven out of eight substitutions. The intermediate peaks are still visible and their intensities are enhanced by combinatorial effects associated with the occupancy of the leucine sites by three different amino acid variants. Table 3.4 shows the amino acid analysis results of the sample (HA1). The ratio of **2** to **3** in the protein was determined to be approximately 0.04.

Figure 3.16 shows the MALDI spectrum of one fragment of HA1 released by trypsin digestion. Purified protein was treated with trypsin, which cleaves the amide bond after lysine and arginine residues. The mass of the wild type fragment shown in Figure 3.16 is 2442 and it has the sequence LKNEIEDLKAEIGDLNNTSGIR. Three leucine residues are present in this peptide. Fragments that correspond to 0, 1, 2, and 3 sites of replacement are observed at masses of 2442.26, 2550.20, 2658.11 and 2766.05, respectively. A different set of tryptic fragments containing two sites of substitution is apparent at slightly higher masses (0: 2772.18; 1: 2880.17; 2: 2988.12)

3.3.5 Characterization of HA1

To characterize the structural features of A1 substituted with **3**, HA1 was expressed on a larger scale (1 L). HA1 was purified via nickel affinity column under denaturation conditions (figure 3.17). Purified proteins were pooled, dialyzed, and lyophilized to a powder. The yield of HA1 was 8 mg/L, a 5-fold decrease compared to that of wild-type A1. The composition of HA1 obtained via large scale expression was verified to be the same as that obtained from small scale expression (figure 15 and 16).

The secondary structure of HA1 is determined by circular dichroism spectroscopy to be >90% α-helical. Ultracentrifugation shows the protein to be predominantly dimeric at concentrations of 10 ~ 100 μM, with a dissociation constant for the dimer-to-tetramer equilibrium of 188 ± 4 μM (as compared to a value of 80.5 μM previously reported for A1). These results suggest minimal disruption of the folding and assembly of the coiled-coil architecture of A1 upon near-complete fluorination of the leucine sites.

The stability of HA1 with respect to thermal and chemical denaturation is significantly improved compared to the wild-type protein and FA1-92 (Figure 3.18, 3.19). The melting temperature of HA1 is elevated from 54°C (for A1) to 76°C; for comparison, the same protein outfitted with **2** melts at 67°C. The free energy of unfolding at 25°C was determined by thermodynamic fitting of the melting curve to be 14.4 kcal/mol, an elevation of 3.7 kcal/mol compared to A1. HA1 also exhibits remarkable resistance to denaturation by urea; the protein is nearly fully folded (as reported by circular dichroism spectroscopy) at 4 M urea, and unfolding remains far from complete even in 8 M solutions of the denaturant, a significant improvement over that of FA1-92. In contrast to the prediction of molecular dynamics simulation discussed in Chapter 2, fluorinating both methyl groups of leucine further enhanced the stability of the

leucine zipper proteins. Juxtaposition of the aliphatic side chains of the leucine residues contributes significantly to the stability of coiled-coil proteins. Replacement of these residues with the highly hydrophobic, fluorinated analog **3** increases the driving force for protein folding and assembly. We are currently investigating the effects of fluorinating more complex proteins, including four-helix bundle cytokines such as interleukin-2 and granulocyte colony stimulating factor (G-CSF).

3.4 Conclusion

The studies presented in this chapter show that fluorination of protein cores can be performed through biosynthetic methods. Replacement of leucine residues at the dimer interface of a leucine zipper protein with either trifluoroleucine or hexafluoroleucine significantly improves the thermal and chemical stabilities of the target protein. The secondary structures and self-assembly properties of the fluorinated proteins are not effected by fluorination, suggesting fluorine atoms induce minimal packing perturbation during protein folding. Trifluoroleucine is recognized by leucyl-tRNA synthetase efficiently and can quantitatively replace leucine residues in recombinant proteins with conventional *E. coli* hosts. The percentage of substitution can be tuned by varying the concentrations of competing leucine in the expression media. The gain in protein stability increases as the substitution level increases. Hexafluoroleucine is a more sluggish analog of LeuRS and can be incorporated *in vivo* upon elevation of the host LeuRS activity. Introduction of hexafluoroleucine into the coiled-coil protein A1 is accompanied by significant increases in the thermal and chemical stability of the folded form of the protein. We believe that these results can be extended to the design of highly stable proteins and to the engineering of protein-protein interactions. The results described here also illustrate the synthetic potential of microbial hosts equipped with elevated aaRS activities. Non-natural amino acids that do not support protein synthesis in conventional hosts can be incorporated efficiently into recombinant proteins in such engineered hosts.

Table 3.1 Wild type and fluorinated proteins produced *in vivo*.

Protein	[Leu] in Culture (μM)[a]	Yield (mg/L)	%Tfl by AAA[b]	MALDI Mass[c] (Tfl sites)
A1-WT	330	40	0	8307.9 (0)
FA1-17	114	27	17	8361.2 (1)
FA1-29	83	25	29	ND[d]
FA1-79	23	21	79	8631.9 (6)
FA1-92	0	20	92	8739.8 (8)

[a] Concentration of L-Leucine. The concentration of **2** in culture was 550 μM (DL mixture). [b] % substitution determined from the diminution of the **1** mole fraction by amino acid analysis. [c] Mass of the most intense signal in the MALDI mass spectrum. Numbers in parentheses indicate corresponding numbers of sites occupied by **2**. [d] ND: not determined.

Table 3.2. Thermodynamic data for the folding of A1-WT and FA1-92.

Protein	T_m[a] (10 μM)	T_m[b] (1 M)	ΔG° (37°C)[c]	ΔΔG°	ΔH_m[d] (1 M)	ΔC_p[e] (1 M)	C_m[f] (0°C)
A1-WT	54	103	-10.7		-70.9	-252	2.8
FA1-92	67	116	-13.1	-2.4	-77.0	-272	7

[a] Midpoint of the thermal denaturation curve at 10 μM protein concentration (PBS, pH 7.4). Units for T_m are °C. [b] Midpoint of the thermal denaturation curve extrapolated to 1 M standard state using non-linear least squares fit. [c] The free energy of folding at 37°C at 1 M standard state. Units for ΔG° and ΔΔG° are kcal/mol of monomers. [d] The enthalpy of folding at the midpoint temperature extrapolated to 1 M standard state; units are in kcal/mol of monomers. [e] Heat capacity change upon folding at 1 M standard state; units are in cal/mol·K. Uncertainties in T_m (1 M), ΔG°, ΔH_m and ΔC_p are ±1.5°C, 1.2, 4.8 kcal/mol, and 120 cal/mol-K, respectively. [f] Midpoint urea denaturation concentration at 0°C in M of urea.

Table 3.3. Kinetic parameters for activation of **1**, **2**, and **3** by *E. coli* LeuRS. [a]

Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (*rel*)
1	3.8±0.3	17.8±5.1	1
2	0.49±0.05	561±121	1/242
3	0.11±0.01	1979±804	1/4100

[a] Leucine was used as the L-isomer; **2** as a mixture of the (*2S,4S*), (*2S,4R*), (*2R,4S*) and (*2R,4R*) forms, and **3** as the D,L-mixture. K_m values are reported in terms of the concentrations of the L-isomers. We have shown previously by ^{19}F NMR spectroscopy that both of the *2S*-isomers of **1** are incorporated *in vivo*; however, we have not evaluated the kinetic parameters for each isomer separately.

Table 3.4 Amino acid analysis of HA1[a], [b].

	Expected		Observed		
Amino Acid	# of Res.	Mole Fractions	pMoles	Mole Fractions	% Err
ASP	9	12.2	4828	11.53	5.22
GLU	13	17.6	7599	18.15	3.29
SER	6	8.1	3492	8.34	2.84
GLY	5	6.8	3147	7.51	11.21
HIS	6	8.1	3352	8.00	1.28
ARG	6	8.1	3360	8.02	1.05
ALA	7	9.5	3743	8.94	5.52
VAL	3	4.1	1585	3.78	6.65
MET	2	2.7	1218	2.91	7.60
ILE	3	4.1	1535	3.67	9.59
LEU	**8**	**10.8**	**1036**	**2.47**	**77.12**
LYS	4	5.4	2049	4.89	9.49
HFL			**3354**	**8.01**	
Total	74	100	41747	100	

[a] Expression conditions: 3 mM of **3** supplied to expression medium; The preinduction incubation period was extended to 120 minutes. [b] Calculation of the fractions of **1**, **2** and **3** in the protein was performed as following:

The combined pmoles of **1** and **3** total 4390. the combined mole fractions of **1**+**3** amount to 10.48% compared to the expected value of 10.8%. We attribute the difference of 0.32% to presence of **2**, which can not be quantified using amino acid analysis (**2** co-eludes with lysine). Calculations based on mole fraction of **1**, **2** and **3** lead to the following distribution of amino acids at the **1** positions:

1: 2.47 / 10.8 = 22.9 %

3: 8.01 / 10.8 = 74.2 %

2: 0.32 / 10.8 = 2.9%

2 / **3**: 0.04 (Figure reported in main text)

A.

MRGSHHHHHHGSMA SGDLENE VAQLERE
VRSLEDE AAELEQK VSRLKNE IEDLKAE IGDLNNT
SGIRRPAAKLN

B.

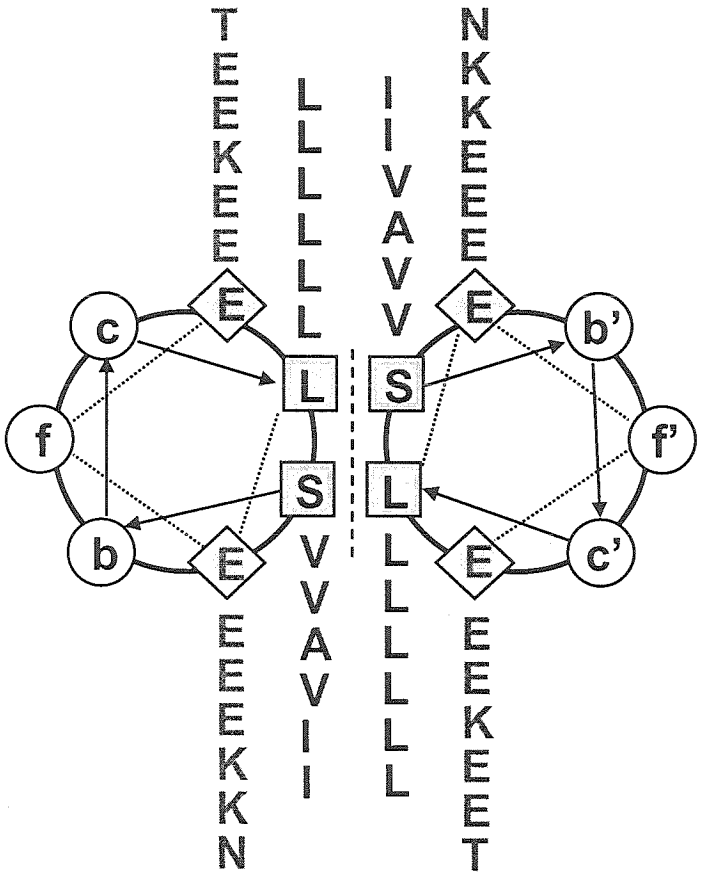


Figure 3.1. **A.** Sequence of target leucine zipper protein A1. The protein contains 74 residues and eight leucines. Seven of the eight leucines occupy the ***d*** position of the heptad repeats. The heptad repeats are underlined. A N-terminus 6xHis tag is installed to facilitate metal affinity chromatography. **B.** The helical wheel representation of the heptad repeats. Only the residues at the ***a***, ***d***, ***e***, and ***g*** positions are shown. Note that ***d*** positions are entirely made up with leucines.

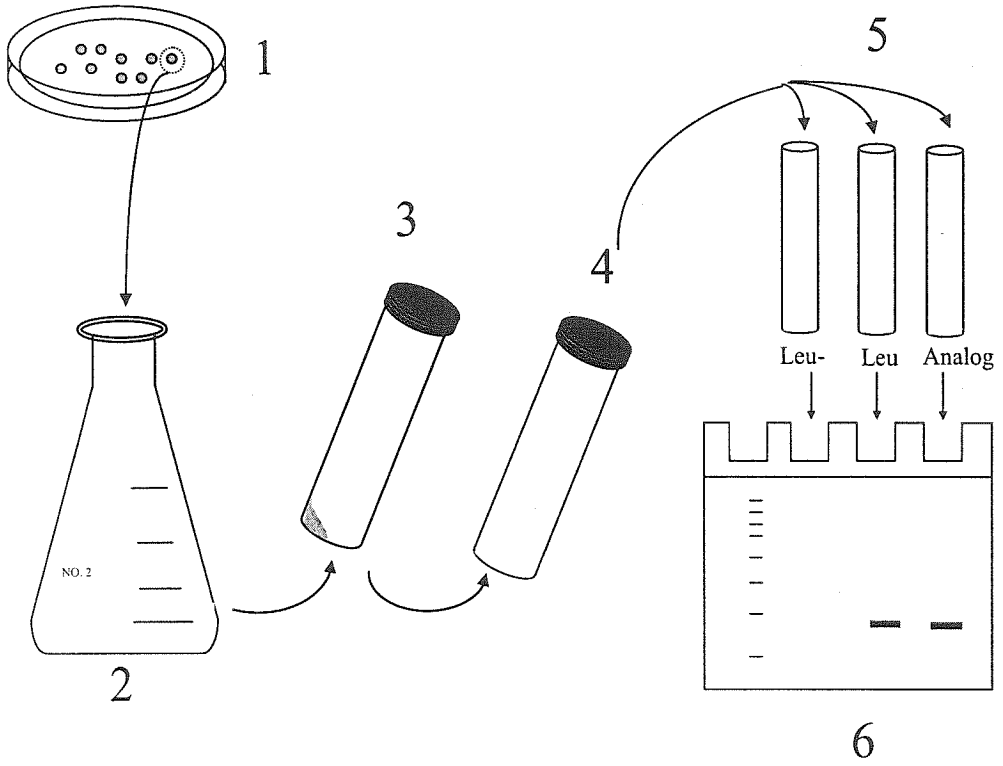
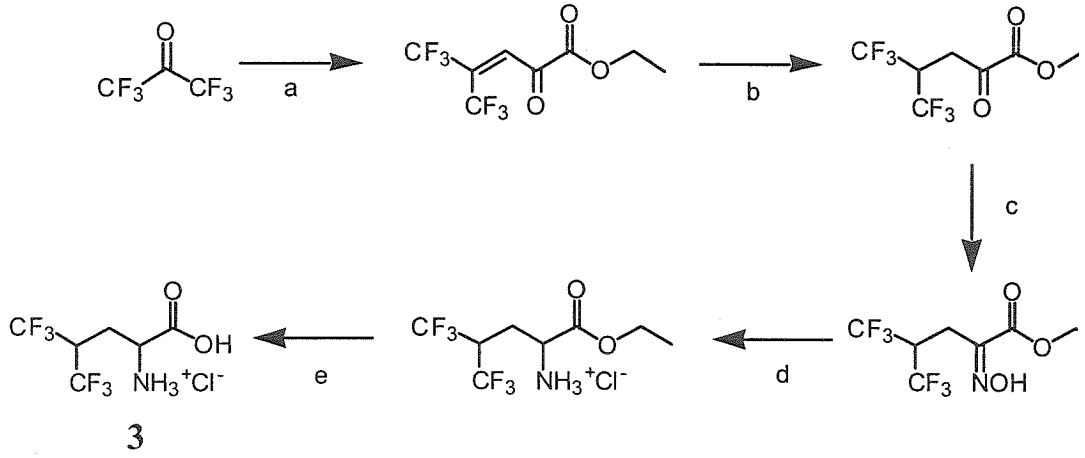


Figure 3.2. The media shift procedure used in this work for assaying leucine analog translational activities. **1.** A single *E. coli* auxotrophic colony is selected. The strain is LAM1000 and a description of its phenotype is available in Materials and Methods. This strain is transformed with pREP4, a repressor plasmid; and with either pQEA1, the regular expression strain, or pA1EL, the LeuRS overexpression strain. **2.** The single colony is grown in M9 medium supplemented with all twenty amino acids till OD_{600} reaches 0.9~1.0. **3.** Cells are collected by centrifugation and washed with 0.9% NaCl three times. The washing steps remove residual leucine. **4.** The washed cells are resuspended in fresh M9 medium contain all of the twenty natural amino acids with the exception of leucine. The volume of culture used is the same as the growth culture, maintaining the cell density at 0.9~1.0. **5.** The cells are distributed to different aliquots. One of aliquot contains no leucine and serves as the negative control; one aliquot is supplemented with leucine and serves as the positive control. The rest of the aliquots are enriched with one of the fluorinated analogs. IPTG is added after a certain period of incubation to induce protein expression. **6.** SDS-PAGE is used to visualize the amount of protein expression in each of the aliquots. Protein accumulation suggests that the specific analog is a suitable surrogate of LeuRS during protein translation.

A.



(a) $PPh_3C_5H_6O_3$, $(C_2H_5)_2O$, -78°C. (b) H_2 (30 psi), PtO_2, RT. (c) NH_2OH, $C_2H_3O_2Na$, 50% EtOH, 50°C. (d) $TiCl_3$, L-tartaric acid, $NaBH_4$, pH 7.0, RT. (e) 6 N HCl, 120°C, o/n.

B.

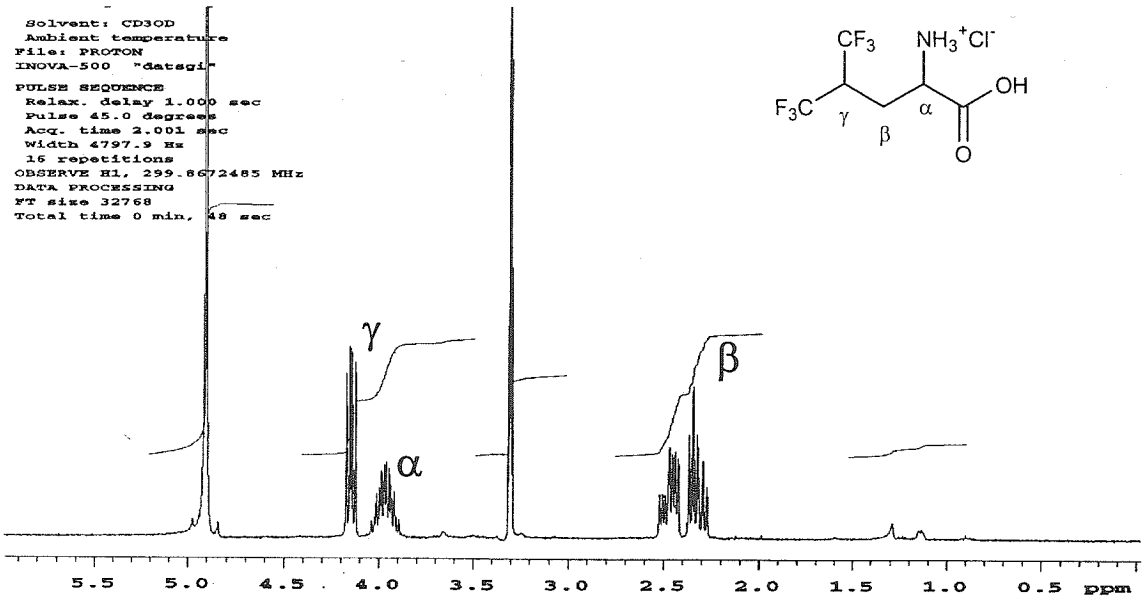


Figure 3.3. A. Synthesis of **3** starting from hexafluoroacetone. The overall yield of the synthesis is 12%. **B.** The ^1H NMR spectrum of hexafluoroleucine hydrochloride.

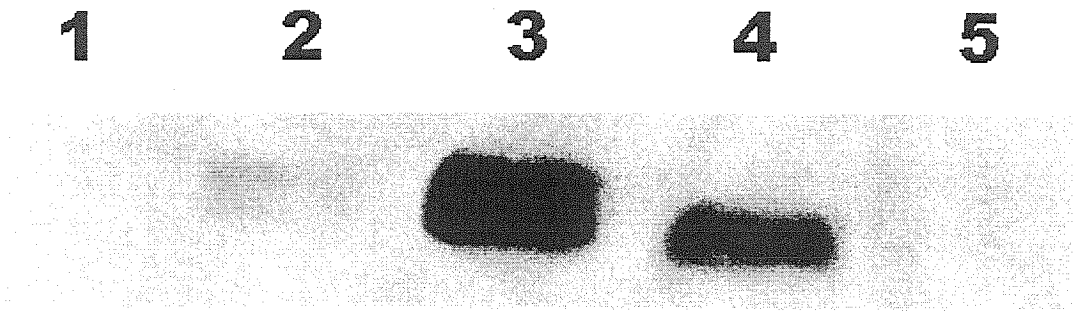


Figure 3.4. Western blot analysis of A1 expression. The expression strain is LAM1000/pQEA1/pREP4. Cultures were supplemented with **1** or with one of the analogs **2** or **3**. Lane 1: uninduced sample; lane 2: induced sample without supplementation; lane 3: induced sample supplemented with **1** (150 μM); lane 4: induced sample supplemented with **2** (220 μM); lane 5: induced sample supplemented with **3** (170 μM). Only **2** supports A1 biosynthesis in the absence of **1**. The more hydrophobic **3** in completely inactive under the assay conditions.

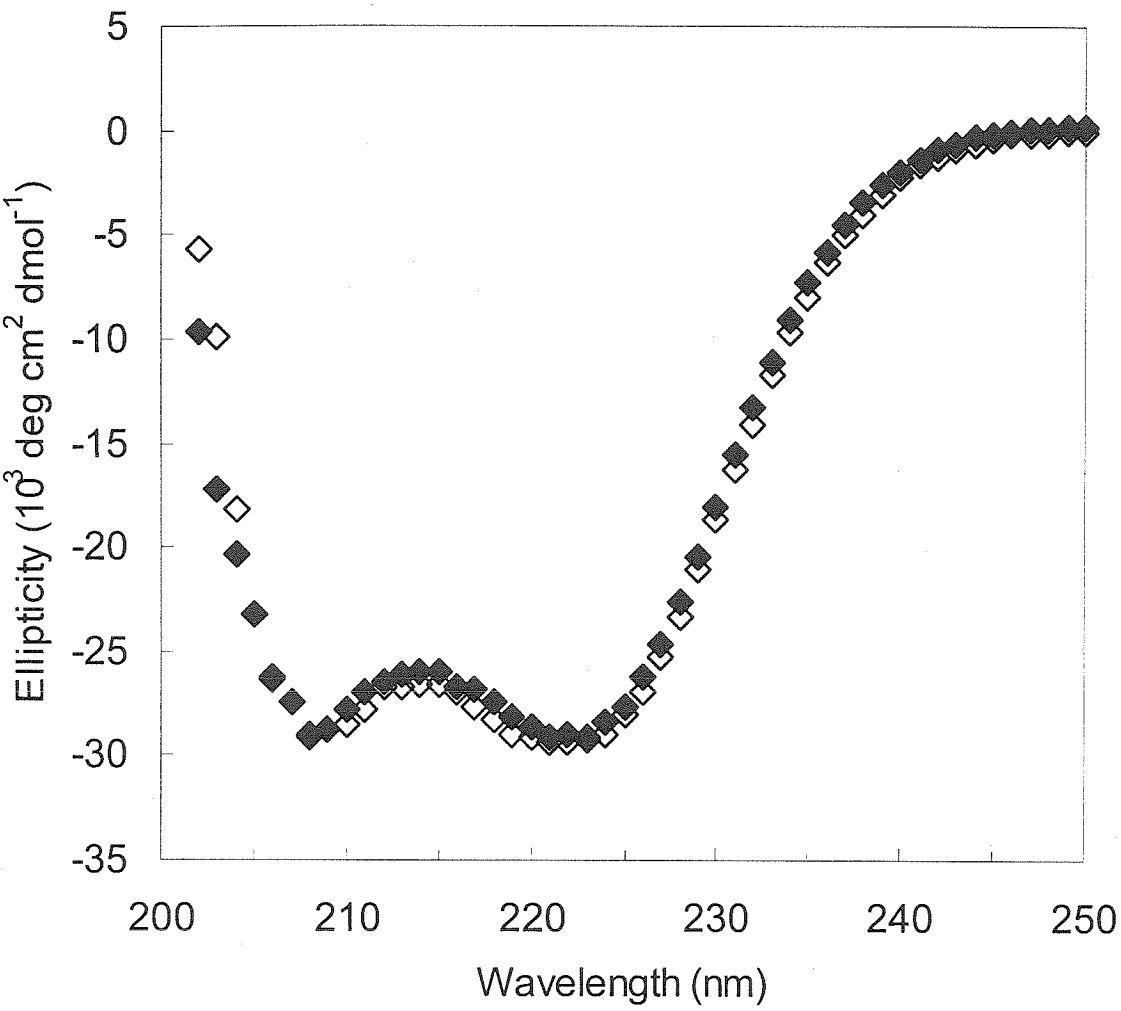


Figure 3.5. CD spectra of A1 (◇) and FA1-92 (◆) at 0°C (10 μM protein concentration, PBS buffer, pH 7.4). The intensities of the minimum at 222 nM suggests > 90% helical structure (Chen, Y. H.; Yang, J. T.; Chau, K. H. *Biochemistry* **1974**, *13*, 3350). The overlapping curves indicate identical secondary structures between the two species.

.

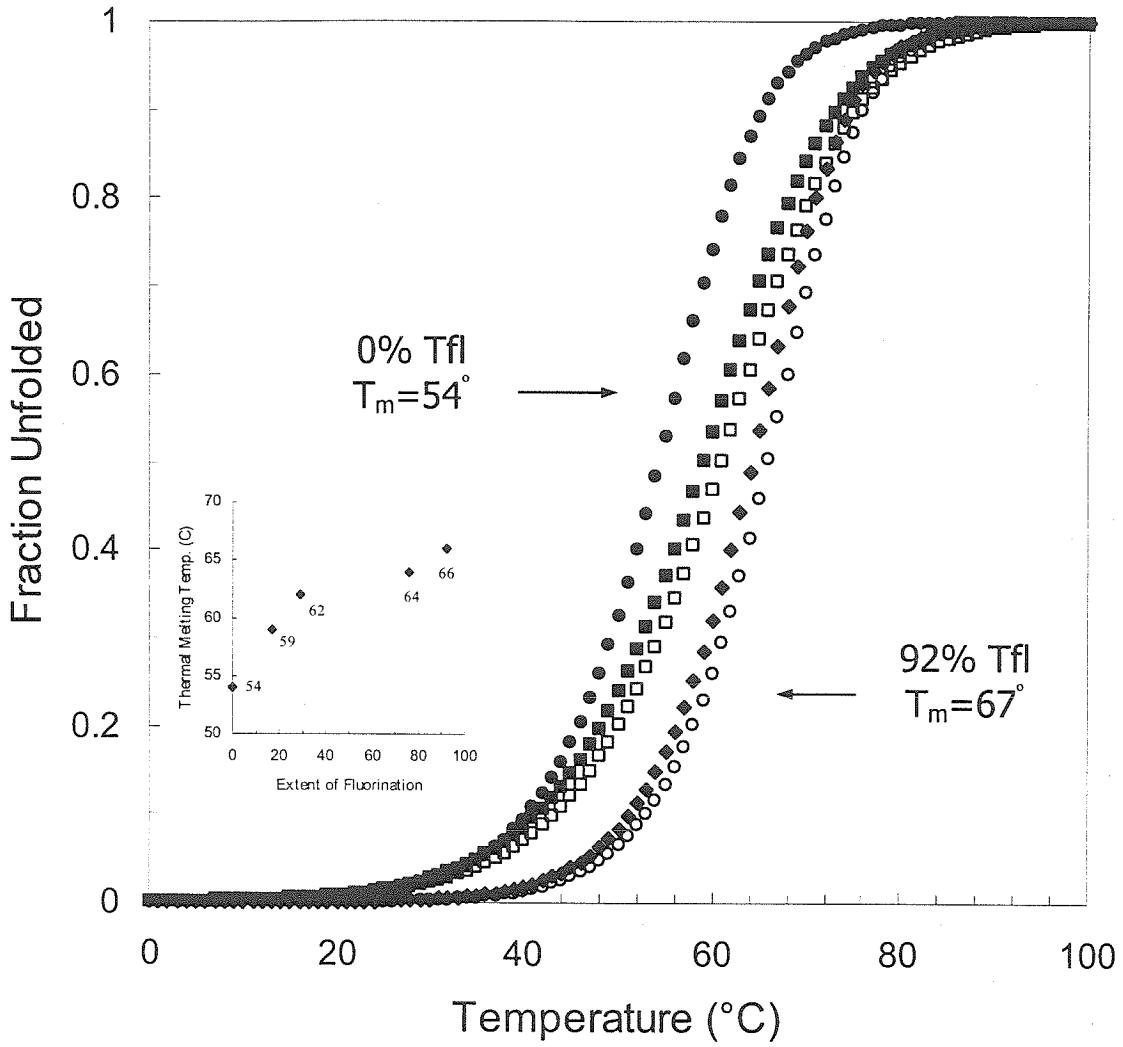


Figure 3.6. Thermal denaturation of partially fluorinated A1 proteins: A1-WT (●), FA1-17 (■), FA1-29 (□), FA1-79 (◆) and FA1-92 (○) (10 μM protein concentration, PBS buffer, pH 7.4). The denaturation of α-helical structure is followed by monitoring the intensity of the 222 nM minima using circular dichroism spectroscopy. The thermal denaturation temperature (T_m) is defined to be the temperature at which 50% of the protein is denatured. The T_m for the curves shown are: A1: 54°C; FA1-17: 59°C; FA1-29: 62°C; FA1-79: 64°C; and FA1-92: 67°C. Inset: T_m as a function of extent of fluorination.

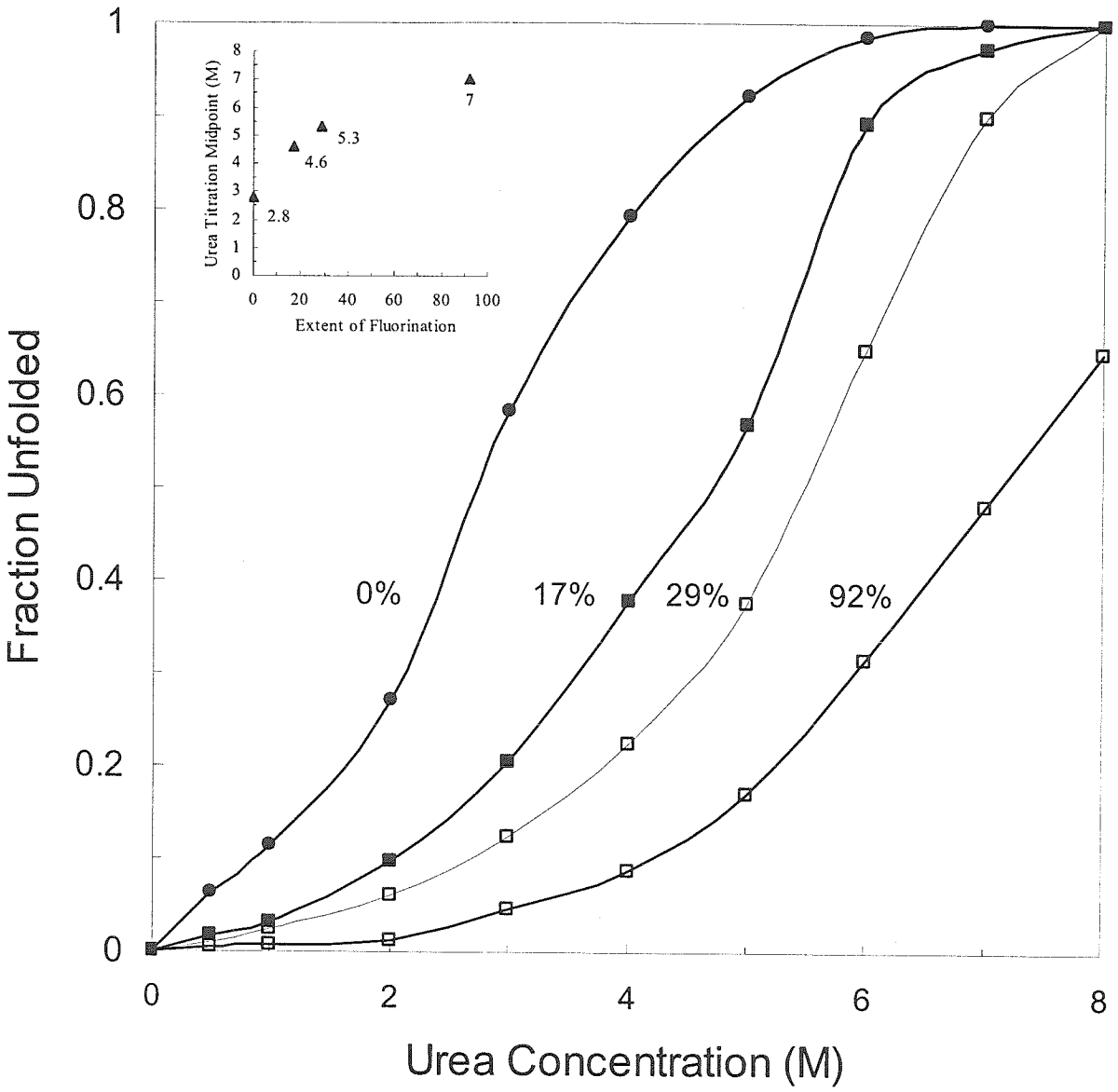
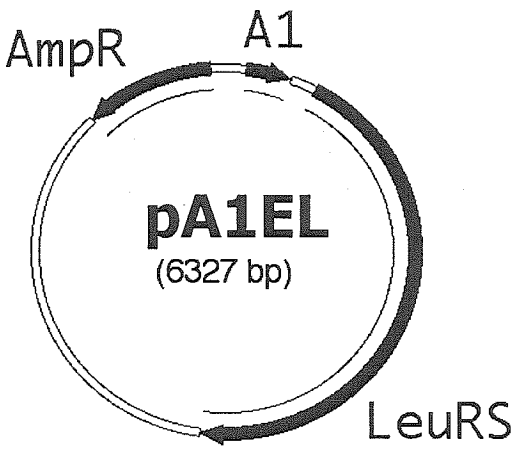


Figure 3.7. Urea titration of WT-A1 (●), FA1-17 (■), FA1-29 (□), and FA1-92 (○) at 0°C. The denaturation of α-helical structure is followed by monitoring the intensity of the 222 nM minima using circular dichroism spectroscopy. The midpoint urea concentration (C_m) is defined to be the concentration of urea at which 50% of the protein is denatured. The C_m for the curves shown are: A1: 2.8 M; FA1-17: 4.6 M; FA1-29: 5.3 M; and FA1-92: 7 M. Inset: C_m as a function of extent of fluorination.

A.



B.

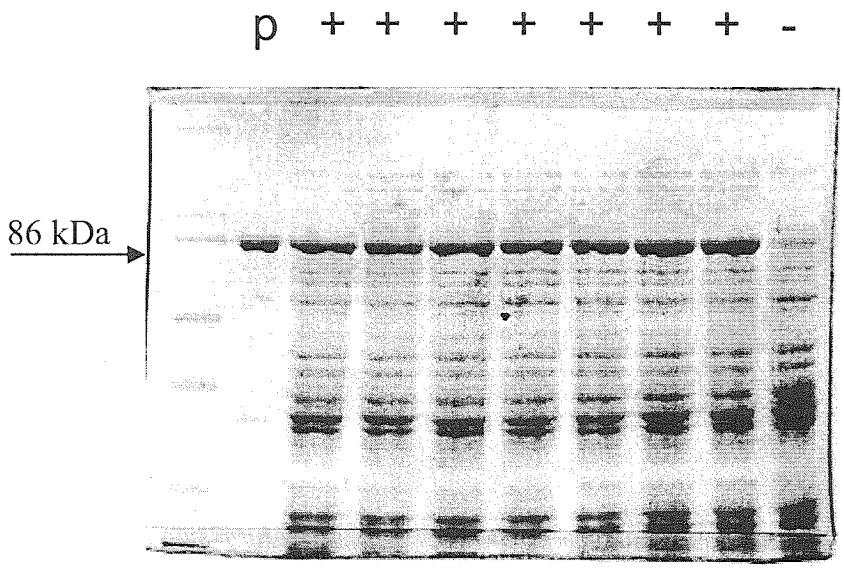


Figure 3.8. A. The map of pA1EL. The *E. coli* leucyl-tRNA synthetase gene is cloned along with its endogenous promoter. The entire cassette (black) was amplified from *E. coli* genomic DNA and inserted into the unique *NheI* site downstream of the A1 gene. The orientation of the insertion was check with *SalI* and the preferred orientation is shown. In this orientation, the LeuRS gene is immediately upstream of the transcription terminator present on pQE vectors. **B.** SDS-PAGE of whole cell lysate of strains carrying the overexpression strain pA1EL (+), carrying the conventional strain pQEA1 (-). Purified LeuRS is shown in lane p. The 86 kDa protein is clearly more abundant in strains bearing pA1EL. The LeuRS activity is increased eightfold in this bacteria host.

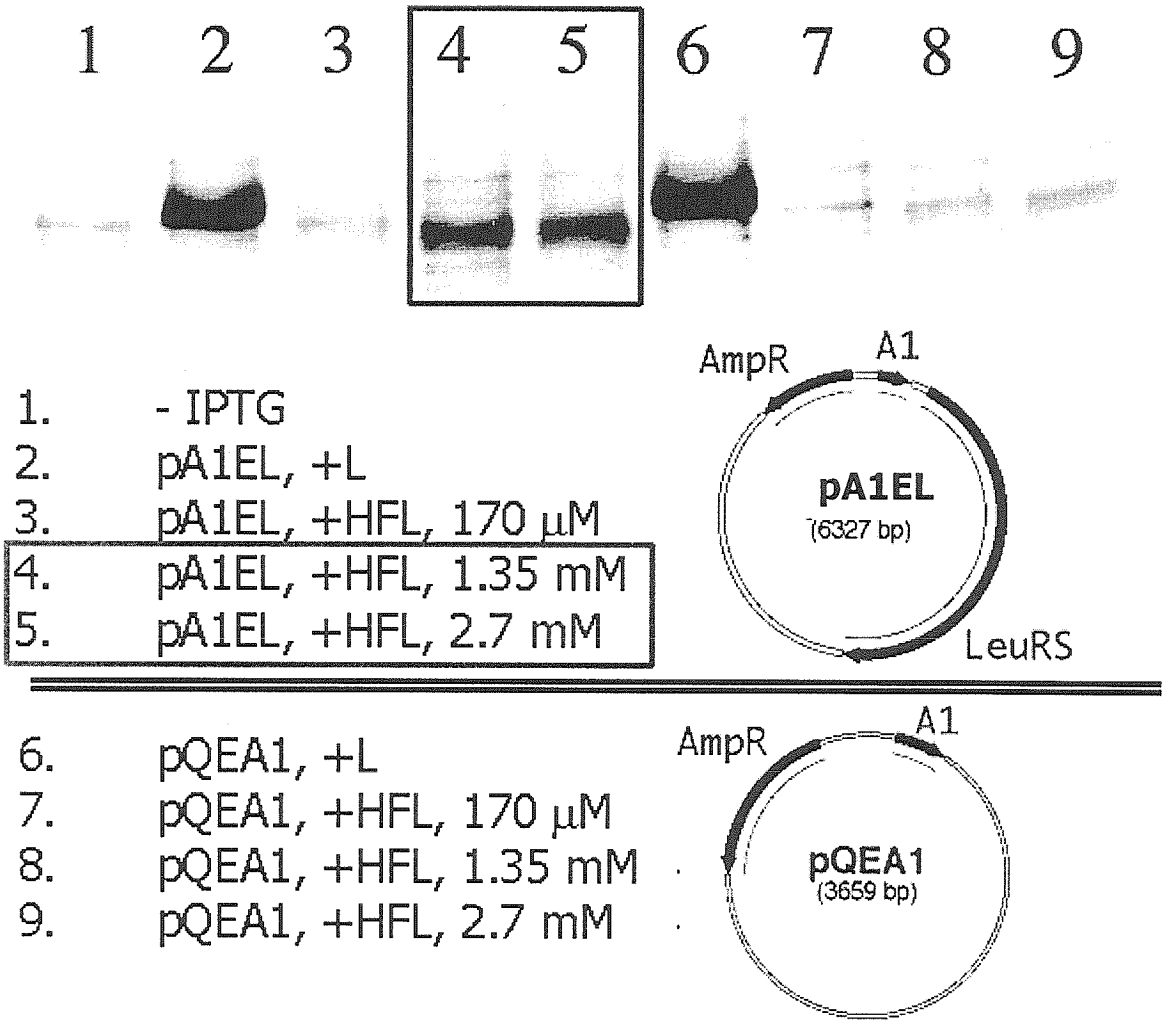


Figure 3.9. Incorporation of hexafluoroleucine into A1 using the LeuRS overexpression strain LAM1000/pA1EL/pREP4. Lanes 1-5: expression results of strains carrying vector pA1EL. 1: No amino acid added; 2: leucine added; 3: **3** at 0.17 mM; 4: **3** at 1.35 mM; 5: **3** at 2.7 mM. Lanes 6-9: expression results of strains carrying conventional vector pQEA1. 6: leucine added; 7: **3** at 0.17 mM; 8: **3** at 1.35 mM; 9: **3** at 2.7 mM. All concentrations are for the L-isomer. The poor analog **3** is incorporated only at elevated LeuRS activity and high amino acid concentrations (lanes 4 and 5).

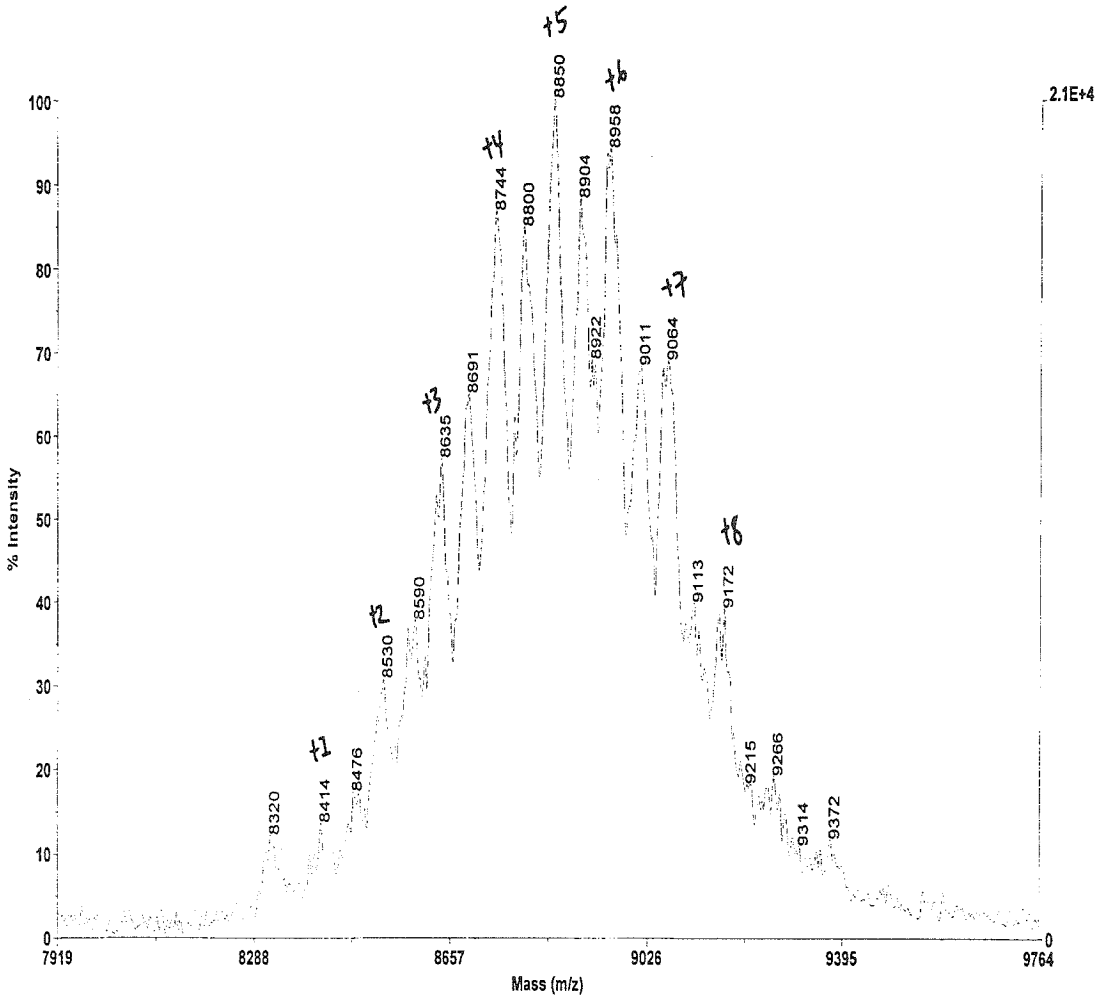


Figure 3.10. Mass spectrum of **3**-containing protein synthesized with 1.35 mM of **3**, and induced after *10* minutes of incubation after media shift. The wild type A1 mass is 8307. Each replacement of **1** with **3** increases the protein mass by 108 mass units. Full substitution of all eight leucines with **3** results in the mass of 9171 (+8). The highest peak observed corresponds to five out of eight substitutions. The surprising observation is that in addition to the expected 108 mass intervals, peaks separated by 54 mass peaks are also observed. The intermediate peaks indicate the mono-trifluoromethyl substituted **2** is also present in addition to the bis-trifluoromethyl substituted **3**. Amino acid analysis of this sample shows the fractions of **1**, **2**, and **3** present at leucine positions in the protein are 35%, 15%, and 50%, respectively.

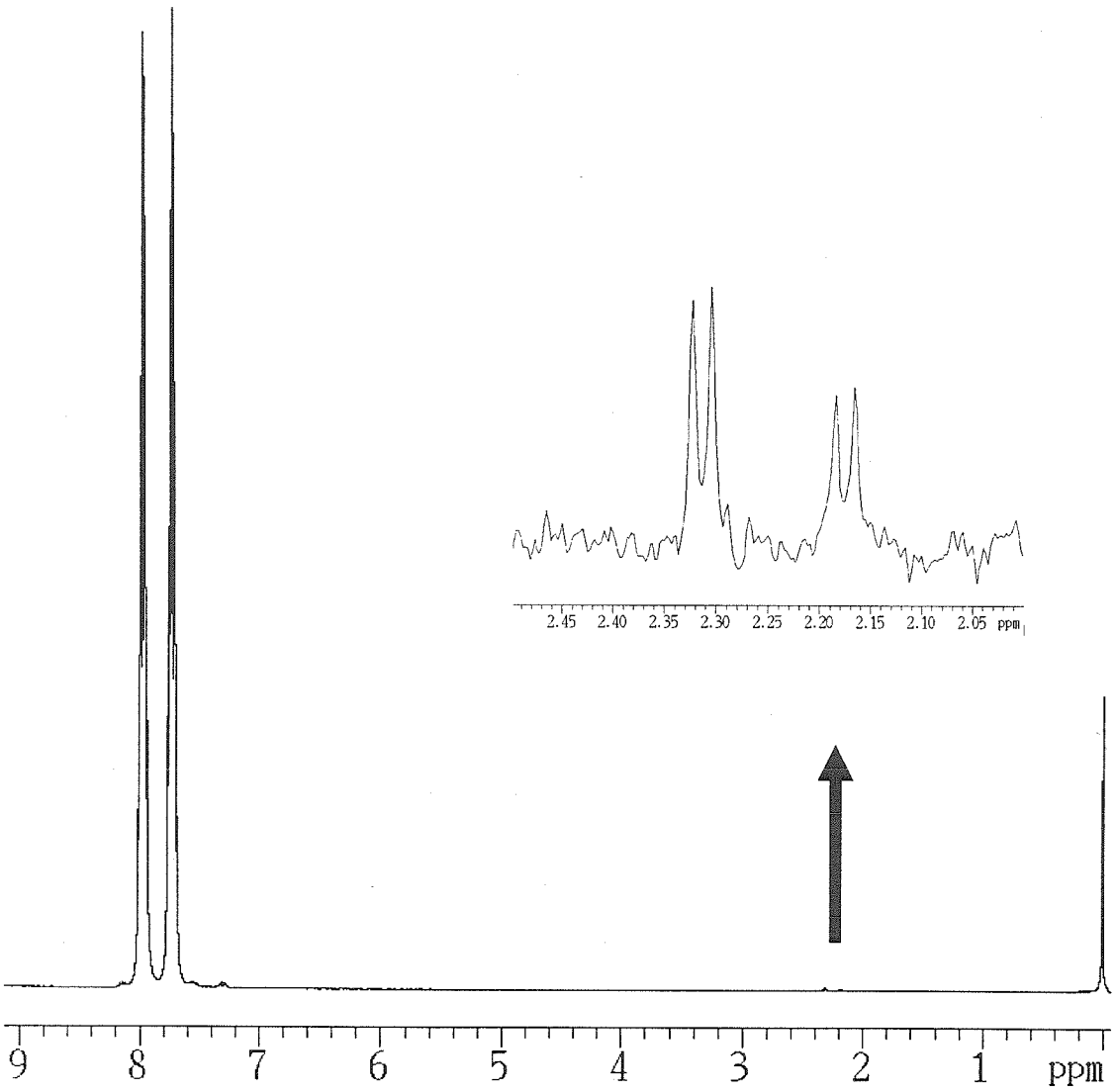


Figure 3.11. ^{19}F NMR spectrum of the sample of **3** used in this study. The chemical shifts are referenced to trifluoroacetic acid. The peaks at δ=7.8 belong to the fluorines of **3**. Expansion of the δ=2.2 ppm region shows that **2** is present as a trace contaminant. Integration of the two sets of peaks indicates that the relative amounts of **2** and **3** are 0.002 and 1, respectively. The contamination is likely to be originated from hexafluoroacetone and is propagated throughout the synthesis. Because of the twenty-fold faster rate of activation, an enrichment of **2** is observed in the target protein. Increase the preinduction incubation period could deplete the media of the contaminant before target protein synthesis.

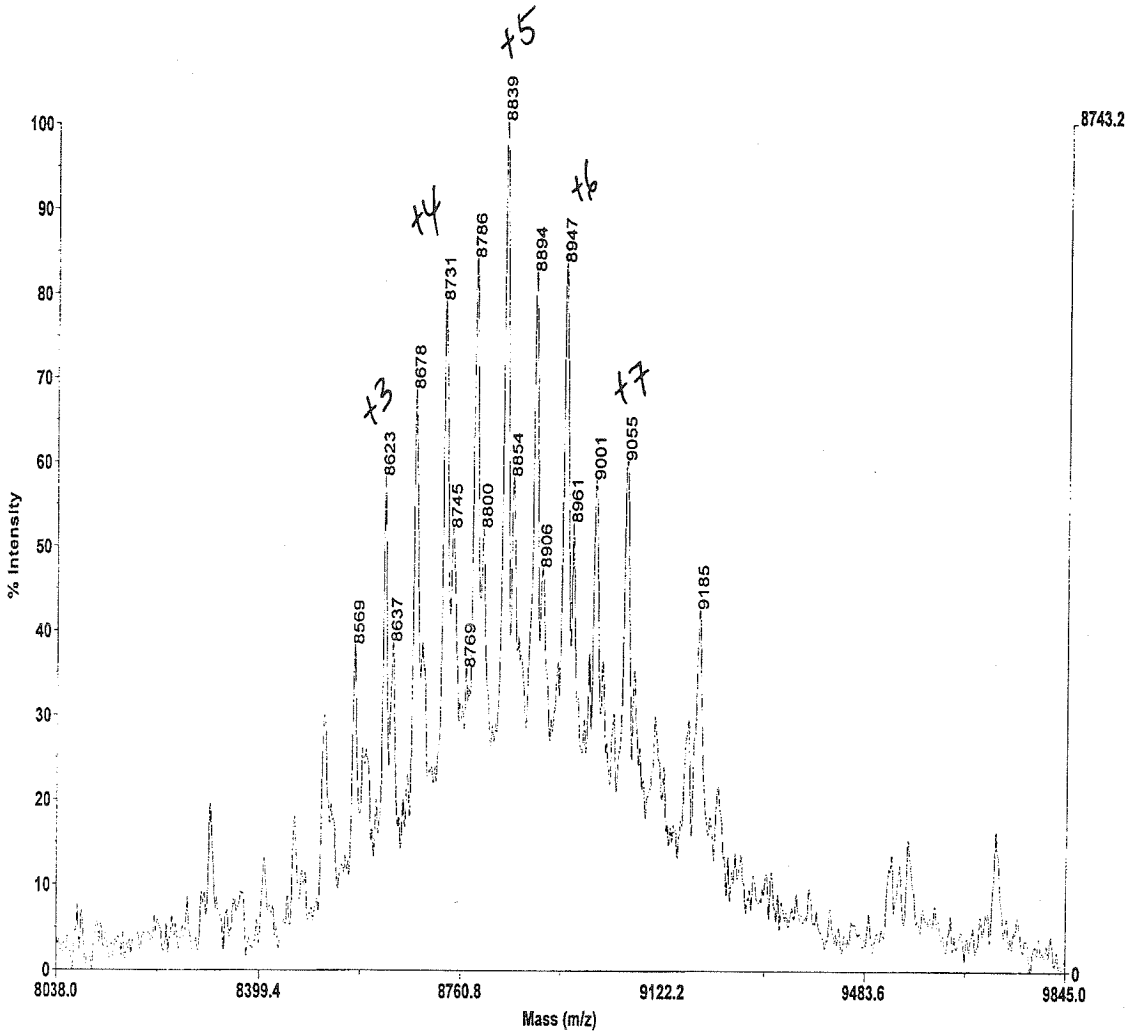


Figure 3.12. Mass spectrum of **3**-containing protein synthesized with 1.35 mM of **3**, and induced after *30* minutes of incubation after media shift. The wild type A1 mass is 8307. Each replacement of **1** with **3** increases the protein mass by 108 mass units. Full substitution of all eight leucines with **3** results in the mass of 9171 (+8). The highest peak observed corresponds to five out of eight substitutions. No significant change in leucine position composition is observed when compared to Figure 10.

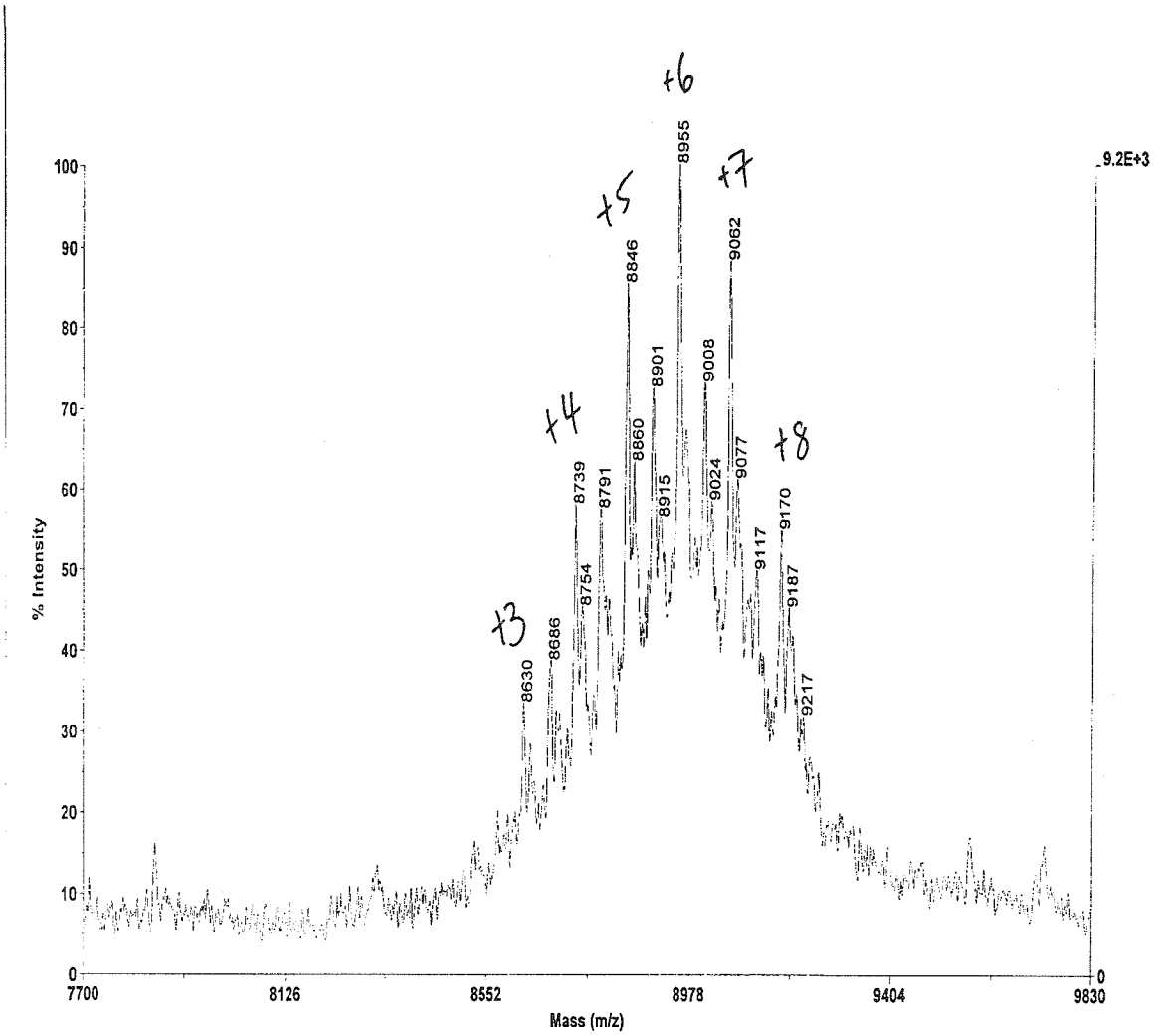


Figure 3.13. Mass spectrum of **3**-containing protein synthesized with 1.35 mM of **3**, and induced after *60* minutes of incubation after media shift. The highest peak observed corresponds to six out of eight substitutions. The amount of **2** in the protein starts to diminish as the incubation time is lengthened. The amount of **3** in the protein increases as compared to that observed in Figure 10.

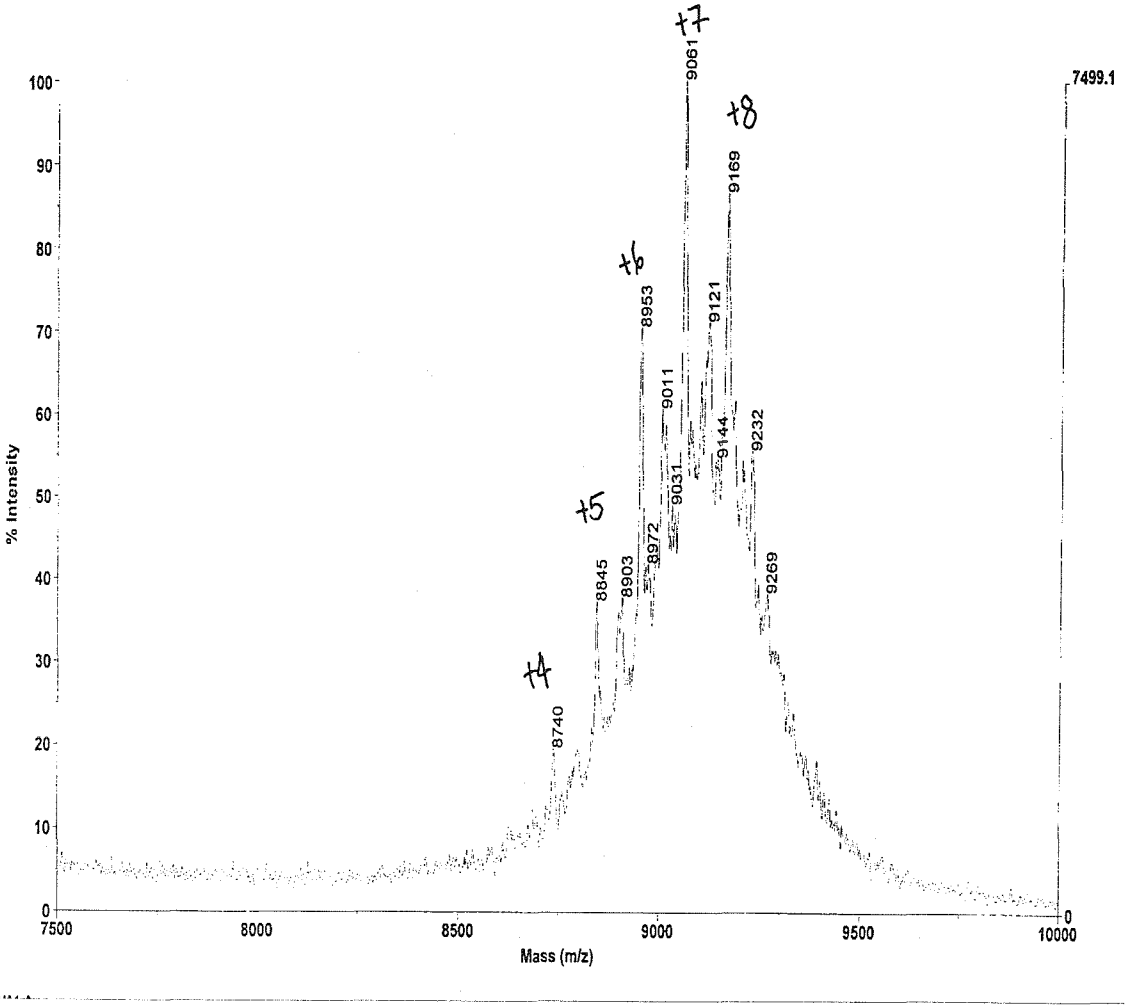


Figure 3.14. Mass spectrum of **3**-containing protein synthesized with 1.35 mM of **3**, and induced after *120* minutes of incubation after media shift. The highest peak observed corresponds to seven out of eight substitutions. In addition, the intensity of the peak correspond to maximum substitutions is also increased. The amount of **2** in the protein is significantly suppressed compared to that observed in figures 11-13. The longer incubation time before protein induction allows the contaminant **2** and residual **1** to be consumed by cells in the biosynthesis of other cellular proteins. The incubation period is not extended beyond 2 hours because cell death was observed.

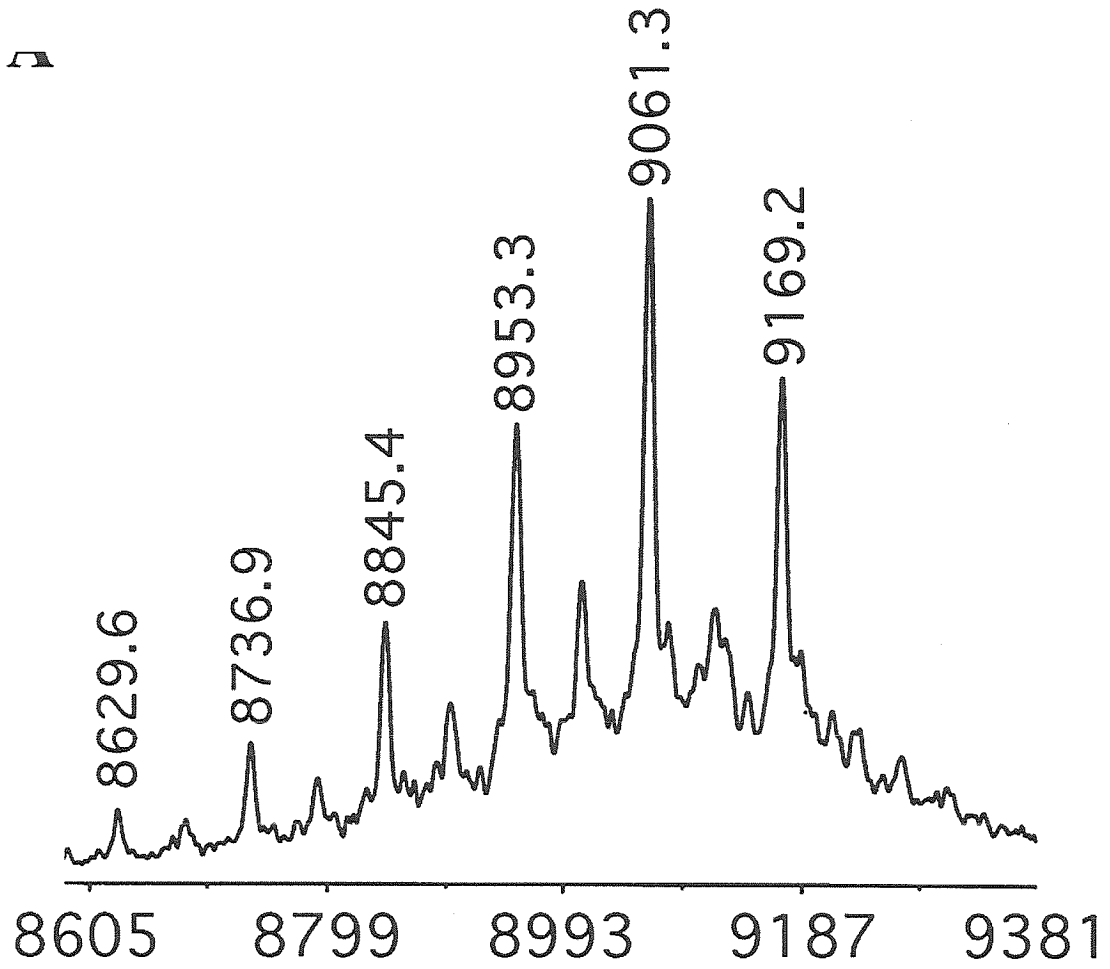


Figure 3.15. MALDI mass spectrum of HA1. HA1 is synthesized using the following condition: 2 hours of preinduction incubation; 3 mM of **3** supplied during A1 expression. The amino acid analysis of HA1 is shown in Table 3.4. The substitution level of **3** in the protein is 75%. The ratio of **2** to **3** in the protein was determined to be approximately 0.04.

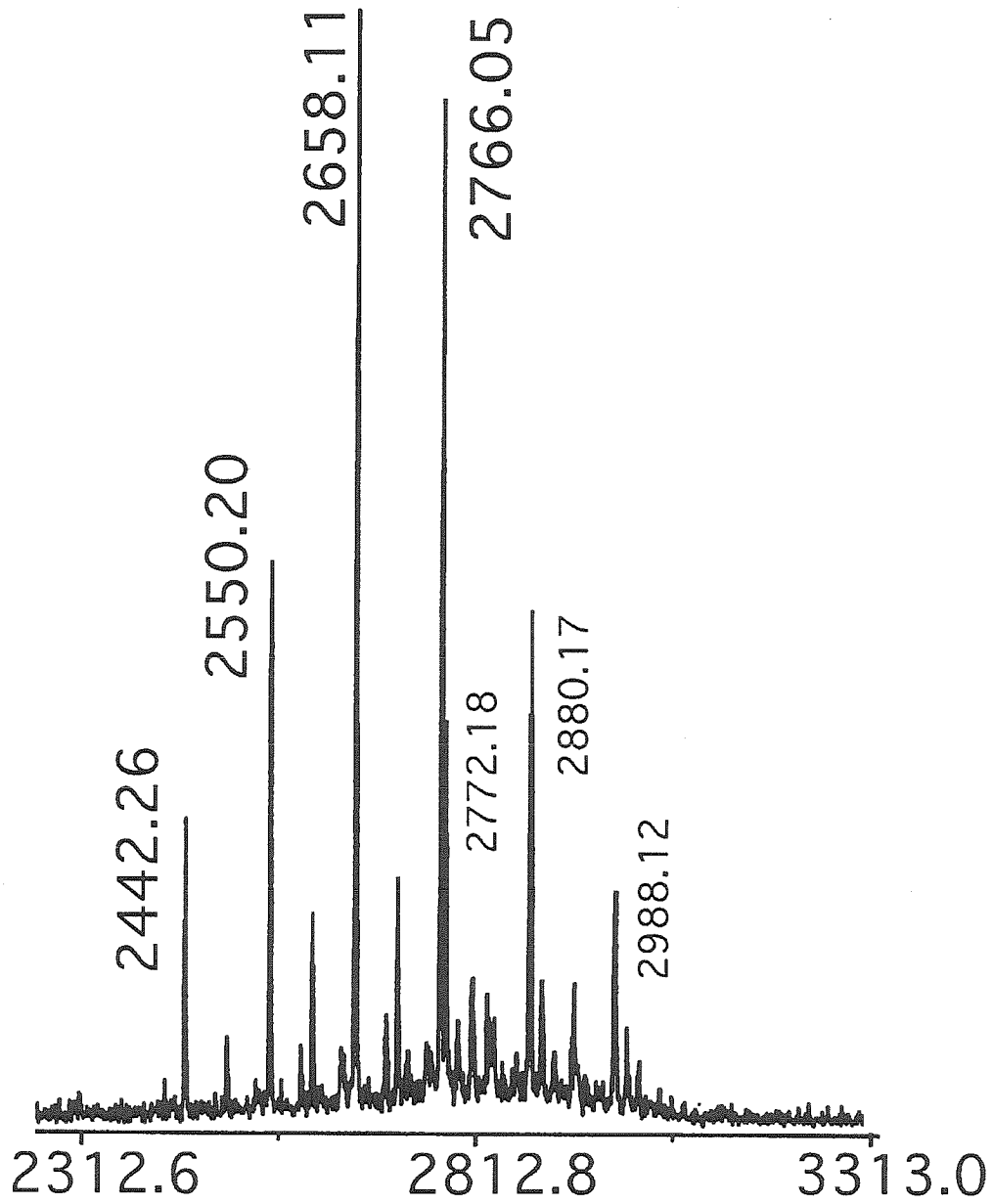


Figure 3.16. MALDI analysis of HA1 following trypsin digestion. Two fragments are present in the spectrum shown. The more abundant tryptic fragment contains the sequence LKNEIEDL-KAEIGDLNNTSGIR, and has three leucine sites. Fragments that correspond to 0, 1, 2, and 3 sites of replacement are observed at masses of 2442.26, 2550.20, 2658.11, and 2766.05, respectively. The other fragment has a wild type mass of 2772 and two sites of leucine substitution.

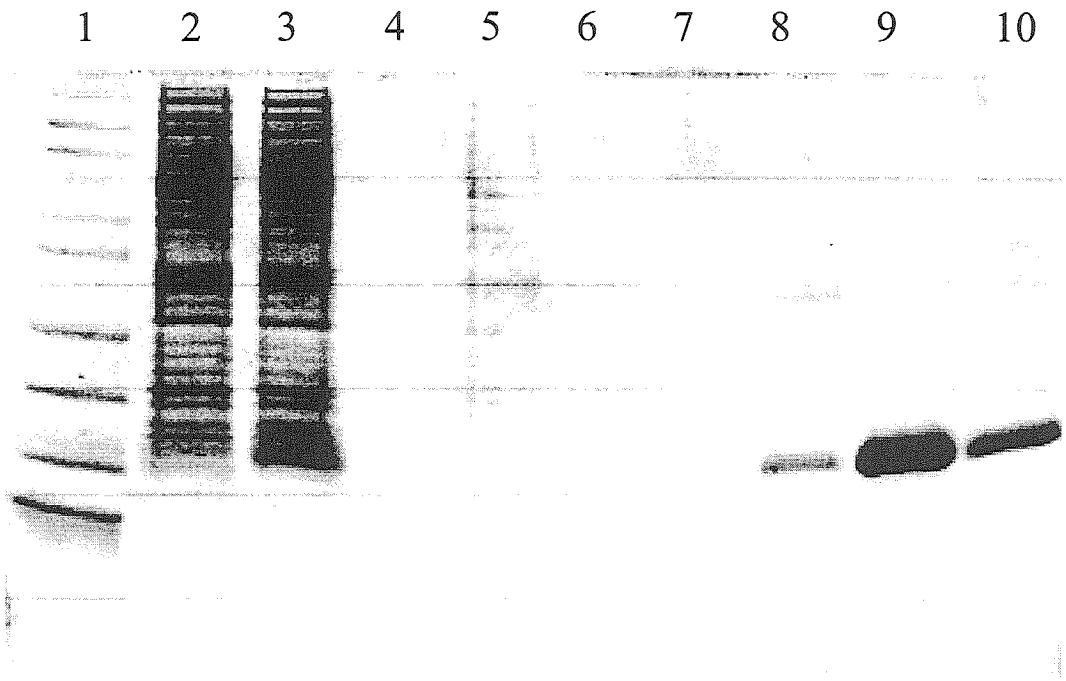


Figure 3.17. Large scale (1 L) expression and purification of HA1. Expression conditions are the same as that in the small scale experiment: 3 mM of **3** supplemented to expression media, followed by two hours of incubation before induction. Lane assignments: 1: molecular weight markers, the last three markers are 14 kD, 6 kD and 2 kD; 2: before induction; 3: three hours after induction of A1 expression; 4: blank; 5: pH 8.0, 8 M urea wash; 6: pH 6.5, 8 M urea wash; 7: pH 5.8, 8 M urea wash; 8: pH 4.5, 8 M urea elution fraction 1; 9: pH 4.5, 8 M urea elution fraction 2; and 10: pH 4.5, 8 M urea elution fraction 3. The three elution fractions were combined, dialyzed and lyophilized to yield A1 in a powder form. The yield of protein synthesis under these conditions is 8 mg/L. Composition of HA1 synthesized under large scale conditions were verified to be the same as that shown in figure 3.15 and 3.16.

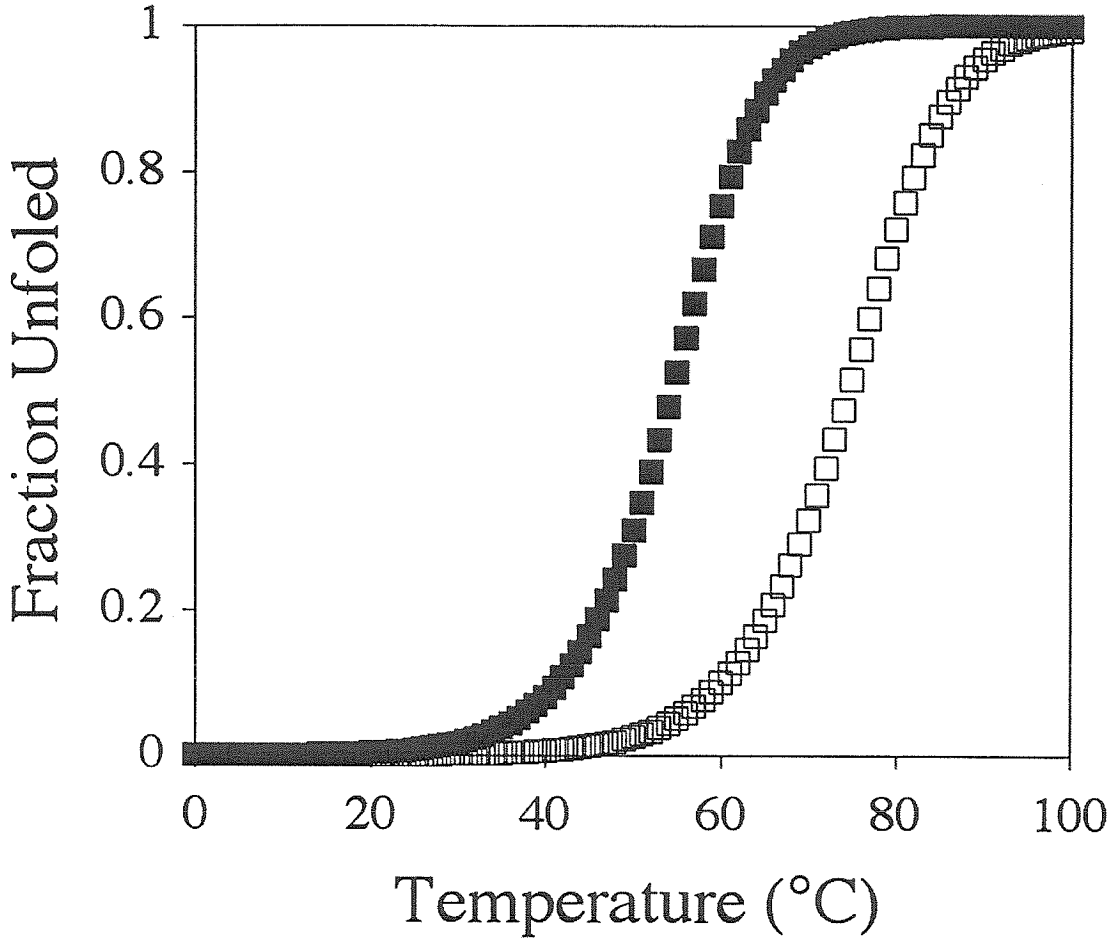


Figure 3.18. Comparison of thermal stabilities between A1 (closed symbols) and HA1 (open symbols). CD spectra were recorded at a protein concentration of 10 μM in PBS buffer, pH 7.4. T_m is increased from 54°C to 76°C upon substitution of 80% of the leucine sites in A1. Global thermodynamic fitting of the HA1 melting profiles shows that the free energy of unfolding is decreased by 3.6 kcal/mol upon fluorination.

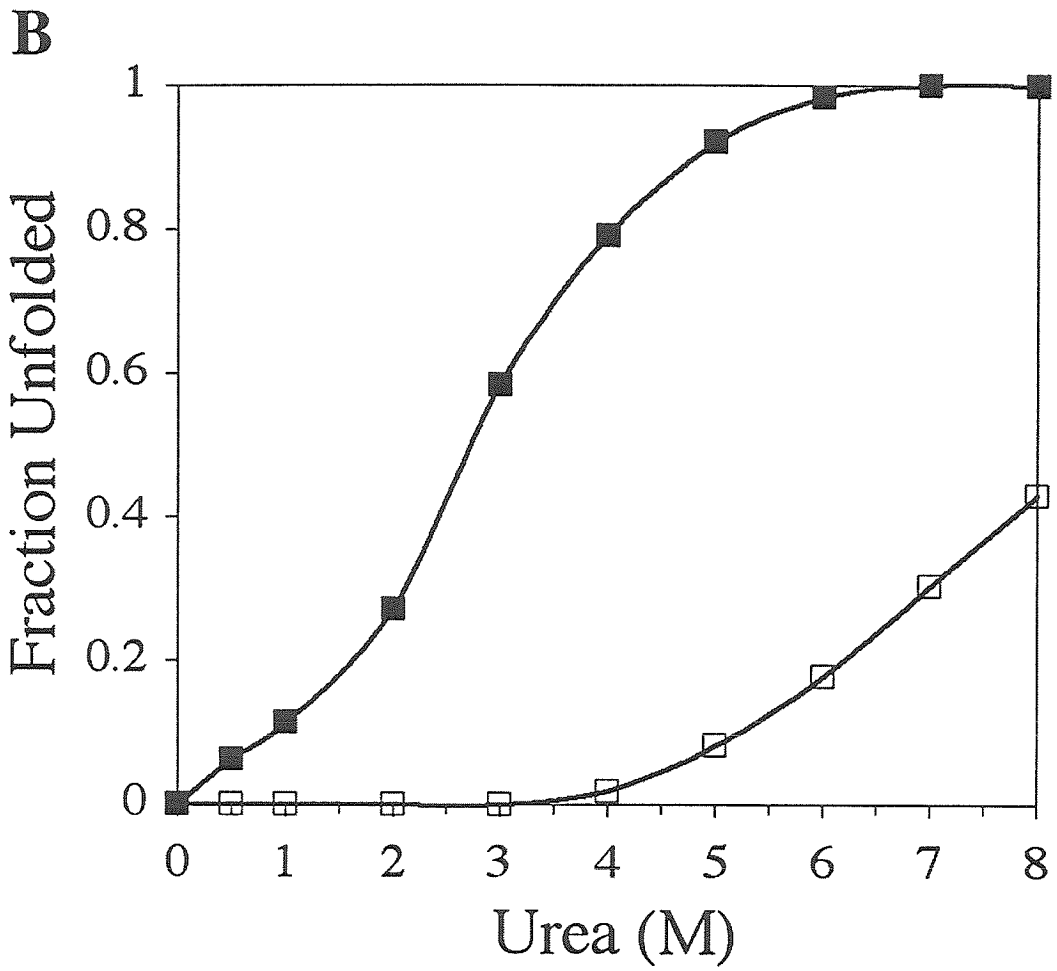


Figure 3.19. Comparison of the leucine zipper resistance to urea induced denaturation between A1 and HA1. CD spectra were recorded at a protein concentration of 10 μM in PBS buffer, pH 7.4. CD reports less than 50% denaturation of HA1 at 8 M urea. The C_m for A1 was previously shown to be 2.8 M (figure 3.7).

3.5 References

1. Tang, Y., Ghirlanda, G., Vaidehi, N., Kua, J., Mainz, D. T., Goddard, I. W., DeGrado, W. F., and Tirrell, D. A. (2001) *Biochemistry 40*, 2790-6.
2. Hecht, S. M. (1992) *Acc. Chem. Res. 25*, 545-552.
3. Cornish, V. W., Mendel, D., and Schultz, P. G. (1995) *Angew Chem Int Ed Engl 107*, 677-690.
4. Liu, D. R., Magliery, T. J., and Schultz, P. G. (1997) *Chem Biol 4*, 685-91.
5. Furter, R. (1998) *Protein Sci 7*, 419-26.
6. Liu, D. R., and Schultz, P. G. (1999) *Proc Natl Acad Sci U S A 96*, 4780-5.
7. Holmgren, S. K., Taylor, K. M., Bretscher, L. E., and Raines, R. T. (1998) *Nature 392*, 666-7.
8. van Hest, J. C., and Tirrell, D. A. (1998) *FEBS Lett 428*, 68-70.
9. van Hest, J. C., Kiick, K. L., and Tirrell, D. A. (2000) *J Am Chem Soc 122*, 1282-1288.
10. Sharma, N., Furter, R., Kast, P., and Tirrell, D. A. (2000) *FEBS Lett 467*, 37-40.
11. Kothakota, S., Fouriner, M. J., and Tirrell, D. A. (1995) *J Am Chem Soc 117*, 536-537.
12. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) *Science 292*, 498-500.
13. Kast, P., and Hennecke, H. (1991) *J Mol Biol 222*, 99-124.
14. Kukhar, V. P., and Soloshonok, V. A. (1995) *Fluorine Containing Amino Acids-Synthesis and Properties.*, Wiley, Chichester.
15. Parsons, J. F., Xiao, G., Gilliland, G. L., and Armstrong, R. N. (1998) *Biochemistry 37*, 6286-94.
16. Feeney, J., McCormick, J. E., Bauer, C. J., Birdsall, B., Moody, C. M., Starkmann, B. A., Young, D. W., Francis, P., Havlin, R. H., Arnold, W. D., and Oldfield, E. (1996) *J Am Chem Soc 118*, 8700-8706.
17. Renner, C., Alefelder, S., Bae, J. H., Budisa, N., Huber, R., and Moroder, L. (2001) *Angew Chem Int Ed Engl 40*, 923-925.
18. Rennert, O. M., and Anker, H. S. (1963) *Biochemistry 2*, 471-475.
19. Kothakota, S. (1993), Ph.D. Thesis, University of Massachusetts, Amherst.
20. McLachlan, A. D., and Stewart, M. (1975) *J Mol Biol 98*, 293-304.
21. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989) *Science 243*, 538-42.
22. O'Shea, E. K., Rutkowski, R., Stafford, W. F., 3rd, and Kim, P. S. (1989) *Science 245*, 646-8.

23. Wendt, H., Baici, A., and Bosshard, H. R. (1994) *J Am Chem Soc 116*, 6073-6074.

24. Gonzalez, L., Jr., Brown, R. A., Richardson, D., and Alber, T. (1996) *Nat Struct Biol 3*, 1002-9.

25. Kenar, K. T., Garcia-Moreno, B., and Freire, E. (1995) *Protein Sci 4*, 1934-8.

26. Betz, S. F., and DeGrado, W. F. (1996) *Biochemistry 35*, 6955-62.

27. Betz, S. F., Liebman, P. A., and DeGrado, W. F. (1997) *Biochemistry 36*, 2450-8.

28. Petka, W. A., Harden, J. L., McGrath, K. P., Wirtz, D., and Tirrell, D. A. (1998) *Science 281*, 389-92.

29. Petka, W. A. (1998), Ph. D. Thesis University of Massachusetts, Amherst.

30. Bottomley, M. J., Robinson, R. C., Driscoll, P. C., Harlos, K., Stuart, D. I., Aplin, R. T., Clements, J. M., Jones, E. Y., and Dudgeon, T. J. (1994) *J Mol Biol 244*, 464-8.

31. Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (1995) *Eur J Biochem 230*, 788-96.

32. Zhang, C., Ludin, C., Eberle, M. K., Stoeckli-Evans, H., and Reese, R. (1998) *Helv. Chim. Acta. 81*, 174-181.

33. XIng, X. C., Fichera, A., and Kumar, K. (2001) *Org. Lett. 3*, 1285-1286.

34. Hoffman, C., Tanke, R. S., and Miller, M. J. (1989) *J. Org. Chem. 54*, 3750-3751.

35. Li, D., Wang, E. D., and Wang, Y. L. (1997) *Acta Biochimica et Biophysica Sinica 29*, 591-596.

36. Kiick, K. L., Weberskirch, R., and Tirrell, D. A. (2001) *FEBS Lett 502*, 25-30.

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

38. Moitra, J., Szilak, L., Krylov, D., and Vinson, C. (1997) *Biochemistry 36*, 12567-73.

39. Hodges, R. S., Zhou, N. E., Kay, C. M., and Semchul, P. D. (1990) *Pept. Res. 3*, 125-137.

40. Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T., and Hodges, R. S. (2000) *J Mol Biol 300*, 377-402.

41. Sandberg, W. S., and Terwilliger, T. C. (1989) *Science 245*, 54-7.

42. Baldwin, E., Xu, J., Hajiseyedjavadi, O., Baase, W. A., and Matthews, B. W. (1996) *J Mol Biol 259*, 542-59.

43. Kono, H., Nishiyama, M., Tanokura, M., and Doi, J. (1998) *Protein Eng 11*, 47-52.

44. Li, T., Guo, N., Xia, X., Wang, E. D., and Wang, Y. L. (1999) *Biochemistry 38*, 13063-9.

45. Mursinna, R. S., Lincecum, T. L., Jr., and Martinis, S. A. (2001) *Biochemistry 40*, 5376-81.

46. Kiick, K. L., van Hest, J. C., and Tirrell, D. A. (2000) *Angew Chem Int Ed Engl 39*, 2148-2152.

47. Hartlein, M., and Madern, D. (1987) *Nucleic Acids Res 15*, 10199-210.

48. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F., and Tirrell, D. A. (2001) *Angew Chem Int Ed Engl 40*, 1494-1496.

CHAPTER 4

Attenuation of the Editing Activity of the *Escherichia coli* Leucyl-tRNA Synthetase Allows Incorporation of Novel Amino Acids into Proteins *in vivo*

This chapter has been submitted to Biochemistry as a full paper. (Authors, Y. Tang and D. A. Tirrell)

4.0 Abstract

Incorporation of unnatural amino acids into proteins is expanding the scope of protein engineering. The fidelity of translation is dependent on the specificity of the aminoacyl-tRNA synthetases (aaRS). The aaRSs that activate the hydrophobic amino acids leucine, isoleucine, and valine employ a proofreading mechanism that hydrolyzes noncognate aminoacyl adenylates and misaminoacylated tRNAs. Discrimination between structurally similar amino acids by these AARSs is believed to operate by a double-sieve principle, wherein a separate editing domain governs hydrolysis based on the size and hydrophilicity of the amino acid side chain. Leucyl-tRNA synthetase (LeuRS) relies on its editing function to correct misaminoacylation of $tRNA^{Leu}$ by isoleucine and methionine. Thr252 of *E. coli* LeuRS has been shown previously to be important in defining the size of the editing cavity. Here we report the isolation and characterization of three LeuRS mutants with point mutations at this position (T252Y, T252L and T252F). The proofreading activity of the synthetase is significantly impaired when an amino acid bulkier than threonine is introduced. The misaminoacylation rate of $tRNA^{Leu}$ by isoleucine and valine increases with increasing size of the amino acid substituent at position 252, and the noncognate amino acids norvaline and norleucine are inserted efficiently at leucine sites of recombinant proteins under conditions of constitutive overexpression of the T252Y mutant in *E. coli*. In addition, the unsaturated amino acids allylglycine, homoallylglycine, homopropargylglycine, and 2-butynylalanine all support protein synthesis. These results demonstrate that programmed manipulation of the editing cavity can allow *in vivo* incorporation of novel protein building blocks.

4.1 Introduction and Background

The fidelity of protein biosynthesis is controlled in large part by the aminoacyl-tRNA synthetases (aaRSs) (*1-3*). The aaRSs catalyze the joining of tRNAs and amino acids in two steps as outlined in Scheme 1: the cognate amino acid is first converted to the aminoacyladenylate, and subsequently transferred to the 3'–end of the corresponding tRNA. The substrate specificity of the aaRSs is remarkable, as each enzyme must distinguish small differences among amino acid side chains in order to maintain accurate translation and cell viability.

The synthetases associated with the aliphatic hydrophobic amino acids, including leucine (**1**), isoleucine (**2**), and valine (**3**), must distinguish among nearly isosteric substrates. All three synthetases are multidomain, monomeric enzymes, and all belong to the Class I synthetase family (*2*). Isoleucyl-tRNA synthetase (IleRS) and valyl-tRNA synthetases (ValRS) achieve amino acid discrimination by a "double-sieve" mechanism (*4-7*). IleRS rejects isomeric (e.g. **1**) or larger amino acids at its active site and rejects smaller amino acids such as **3** at its editing site (*5*). As a result, IleRS mischarges **3** at a frequency of only 1 in 3000, even though **3** is activated 180 times more slowly than **2** (*8*). In analogous fashion, ValRS rejects larger amino acids such as **1** and **2** at its synthetic site and proofreads threonine at its editing site based on the hydrophilicity of the threonine hydroxyl group (*7*). The crystal structures of IleRS (*9, 10*), ValRS (*11*) and LeuRS (*12*) have been elucidated, and each shows the editing site embedded in a readily identified editing domain. In all three structures, the synthetic active sites adopt the characteristic Rossmann fold found in all Class I synthetases. The editing domains are inserted into the Rossmann folds via flexible linkers and form so-called connective polypeptide 1 (CP1) domains (*13*). Both pretransfer editing of misadenylated amino acids and posttransfer editing of misaminoacylated tRNAs by the editing domains of IleRS and ValRS have been observed (*7, 14*). The CP1 domain in ValRS is outfitted with two subsites for recognition of threonine (*11*): one specifically for Thr-tRNA^{Val} in posttransfer editing, and a putative pocket for binding Thr-AMP in pretransfer editing. Bishop et al. recently suggested a postpre-prepre editing mechanism (*15*), in which posttransfter editing primes the synthetase toward proofreading and is the dominant pathway under *in vivo* conditions. Mutagenesis of residues involved in the proofreading processes of both IleRS and ValRS results in the accumulation of misaminoacylated cognate

tRNAs (*9, 15, 16*). Amino acid **3** was appended to $tRNA^{Ile}$ *in vitro* when D342 of *E. coli* IleRS, a residue critical for the translocation of Val-$tRNA^{Ile}$ to the editing site, was mutated (*15*). The nonnatural amino acid 2-aminobutyric acid (αbu) was inserted into more than 20% of valine positions in cellular proteins *in vivo*, by ValRS mutants that were isolated by screening mutagenized *E. coli* strains (*17*). The mutations were all located within the CP1 region of the synthetase.

Leucyl-tRNA synthetase (LeuRS) is an 860-amino acid monomeric protein (*18*). Unlike the well-studied proofreading mechanisms of IleRS and ValRS, the "double-sieve" behavior of LeuRS is poorly understood. The CP1 domain in LeuRS is approximately two hundred amino acids in length (*12*), and the residues involved in proofreading have not been identified. The relative contributions of pre- and posttransfer proofreading in LeuRS are also not known (*19*). Chen et al. found that disruption of the editing domain, either by site-directed mutagenesis of alanine 293 (*20*) or by PCR domain insertion (*21*), results in formation of Ile-$tRNA^{Leu}$ and Met-$tRNA^{Leu}$ *in vitro*.

Residues in a threonine-rich region in *E. coli* IleRS CP1 are crucial to IleRS proofreading activity (*9*). Using alanine scanning mutagenesis of a homologous region in the CP1 domain of *E. coli* LeuRS, Mursinna et al. have provided the first insight into the structural basis of LeuRS editing (*22*). They identified T252 to be an essential residue in the LeuRS posttransfer editing pathway. Alignment of LeuRS CP1 domain sequences from different organisms shows this threonine is rigorously conserved among species (Figure 4.1). The hydroxyl side chain of T252 forms a hydrogen bond with a water molecule, as observed in the crystal structure (12). The editing domain is shown in Figure 4.2 in a side-view representation with T252-water complex shown in space-filled model. The oxygen atom on the threonine side chain is shown in red and the oxygen atom of the hydrogen bonded water molecule is shown in cyan. The side chain of T252 and the putative posttransfer editing cavity are oriented towards the synthetic active site (partially shown in grey). Figure 4.3 shows the molecular surface representation of the cavity. It can be clearly seen that the cavity consists of a groove in which a linear chain, such as the terminal methylene groups in **2** and the thioether side chain of **4**, can be accommodated. The water molecule is shown in blue and it occupies one side of this groove. The importance of T252 (not visible in this representation) in preventing **1** from binding to this pocket can be visualized by manually removing the coordinates of the γ carbon, γ oxygen of T252, and the water molecule in the structure file (so now this residue is Ala). The

remodeled surface (Figure 4.3, bottom) reveals an additional groove juxtaposed to the existing cavity, which could accommodate the isobutyryl side chain of **1**. Indeed, purified LeuRS-T252A rapidly hydrolyzed Leu-tRNA^{Leu} *in vitro* (*22*). Wild type LeuRS proofreads against **2** and methionine (**4**), both of which lack branching at the γ carbon atom. T252 thus facilitates substrate recognition at the editing site based on the sizes of substituents at the γ carbons of amino acid side chains during posttransfer editing. Amino acids with a tertiary γ carbon, such as **1**, are too bulky to fit in the pocket, while those with secondary γ carbons can fit and are hydrolyzed away from the misaminoacylated tRNA^{Leu}.

Our laboratory is interested in the use of nonnatural amino acids *in vivo* for applications in protein engineering and biomaterials synthesis (23-32). Of special interest are amino acids with novel physical (*24, 25*) and chemical properties (27, 31, 32). For example, incorporation of fluorinated analogs of leucine, such as trifluoroleucine and hexafluoroleucine, has been shown to stabilize coiled-coil domains with respect to thermal or chemical denaturation *(24, 25, 33)*. In addition, we have tested the *in vivo* translational activities of several other amino acids that are structurally similar to leucine, including norvaline (**5**) and norleucine (**6**). *In vivo* protein expression studies showed that neither of the straight-chain analogs supported protein synthesis in cultures depleted of **1**. Based on our interpretation of the LeuRS editing mechanism, we attribute the lack of translational activities of **5** and **6** to proofreading by LeuRS, similar to that observed for **4**. An analogous mechanism controls the editing specificity of ValRS, which rapidly hydrolyzes αbu-tRNA^{Val}, while leaving Val-tRNA^{Val} intact (*5*).

In this report, we present a successful attempt to disrupt the editing function of LeuRS through directed mutagenesis of the key residue T252. The premise of this work was that the editing activity of LeuRS should be attenuated by replacement of T252 with a bulkier amino acid residue. We studied *in vitro* aminoacylation kinetics of three T252 LeuRS mutants and discovered that these enzymes indeed exhibit impaired editing activity toward natural amino acids **2** and **3**. One such mutant, T252Y, was chosen for subsequent *in vivo* studies. When this mutant was overexpressed in *E. coli,* the previously translationally silent amino acids **5** and **6** were incorporated into recombinant proteins with high efficiency. In addition, we also tested the translational activities of unsaturated analogs allylglycine (**7**), propargylglycine (**8**), α-cyanoalanine (**9**), homoallylglycine (**10**), homopropargylglycine (**11**), and 2-butynylglycine (**12**). Analogs

10-12 have been tested previously as methionine surrogates *(30, 32)*. We find that the T252Y LeuRS mutant enables incorporation of **7**, **10-12** into *E. coli* proteins.

4.2. Materials and Methods

4.2.1 Materials.

Amino acids **1-8** were purchased from Sigma; **9** was purchased from ICN (Irvine, CA). Amino acids **10-12** were prepared as described by van Hest et al. (*32*), i.e. by alkylation of diethyl acetamidomalonate by the alkyl tosylate, followed by base hydrolysis, decarboxylation and enzymatic deacylation. [^{3}H]-labeled amino acids were purchased from Amersham Pharmacia Biotech (Piscataway, NJ). [^{32}P]-labeled sodium pyrophosphate was purchased from NEN Life Sciences. Oligonucleotides were synthesized at the Caltech Biopolymer Synthesis Center. General cloning was performed in XL-1 blue cells. Expression strain SG13009 was purchased from Qiagen.

4.2.2. Cloning and Mutagenesis.

Cloning and Mutagensis. The LeuRS gene was cloned directly from *E. coli* genomic DNA using four-primer PCR. The flanking primers were 53Sph: 5'- GGACCACTGG-CTGGCATGCAAGAGCAATAC-3' and 35Hind: 5'- CGCTTCCTCCCAAGCTTAGCCAACGACC-3'. The introduced restrictions sites are underlined. The remaining pair of complementary oligos was designed to carry the desired mutations at position 252, 53*NNN*: 5'- CTACCCGCCCGGAC*NNN*T-TTATGGGTTGTAC-3', where *NNN* codes for the desired amino acid (WT: *ACC*; T252Y: *TAC*; T252L: *CTC*; T252F: *TTC*). 35*NNN* is the oligonucleotide that is complementary to 53*NNN*. In the first reaction, primers 53Sph and 35*NNN* were used to yield a 780 bp product. In the second reaction, primers 53*NNN* and 35Hind were used to yield a 1900 bp product. The products were then mixed and subjected to further amplification in the presence of 53Sph and 35Hind. *Pwo* polymerase was used throughout to minimize random mutations. The resulting 2600 bp DNA fragment was gel-purified and digested overnight with *SphI* and *HindIII*. The digested fragment was ligated into the expression plasmid pQE32 to yield p32leus, p32T252Y, p32T252L, and p32T252F, which encode the wild-type, T252Y, T252L, and T252F enzymes,

respectively. The cloned enzymes contained the N-terminal leader sequence MRGSHHHHHGIR. The fidelity of PCR cloning and the presence of the desired mutations were checked by DNA sequencing. An *MslI* restriction site is removed upon successful mutagenesis and was used as a simple indicator of mutation at position T252.

4.2.3. Synthetase Purification.

SG13009 cells carrying the pREP4 repressor plasmid were transformed with p32leus, p32T252Y, p32T252L and p32T252F, respectively, to yield the expression strains. Protein expression was induced at OD_{600} = 0.5 with 1 mM IPTG. After three hours, the cells were harvested and lysed with sonification. The enzymes were purified using Ni-NTA agarose resins under native conditions according to the manufacturer's instructions (Qiagen). The proteins were purified to >95% as indicated by SDS-PAGE. Proteins were stored in Buffer A (50 mM Tris-HCl, 1 mM DTT)/50% glycerol. Aliquots were flash frozen in liquid nitrogen and stored at -80°C. The concentration of each enzyme was determined by measuring the absorbance at 280 nm under denaturing conditions.

4.2.4. ATP-PP_i Exchange Assay.

The assay was performed according to literature procedures (*16*). The assay buffer contained 50 mM HEPES (pH 7.6), 20 mM $MgCl_2$, 1 mM DTT, 2 mM ATP and 2 mM [^{32}P]-PP_i (0.5 TBq/mol). The concentration of enzyme was 75 nM. The amino acid concentration ranges varied depending on the activity of the enzyme toward the substrate (**1**: 10-500 μM; **5, 6**: 50-5000 μM; **2-4**: 200-8000 μM). Aliquots (15 μL) were quenched in 500 μL quench solution (200 mM PP_i, 7% w/v $HClO_4$ and 3% w/v activated charcoal). The charcoal was washed twice with 10 mM PP_i, 0.5% $HClO_4$ and counted. The results reported in Table 4.1 are averages from triplicate experiments.

4.2.5. Aminoacylation Assay.

Aminoacylation assays were carried out as described previously (16, 34, 35). Assays were performed at 37°C in 30 mM HEPES, pH 7.4, 10 mM $MgCl_2$, 1 mM DTT and 2 mM ATP. Purified *E. coli* tRNA mixture (Roche Biochemical) was used in the assay at a final concentration of 12 mg/ml. tRNA

solution was treated at 80°C for 3 minutes and allowed to cool slowly to 37°C before use. For reactions involving **1**, 20 μM [^{3}H]-Leu (2000 dpm/pmol) and 10 nM LeuRS were added. For assays involving noncognate substrates **2** and **3**, 20 μM [^{3}H]-aa (6000 dpm/pmol) and 500 nM LeuRS were added. Reactions were initiated by adding enzyme; 10 μl aliquots were quenched on filter disks pretreated with 5% trichloroacetic acid (TCA) and 100 μM of the corresponding amino acid. The filter disks were washed with cold 5% TCA three times, and counted.

4.2.6. Expression Plasmids.

Plasmid pA1EL (a derivative of pQE9 (Qiagen)) was used as the template for site-directed mutagenesis (*25*). The synthetic leucine zipper protein A1 (36) was inserted at the *BamHI* restriction site of pQE9 to yield pQEA1. The *E. coli leuS* gene with its endogenous promoter was cloned from *E. coli* genomic DNA and inserted into the *NheI* site of pQEA1 to yield the template pA1EL. The Quickchange protocol (Stratagene) was used to introduce the mutation at position T252 in the *leuS* gene. Primers 53TAC and 35GTA were used in the reaction to generate the tyrosine mutation. The integrity of the entire plasmid was verified through DNA sequencing. The plasmid carrying the overexpressing T252Y mutant was named pA1T252Y. The level of LeuRS overexpression was verified using SDS-PAGE of whole cell lysates from overnight cultures.

4.2.7 Analog Incorporation Assay.

The leucine auxotrophic strain LAM1000 (24) was transformed with pA1EL or pA1T252Y, and with pREP4 to yield the A1 expression strains LAM1000/pA1EL or LAM1000/pA1T252Y. Growth and expression were performed in supplemented M9AA media (M9 mediium, amino acids at 40 mg/L, 1 mM $MgSO_4$, 1 mM $CaCl_2$, 0.4 wt% glucose, 5 μg/ml thiamin, 200 μg/ml ampcillin and 25 μg/ml kanamycin). M9AA (200 mL) was inoculated with 1 mL of an overnight culture of the expression strain. The cells were grown to an OD_{600} between 0.9 and 1.0., pelleted and washed with cold 0.9% NaCl three times. The cells were then resuspended in fresh M9AA medium with 16 natural amino acids (Leu, Met, Ile and Val were not added). Aliquots (10 mL) of the resuspended cells were added to different test tubes, each containing one of the amino acid analogs at a concentration of 320 mg/L. Leucine was added to a separate 10 mL

solution at 40 mg/L as a positive control. After 10 min, 1 mM IPTG was added to induce protein expression. After three hours, the cells were collected by centrifugation (5000 g, 10 min, 4°C) and resuspended in 600 μL of Buffer B (8.0 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris, pH 8.0) and frozen at -80°C. Whole cell lysates were analyzed by SDS-PAGE.

4.2.8. Protein Composition Analysis.

The target protein A1 in 600 μL of whole cell lysate was purified on a Ni-NTA spin column (Qiagen) according to the manufacturer's instructions. A1 protein was eluted in 400 μL Buffer B, pH 4.5. A portion of the eluent (10 μL) was diluted in 450 μL of 50 mM $(NH_4)_2CO_3$. The pH was adjusted to optimal trypsin working pH (8.0). Trypsin stock solution (5μL, 20 μg/200 μL) was added and the sample was incubated at room temperature overnight. The reaction was quenched by addition of 2 μL trifluoroacetic acid (TFA). The reaction mixture was subjected to C18 ZipTip (Millipore) purification and peptide fragments were eluted with 3 μL of 0.1 TFA, 50% CH_3CN. One μL of eluent was used for MALDI analysis. The remaining 350 μL of spin column eluent was dialyzed against water extensively and lyophilized to a fluffy powder. The powder was sent directly for MALDI and amino acid analysis.

4.3. Results and Discussion

As shown in Figure 4.3, the T252-H_2O complex is positioned at the entrance of a putative editing site. Replacing T252 with a larger amino acid could block the binding of misaminoacylated tRNA species and result in relaxed LeuRS substrate specificities *in vivo*. We replaced T252 with bulkier residues leucine, phenylalanine, and tyrosine. The van der Waals volumes of the side chains of these amino acids are 76 $Å^3$, 87 $Å^3$, and 93 $Å^3$, respectively (the volume of threonine side chain is 45 $Å^3$) (37). The mutations introduce different degrees of steric hindrance to the editing groove, and could result in varying rates of $tRNA^{Leu}$ misaminoacylation.

4.3.1 Cloning and Mutagenesis

Genes encoding the wild-type LeuRS and the mutants T252L, T252F, and T252Y were obtained from four-primer PCR with *E. coli* genomic DNA as the template. Enzymes were expressed in *E. coli* by construction of individual pQE32 derivatives containing the gene in frame with the N-terminal His_6 tag from the vector. The enzymes were purified under native conditions using Ni^{2+} affinity chromatography to > 95% purity as estimated by SDS-PAGE. The typical yields of enzymes were c.a. 4 mg/L.

4.3.2. Amino Acid Activation in vitro

We examined the kinetics of the ATP-dependent amino acid activation reaction using purified, wild-type His_6-synthetase. We compared the rates of activation of canonical amino acids **1-4** and noncanonical amino acids **5** and **6** by using an ATP-PP_i exchange assay. The kinetic parameters are listed in Table 4.1. The K_m reported for **1** is consistent with that reported in the literature (38), while the k_{cat} value is slightly lower. Compared to that of **1**, the catalytic efficiencies of LeuRS towards natural, noncognate substrates **2**, **3**, and **4** were decreased by factors of 3300, 9600 and 5200, respectively. Under normal cellular conditions, these amino acids do not pose significant threats to LeuRS fidelity. In contrast, LeuRS displayed higher activities towards the noncanonical analogs **5** and **6**. Hydrocarbon side chains are more energetically favored in the LeuRS active site, as illustrated by the five-fold increase in the k_{cat}/K_m of LeuRS towards **6** compared to the isosteric **4**. The 50-fold decrease in the k_{cat}/K_m of LeuRS towards **2** relative to **5** suggests that a branched methyl group at the β carbon significantly decreases substrate binding affinity.

Both **5** and **6** displayed remarkably high affinity for the enzyme. Substituting the isobutyryl side chain present in **1** with a smaller propyl group found in **5** decreased the rate by only 120-fold. A comparable 200-fold decrease was noted in ValRS activity when the isopropyl side chain of **3** was replaced with the ethyl group found in αbu (*5*). The proofreading mechanism of ValRS increases its selectivity for **3** over αbu by an additional 60-fold, resulting in a final discrimination factor of 12,000 (39). *In vivo* translation studies have shown that while trifluoroleucine, an analog that is activated by LeuRS 250-fold slower compared to **1** can substitute for **1** in proteins quantitatively (24) **5** is completely inactive during biosynthesis. (A conflicting report shows **5** can be inserted at leucine positions in human hemoglobin at <

1% substitution rate, (40)). The comparable activation rates of **5** and trifluoroleucine *in vitro* and the differences in their translation activities *in vivo* strongly indicate that LeuRS achieves additional selectivity over **5** through an editing step. Similarly, the linear, four-carbon side-chain analog **6** is also subject to proofreading by LeuRS. The activities of mutant LeuRS enzymes toward **1, 5** and **6** were essentially identical to these reported for the wild type enzyme, in agreement with previous conclusion that activities at the synthetic site and the editing site are mutationally exclusive (*16*).

4.3.3. Aminoacylation Kinetics of Mutants

The abilities of the mutant LeuRS enzymes to aminoacylate $tRNA^{Leu}$ with amino acids **1**, **2** or **3** were measured by standard TCA precipitation assays and the results are shown in Figure 4.4. All four enzymes aminoacylate $tRNA^{Leu}$ with **1** at comparable rates, indicating mutations of T252 in the CP1 domain have no effect on the rate of **1**-$tRNA^{Leu}$ synthesis. For reactions containing **2** or **3**, high concentrations (500 nM) of LeuRS were used to facilitate detection. Wild type (filled circles) enzyme catalyzed no detectable levels of **2**-$tRNA^{Leu}$ and **3**-$tRNA^{Leu}$ synthesis. This is expected as any misaminoacylation by the synthetic active site is to be corrected by the proofreading mechanism of the CP1 domain. The three mutants in this study all displayed impaired editing activity toward **2** and **3**, confirming the importance of T252 in the editing process. Poor activation of **3** in the synthetic active site results in the lower rates of **3**-$tRNA^{Leu}$ synthesis as compared to that of **2**-$tRNA^{Leu}$. The rate of misaminoacylation increases as the size of residue 252 increases. T252Y mutant (filled squares) has the fastest rate of misaminoacylation, followed by T252F (empty triangles) and T252L (filled diamonds). The exact orientations of the substituted side chains are unknown without crystallographic data. The observed correlation between side chain volumes (Tyr > Phe > Leu > Thr-H_2O) and misaminoacylation rates partially supports our initial hypothesis that a larger amino acid at the 252 position would block the binding of misaminoacylated $tRNA^{Leu}$ at the editing site shown in Figure 4.3, and reduce the ability of LeuRS to proofread.

4.3.4. In vivo properties of mutant synthetases

The T252Y mutant displayed the highest activity in misaminoacylating $tRNA^{Leu}$ based on the *in vitro* aminoacylation assay. Overexpression of this mutant could lead to accumulation of misaminoacylated $tRNA^{Leu}$ *in vivo*, and result in the incorporation of amino acid analogs that were otherwise edited by the wild type enzyme. The expression plasmid pA1EL was the template for site-directed mutagenesis. The construction and application of this plasmid has been previously reported (*25*). It contains a gene encoding for a 74-residue synthetic leucine zipper protein. This protein contains eight leucine residues, of which six can be easily detected using tryptic mass spectrometry (Figure 4.5). The plasmid carries a copy of the wild type *leuS* gene, which encodes *E. coli* LeuRS under its endogenous, constitutive promoter. *E. coli* host transformed with pA1EL has elevated LeuRS activity compared to untransformed hosts. We used this overexpression strain to incorporate the sluggish amino acid hexafluoroleucine in place of **1** (*25*). The T252Y mutation was introduced via site-directed mutagenesis. The resulting plasmid, pA1T252Y was transformed into the leucine auxotroph strain LAM1000/pREP4 to yield the overexpression strain LAM1000/pA1T252Y/pREP4. LAM1000 contains the wild-type *leuS* gene on its chromosome.

We compared levels of target protein expression in host strains carrying either pA1EL or pA1T252Y. To test the ability of an analog to substitute for **1** during protein translation, we adopted the medium shift procedure to deplete the expression medium of **1**. Cells were first grown in M9 medium supplemented with all twenty natural amino acids. At the time of protein induction, cells were transferred to a new medium containing nineteen amino acids, either without **1** (negative control); with **1** (positive control) or with one of the analogs. Protein expression was induced with IPTG and continued for three hours. The levels of target protein accumulation were visualized with SDS-PAGE and are shown in Figure 4.6. As expected, when the wild type synthetase was overexpressed with pA1EL, only cognate substrate **1** supported protein synthesis (lane 2, Fig. 4A). Neither of the canonical, nor the noncanonical analogs (lanes 3-6) supported detectable levels of protein expression.

The background level expression (lane 1) in Figure 4.6A is nearly undetectable. This is expected since in the absence of **1**, none of the other nineteen amino acids is joined to $tRNA^{Leu}$ at a substantial rate. In contrast, when we performed the same experiments with strains that overexpress the T252Y mutant,

significant protein expression was observed, even in media depleted of **1** (not shown). Composition analysis of protein expressed under these conditions shows that natural amino acids **2-4** are incorporated at positions normally occupied by **1**, indicative of the inability of the LeuRS mutant to proofread misaminoacylated $tRNA^{Leu}$ *in vivo*. In order to decrease background expression, we opted to supplement M9 medium with only sixteen natural amino acids (without **1-4**), in addition to the analogs of interest.

Figure 4.6B shows the levels of protein expression supported by the different analogs in the host LAM1000/pA1T252Y/pREP4. Exclusion of amino acids **2-4** from the expression medium diminished the levels of background expression (lane 1). We attribute the residual expression observed in this experiment to the incorporation of **2-4** accumulated via cellular biosynthesis (LAM1000 is only auxotrophic in **1**). Analogs **2-6** all supported substantial levels of protein synthesis in the absence of **1** when the mutant LeuRS was overexpressed. The amounts of protein obtained correlate well with the results of *in vitro* ATP-PP_i exchange experiments. The most robust analogs **5** and **6** supported protein synthesis at almost the same level as that of **1**. Protein expression levels in media supplemented with more sluggish analogs **2-4** were visibly reduced.

We do not fully understand the differences in the electrophoretic mobilities of A1 proteins upon incorporation of the analogs. We have previously observed similar changes in electrophoretic mobility of A1 upon incorporation of trifluoroleucine (*24*). The integrity of each purified protein was confirmed with MALDI-TOF mass spectrometry (data not shown), thus excluding the possibilities of premature translation termination and posttranslational modification. Interestingly, the substituted proteins migrated more slowly with decreased amino acid side-chain hydrophobicity. For example, the average hydrophobicity (a quantity defined to be the required energy to transfer the amino acid from a hydrophobic to an aqueous environment, relative to that of glycine) of **1**, **2**, and **3** are 2.4, 2.9, and 1.4 kcal/residue, respectively (41). Figure 4.6B shows that the relative mobilities of A1 containing these amino acids correlates directly with these values (**2**-A1 > **1**-A1 > **3**-A1). A1 substituted with the more hydrophobic analogs trifluoroleucine or hexafluoroleucine also migrated faster than the wild-type protein (25).

4.3.5. Protein Amino Acid Analysis

Compositions of A1 protein expressed under different conditions of amino acid supplementations were determined by amino acid analysis. The proteins were purified via Ni-NTA columns under denaturing conditions. Table 4.2 lists results for proteins expressed in media containing **1-6**. The expected individual mole fractions of **1-4** are listed and compared to the observed mole fractions of these amino acids. The combined mole fractions of these amino acids and the nonnatural amino acids (in samples containing **5** or **6**) are also compared to the expected total value. Canonical amino acids **2**, **3**, and **4** were inserted at leucine positions using the overexpression mutant strain at substitution rates of 45%, 54%, and 44%, respectively. These numbers are calculated by noting the percentage drop in the amount of **1** detected. Corresponding increases in the mole fractions of the expected analogs are observed, confirming the presence of these natural amino acids at positions in the protein other than those specified by their respective codons. The mole fraction totals of **1-4** in the protein decreased slightly in proteins synthesized in **2**, **3** or **4** supplemented medium. We initially suspected the observed decrease in total mole fractions is due to substitution of **1** with other natural amino acids, such as threonine. However, no evidences of $tRNA^{Leu}$ misaminoacylation by other amino acids were detected using mass spectrometry analysis.

The amino acid analysis results for proteins containing the unnatural amino acids **5** and **6** are also shown in Table 4.2. Both amino acids elute separately during analysis and can be quantified directly. As expected from the intensities of the SDS-PAGE bands, most of the leucine positions are occupied by the unnatural surrogates. Based on the levels of leucine depletion, **5** and **6** are incorporated at 79% and 91%, respectively. Although the observed mole fractions of **2** and **3** in these samples match the expected values, the mole fraction of **4** is significantly lower. The total mole fractions of **1-6** in the protein, however, matched the expected total mole fractions of **1-4** in the wild type protein (errors of 2.3 and 0.9%, respectively). These results suggest that in addition to occupying the leucine positions in A1, **5** and **6** substituted for **4** at ATG codons with rates of 68% and 92%, respectively. This result is not surprising since we have reported previously that both amino acids are efficient analogs of **4** (*30*). Tryptic mass spectrometry provided additional evidences that these two unnatural amino acids replace **1** and **4** in the target protein, as discussed below.

4.3.6. Protein Typtic Mass Spectroscopy

Tryptic digest of A1 yields several peptide fragments that can be easily detected by MALDI-MS as shown in Figure 4.5. For most of these studies, we focused on fragment 2: N-LKNEIEDLKAEIGDLNNTSGIR-C. There are three leucine sites in this fragment; the unsubstituted mass is 2442. The spectra of peptide fragments that contained amino acids **1-6** are shown in Figure 4.7. Peptides containing **2** (**B**) and **6** (**F**), which have the same molecular weight as **1** (**A**), displayed a single peak in MALDI analysis. The anticipated differences in peptide masses can be clearly detected for the other analogs. For example, upon each substitution of **1** by **4** in this fragment, the mass is increased by 18. Four peaks separated by 18 mass units are observed by MS (**D**), each corresponding to 0, 1, 2, and 3 sites of substitution by **4**. Similar patterns are observed for **3** (-14 mass difference, **C**) and **5** (-14, **E**). Figure 4.8 shows the tryptic MS pattern for peptide fragment 1 containing **1**, **5**, and **6**. This peptide contains two leucines and one methionine and has a wild-type mass of 2772. The peak with the exact mass was observed in the wild type fragment (**A**). When **5** (**B**) or **6** (**C**) is supplemented during protein expression, mass shift of this fragment indicates that these analogs replaced both **1** and **4** in the target protein, agreeing with amino acid analysis results.

4.3.7. Incorporation of Other Analogs

The incorporation of **5** and **6** expands the repertoire of aliphatic amino acids available for protein biosynthesis. However, the inert hydrocarbon side chains of these analogs are inaccessible to chemoselective protein modifications. We investigated the abilities of analogs **7-12** to replace **1** during protein synthesis. Each of these analogs contains an unsaturated moiety, which can be activated through ruthenium-mediated metathesis (42, 43) or palladium-catalyzed coupling reactions. Analogs **7-9** are isosteric to the three-carbon side chain analog **5**; analogs **10-12** are isosteric to the four-carbon side chain analog **6**. None of the analogs supported A1 synthesis when the wild type LeuRS is overexpressed (not shown). Figure 4.6C shows the expression results when the T252Y mutant was overexpressed. All of the analogs except **9** and **10** (lanes 4 and 5) supported protein expression at levels higher than background (**9** is not incorporated despite the higher intensity of the protein band. Composition analysis shows that **4** is incorporated in place of **1** in media supplemented with **9**, see below).

Proteins expressed in media supplemented with analogs **7-12** were analyzed with tryptic mass spectrometry as before. The MS spectra of tryptic fragment 2 are shown in Figure 4.9. Incorporated analogs in these fragments can be clearly identified: **7** (-16 mass difference per substitution, **A**), **10** (-2, **D**), **11** (-4, **E**) and **12** (-4, **F**). Instead of observing a decrease of 18 mass units upon each substitution of **1** with **8**, we detected species with incremental *increases* in mass of 18 units, a pattern similar to that observed with **4**. Propargylglycine is known to inhibit γ-cystathionase, which is involved in the conversion of homocysteine to cysteine (44, 45). Upon blocking this biosynthetic pathway, homocysteine is shuttled through the methionine biosynthetic pathway, leading to the intracellular accumulation of **4**. Hence it is likely that while **8** is not directly incorporated into proteins under our experimental conditions, the indirect accumulation of **4** supported the apparent protein synthesis in Figure 4.6C (lane 4). We also examined the composition of the trace amounts of protein recovered from cells treated with **9** (**C**). A single peak corresponding to the wild type mass was observed, instead of peaks separated by 16 mass units as a result of **9** incorporation, confirming the previous observation that **9** is not a surrogate of LeuRS. Analogs **7, 10-12** were all previously shown to be methionine analogs (30). As a result, tryptic fragment 1 containing **7**, **10-12** all show mass distributions characteristic of analog incorporation at both **1** and **4** sites, similar to that observed for fragments containing **5** and **6** (Figure 4.8).

Assuming equal probability of incorporating an analog at all leucine positions in all copies of A1, the relative intensities of the tryptic fragments in Figures 4.7 and 4.9 can be described by the terms of the polynomial $((1-p) + p)^n$, where p is the fraction of leucine sites occupied by an analog and n is the number of leucine sites in the fragment.

We assessed the validity of this approach by comparing the values of p obtained from a least squares fit of the spectra in Figure 4.7 with those determined by amino acid analysis. For fragment 2, the intensities of the peaks corresponding to 0, 1, 2 and 3 sites of substitution are $(1-p)^3$, $3(1-p)^2p$, $3(1-p)p^2$ and p^3, respectively. For samples containing **3**, **4** and **5**, the values of p obtained from fitting the spectra are 41, 66 and 82%, respectively, as compared to 44, 54 and 79% obtained from amino acid analysis. One should not expect better agreement; the MALDI method is not rigorously quantitative, assumption of a single p may not be valid, and the cellular amino acid pools change over the course of the experiment.

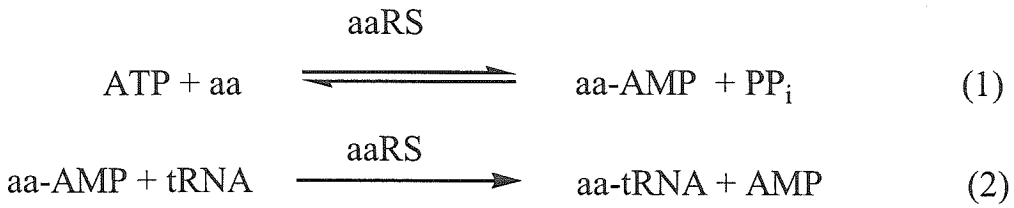
Nevertheless, the MALDI spectra provide a rough estimate of level of substitution. The values obtained in this way were 76, 37, 74 and 37%, for analogs 7, 10, 11 and 12, respectively.

We also examined the isoleucine analogs 5,5,5-trifluoroisoleucine, 2-amino-3-methyl-5-pentenoic acid and 2-amino-3-methyl-5-pentynoic acid. None of these analogs supported measurable levels of protein synthesis. For these substrates, the observed lack of translational activity is likely to be due to slow activation by the synthetic site. Similar results were obtained for the valine derivatives 4,4,4-trifluorovaline and αbu. The five carbon side-chain analog 2-aminoheptanoic acid, also displayed no translational activity as an analog of **1**. We attribute the lack of incorporation to weak binding at the synthetic active site of LeuRS. The back wall of the leucine binding pocket is formed in part by a histidine residue (position 537 in *E. coli* LeuRS; position 545 in *T. thermophilus* LeuRS), which limits the maximum tolerable number of carbon atoms on the side chain to four. Amino acids with longer side chains must adopt entropically unfavorable conformations to be accommodated in the active site. Mutation of this histidine residue to a smaller amino acid to create an enlarged and more flexible pocket, along with the editing domain mutation T252Y, might allow the incorporation of amino acids with longer side chains.

4.4 Conclusions

Our motivation for this study is to identify a suitable host strain for introducing the noncanonical amino acids shown in Scheme 2 *in vivo*. Our research laboratory and others (46-50) have established a variety of techniques for introducing unnatural functionalities, including overexpressing the wild type synthetase to elevate host aminoacylation activity (*25, 26*), introducing active site mutations (*28, 29*) to allow more efficient amino acid activation, etc. Our results reported here, along with the previous work on ValRS editing mechanism (*17*), show that rational manipulation of residues that modulate editing cavity geometry is a new and powerful tool for the incorporation of novel amino acids *in vivo*. Amino acids such as **7**, **10-12** that contain unsaturated functionalities, once incorporated in a protein, are attractive orthogonal sites for performing protein-protein, protein-polymer conjugations. We are currently investigating these chemoselective protein derivatization methods using recombinant proteins produced in the host described here.

Scheme 1



Scheme 2

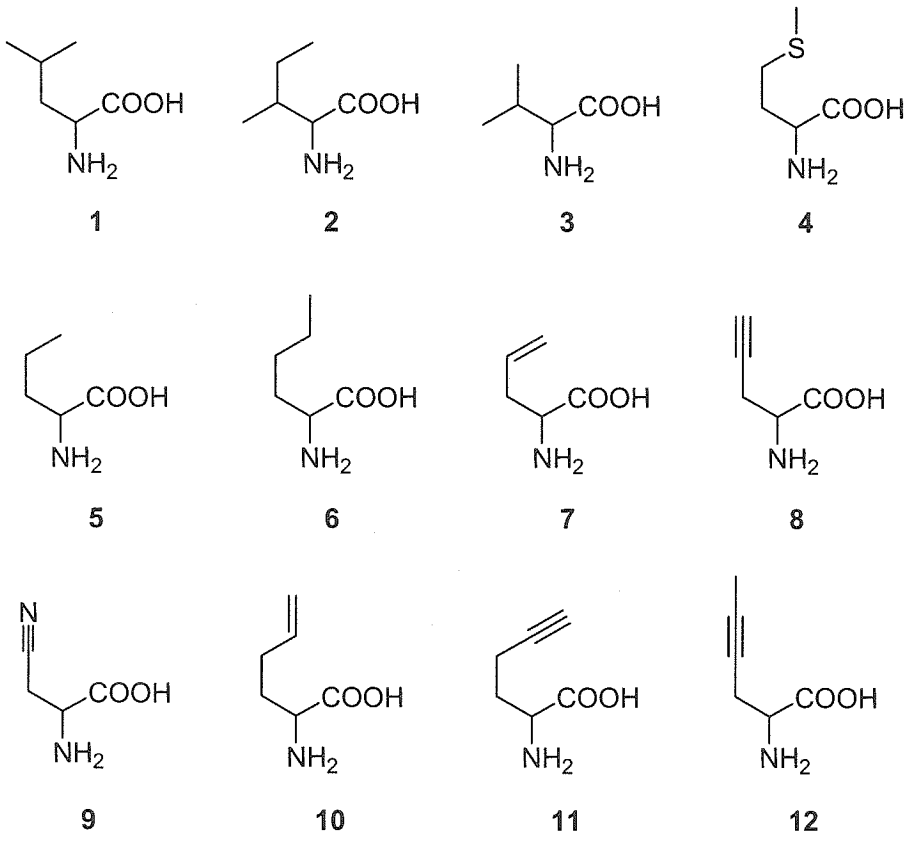


Table 4.1. ATP-PP_i exchange kinetic parameters of wild-type 6xHis-LeuRS toward canonical (**1-4**) and non-canonical (**5, 6**) amino acids.

Amino Acid	K_m (μM)[a]	k_{cat} (s^{-1})	k_{cat}/K_m (rel)
1[b]	18±1.7	2.2±0.02	1
2	2568±555	0.06±0.002	1/5231
3	2356±320	0.03±0.0008	1/9599
4	2178±554	0.08±0.003	1/3327
5	1155±368	1.24±0.06	1/114
6	2516±659	0.22±0.02	1/1397

[a] Concentrations are quoted for L-amino acids. [b]The kinetic parameters of the T252Y mutant were determined to be essentially the same as those for the wild type enzyme for **1** (K_m = 20±2.3 μM, k_{cat} = 1.92±0.03 s^{-1}).

Table 4.2. Results of amino acid analyses of target proteins expressed in media supplemented with **1**, **2**, **3**, **4**, **5** or **6**. The values reported are mole fractions of amino acids.

		Amino Acid Supplement[a]											
Amino acid in protein		**1**		**2**		**3**		**4**		**5**		**6**	
	Exp[b]	Obs[c]	Δ%[d]	Obs	Δ%	Obs	Δ%	Obs	Δ%	Obs	Δ%	Obs	Δ%
1	10.8	**10.9**	**+0.9**	5.9	-45	6.0	-44	5.0	-54	2.3	-79	0.97	-91
2	4.1	3.8	-7.3	**7.2**	**+76**	3.8	-7.3	4.2	+2.4	4.1	0	3.8	-7.3
3	4.1	4.2	+2.4	4.5	+9.7	**6.7**	**+63**	4.7	+14	**4.3**	+4.9	4.1	0
4	2.7	2.4	-11	2.5	+7.4	2.6	-3.7	**6.3**	**+133**	0.87	-68	0.22	-92
5	0	ND[e]	-	ND	-	ND	-	ND	-	**10.6**	-	ND	-
6	0	ND	-	ND	-	ND	-	ND	-	ND	-	**12.4**	-
Total[f]	21.6	21.3	-1.4	20.1	-6.9	20.2	-5.2	20.2	-5.2	22.1	+2.3	21.4	-0.9

[a]**1** is supplemented at 40 mg/L; other amino acids are supplemented at 320 mg/L. [b]Expected mole fractions based on composition of **1-4** in A1. [c]Observed mole fractions. [d]Percentage deviation between observed mole fractions to that of expected. [e]Not detected. [f]Sum of mole fractions corresponding to **1-6**.

A. aeolicus	263	Y	I	D	V	F	T	T	R	P	D	T	V	F	G	A	T	F	V	V	L	A	P
Synechocystis sp.	244	-	I	A	V	F	T	T	R	P	D	T	V	Y	G	V	T	Y	V	V	L	A	P
E. coli	242	T	L	T	V	Y	T	T	R	P	D	T	F	M	G	C	T	Y	L	A	V	A	A
B. subtilis	240	S	F	T	V	F	T	T	R	P	D	T	L	F	G	A	T	Y	T	V	L	A	P
B. burgdorferi	248	K	I	K	V	F	T	T	R	P	D	T	I	F	G	I	T	Y	L	V	I	A	P
S. cerevisiae	309	Y	F	V	A	A	T	L	R	P	E	T	M	Y	G	Q	T	C	C	F	V	S	P
N. crassa	328	Y	L	C	P	A	T	L	R	P	E	T	M	Y	G	Q	V	C	C	F	V	G	P
C. elegans	284	Y	L	V	A	A	T	L	R	P	E	T	M	Y	G	Q	T	N	C	Y	L	H	P

Figure 4.1. Alignment of LeuRS sequences flanking the T252 position from eight difference organisms. The shaded threonine residue is absolutely conserved across species. Sequences immediately adjacent to T252 are also highly conserved. Sequence source: http://www.expasy.ch.

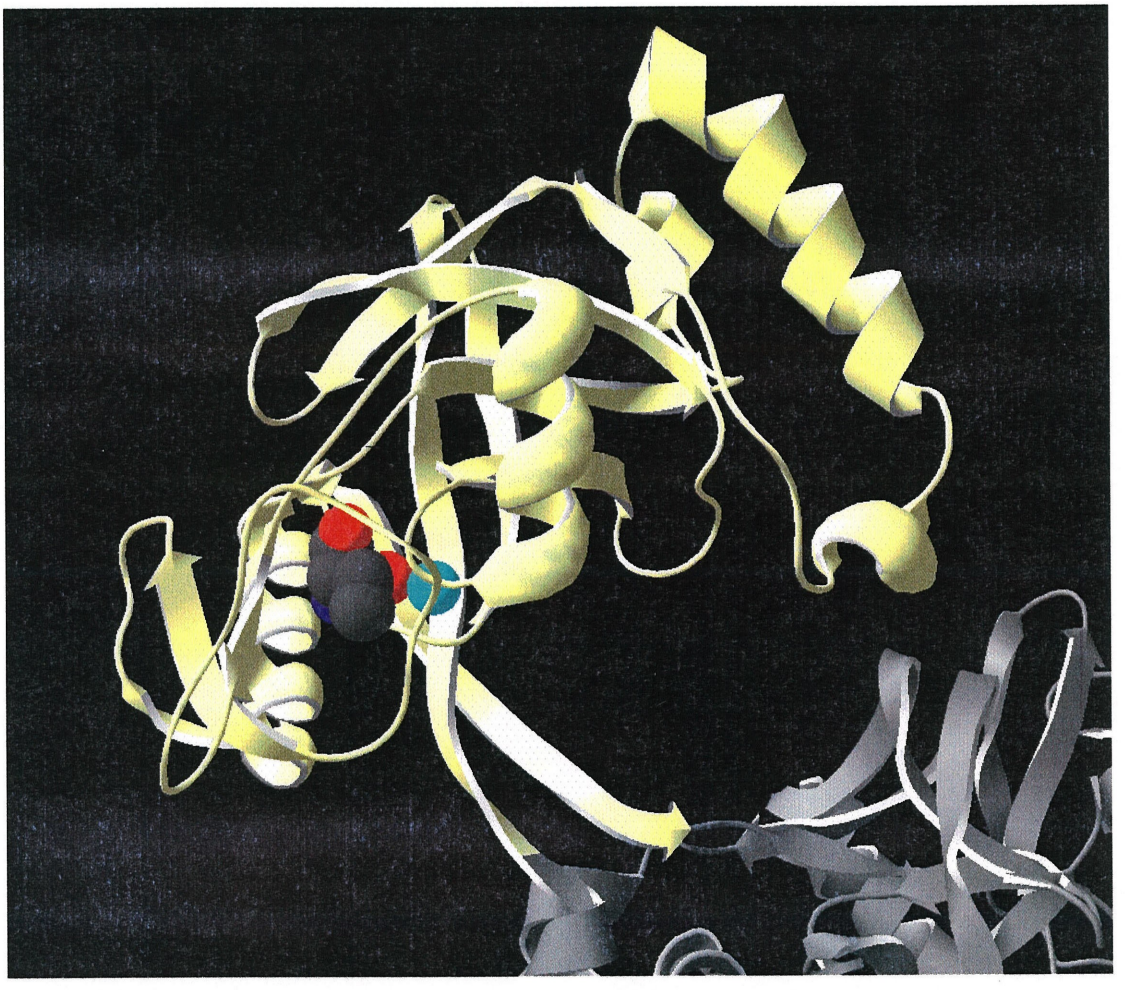


Figure 4.2 Crystal structure of the *thermophilus* LeuRS CP1 domain. The CP1 domain is shown in yellow. Rest of the LeuRS is visible at the bottom-right corner in grey. The conserved T252 residue is shown in space filled model. The oxygen on its side chain is shown in red. The water molecule that hydrogen bonds to this oxygen is shown in cyan. In this view, the T252 side chain is oriented towards the synthetic active site (grey portion).

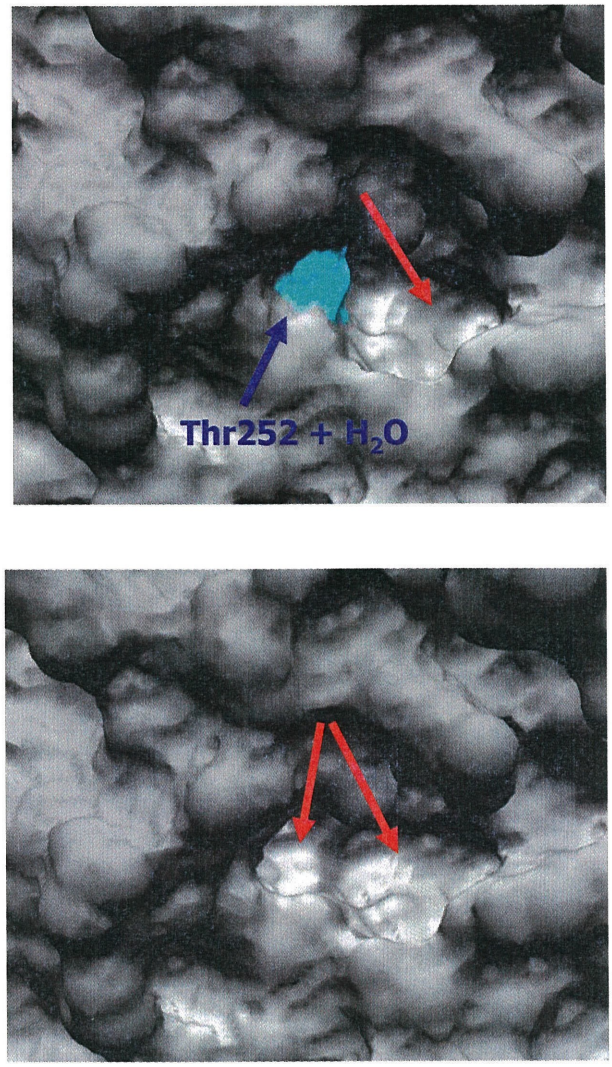


Figure 4.3. Close-up view of the *thermophilus* LeuRS putative editing pocket. Molecular surfaces are generated using Swiss-Pdbviewer. Top panel: wild-type LeuRS cavity. The water molecule that hydrogen bonds to T252 is shown in blue. T252 is not visible in this representation. The binding groove adjacent to the water molecule (red arrow) is where straight chain analogs such as **7** could bind. Bottom panel: T252A LeuRS cavity. This representation is generated by deleting the coordinates of T252 γ carbon and γ oxygen and the water molecule. This mutation creates a second binding groove (additional red arrow) where the γ methyl substituent of **1** can be accommodated. Mursinna et al. have shown this mutant can hydrolyze its cognate substrate **1** rapidly from Leu-$tRNA^{Leu}$.

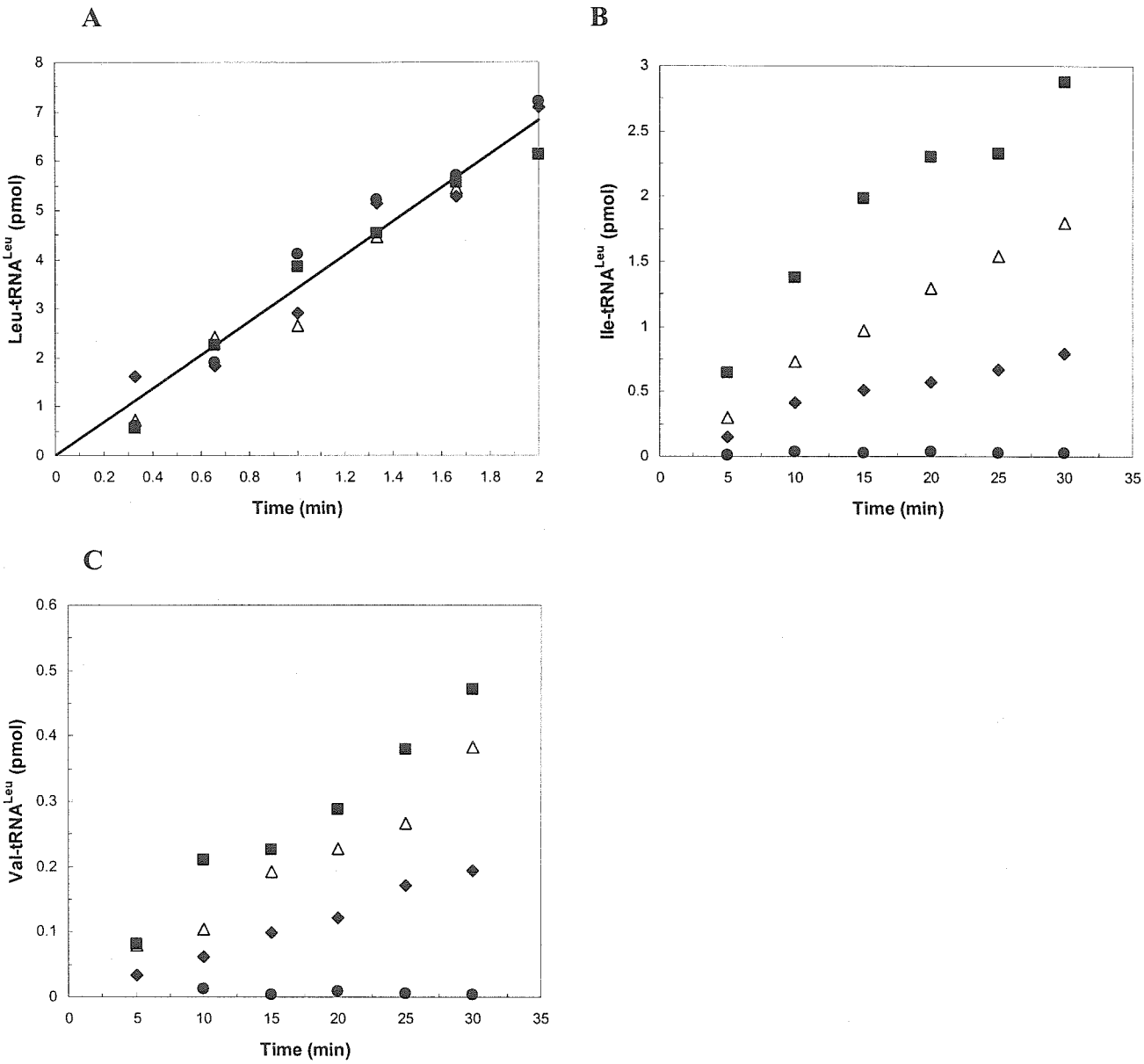


Figure 4.4. Aminoacylation of tRNA^{Leu} by wild-type and mutant LeuRS enzymes at 37°C. Each enzyme was evaluated for its ability to aminoacylate tRNA^{Leu} with **1** (**A**), **2** (**B**) and **3** (**C**). (●): wild-type enzyme; (■): T252Y; (△): T252F; (◆): T252L. All enzymes displayed comparable rate of leucine aminoacylation (10 nM enzyme, 20 μM **1**), demonstrating point mutations in the editing cavity does not affect the kinetics of this reaction. While the wild type does not misaminoacylate **2** or **3**, all three mutant enzymes show different levels of aminoacylation under these conditions (500 nM enzyme, 20 μM analog). The rate of misacylation increases as larger amino acids are chosen to occupy the T252 position. The mutant T252Y is chosen for the *in vivo* incorporation studies.

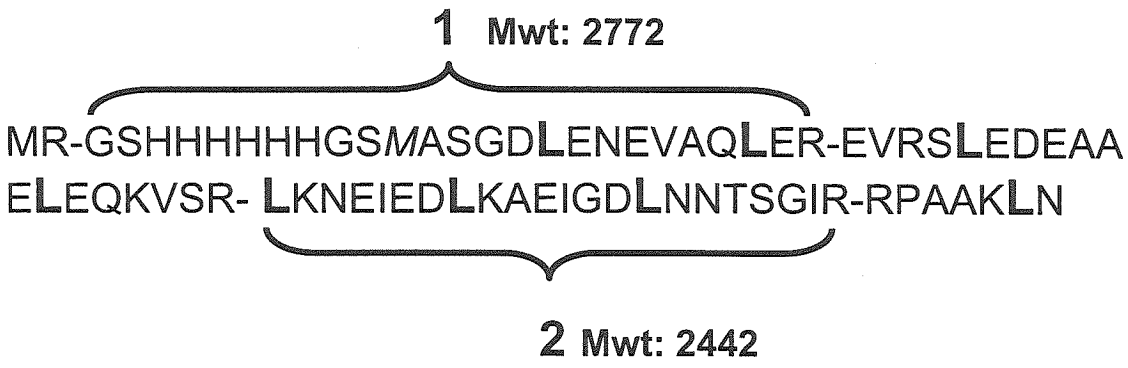
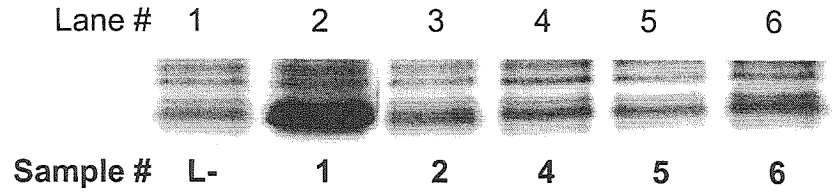
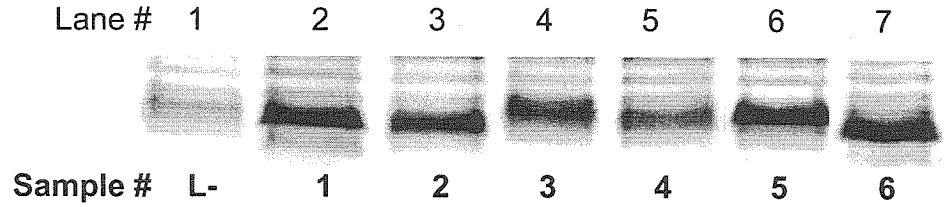


Figure 4.5. Target protein A1 amino acid sequence. The protein contains 74 residues and eight leucines. Two tryptic fragments are detectable by MALDI-MS and are labeled as shown. Fragment 1 contains 2 leucines and one methionine and has an expected mass of 2772. Fragment 2 contains three leucines and has an expected mass of 2442.

A.



B



C

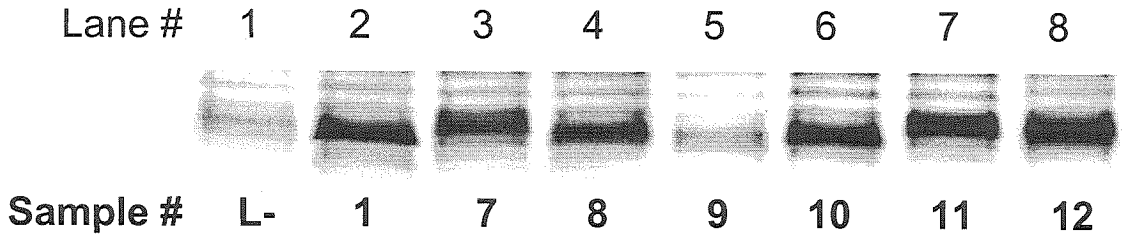


Figure 4.6. *In vivo* incorporation of unnatural amino acids by the wild type and mutant LeuRS synthetase. Proteins were visualized by commassie staining of SDS-PAGE gel. The amino acids of interest were supplemented at 320 mg/L of the L isomer. Lane numbers are indicated at the top of the gel and analog numbers are shown below the gel. L⁻: induction without any of the amino acids in scheme 1. **A**: Protein expression profiles for the wild type overexpression strain LAM1000/pA1EL, none of the supplied amino acids supported protein synthesis except **1** (lane 2); **B**: Protein expression profiles for the LAM1000/pA1T252Y overexpression strain. Previously translationally inactive analogs **2-6** (lanes 3-7) show bands corresponding to target protein accumulation. **C:** Protein expression profiles for unsaturated analogs **7-12.** Analogs **7, 10-12** support significant levels of protein synthesis in the absence of **1.** Culture supplemented by **8** (lane 4) shows protein expression and is verified to be incorporation of **4** instead (see text). Analog **9** (lane 5) does not support detectable levels of protein expression under these conditions.

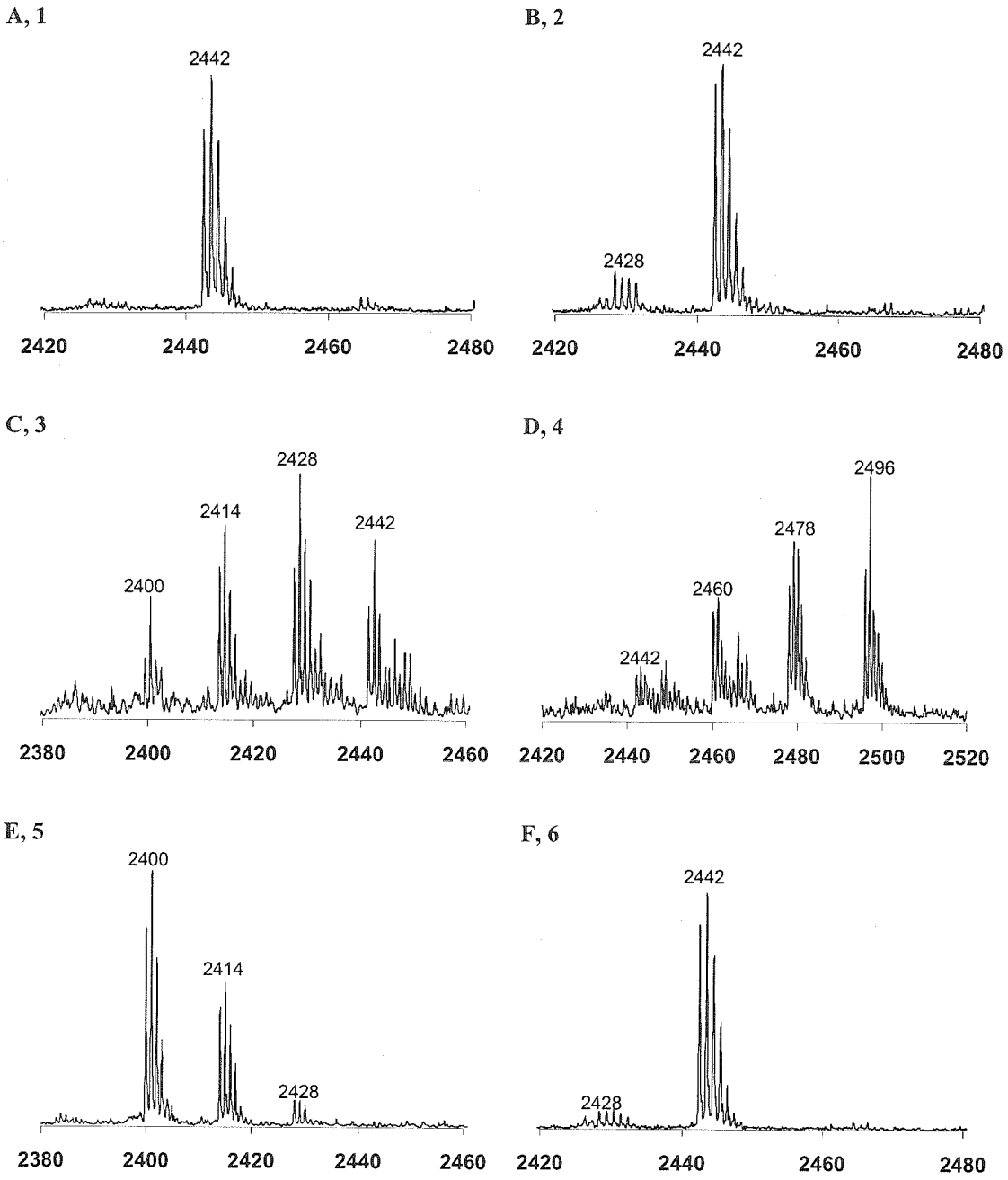
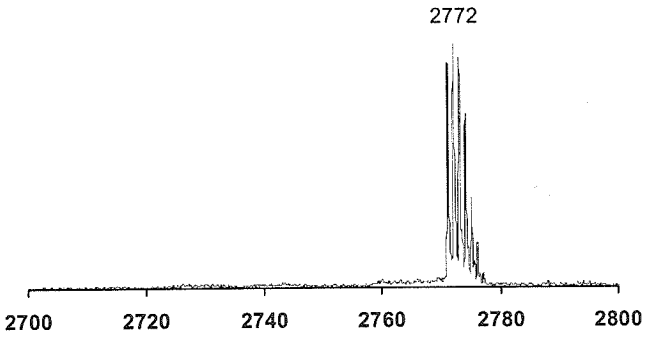
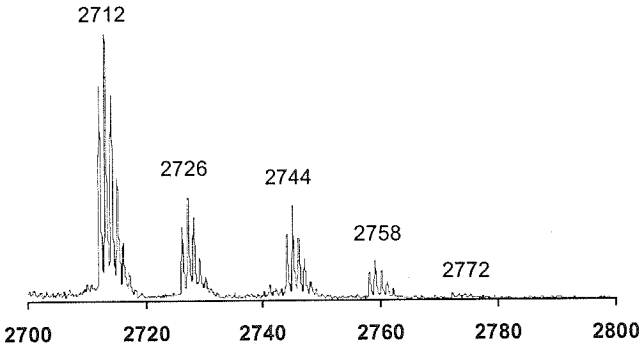


Figure 4.7. MALDI-MS of tryptic peptide fragment 2 containing (A): **1**; (B): **2**; (C): **3**; (D): **4**; (E): **5**; (F): **6**; The peptide has the sequence LKNEIEDLKAEIGDLNNTSGIR. The expected mass of the fragment is 2442. Three sites of substitutions are possible.

A.



B.



C.

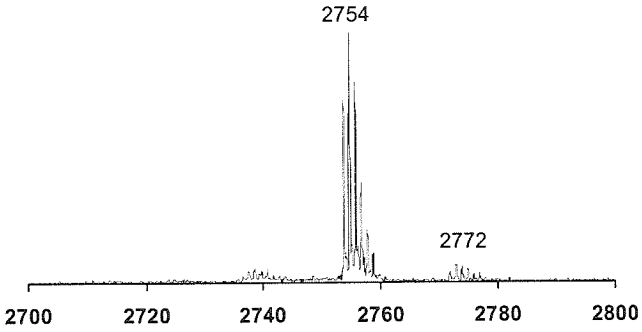


Figure 4.8. MALDI-MS of tryptic peptide fragment 1 containing (**A**): **1**; (**B**): **5** and (**C**): **6**. The sequence of this fragment is GSHHHHHHGS*M*ASGDLENEVAQLER and the wild type mass is 2772 (**A**). Two sites of **1** substitution and one site of **4** substitution are present. For **5** and **6**, the decreases in mass observed with this fragment is due to the substitution of the lone **4** position with these amino acids, in addition to occupying the positions corresponding to **1**. The patterns observed agree well with amino acid analysis results (Table 4.2).

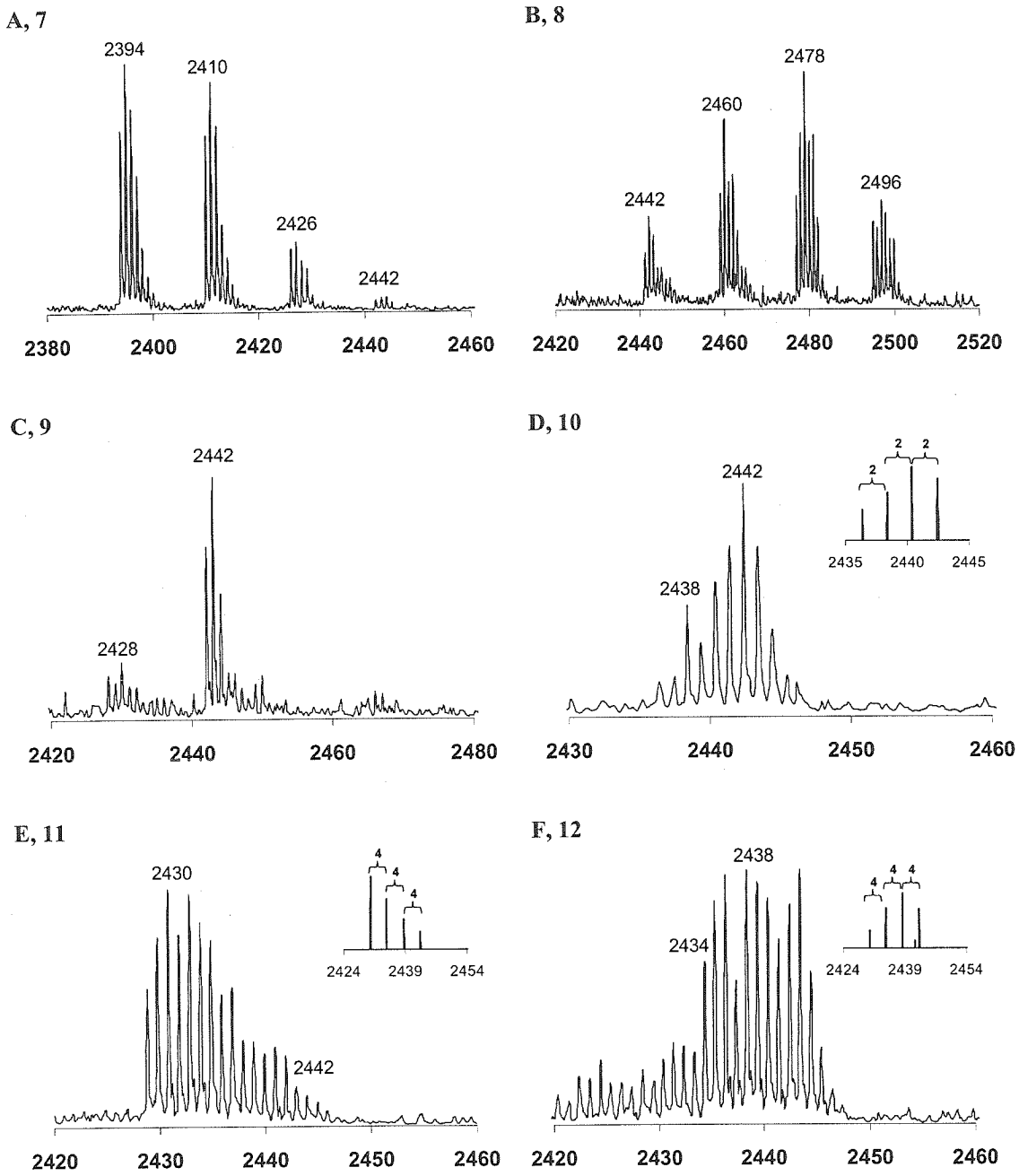


Figure 4.9. MALDI-MS of tryptic peptide fragment 2 containing unsaturated amino acid analogs of **1**, (**A**): **7**; (**B**): **8**; (**C**): **9**; (**D**): **10**; (**E**): **11**; (**F**): **12**; Analog **8** is not incorporated. The unexpected increase of 18 mass units in 7B is attributed to the incorporation of **4** (see text). Analog **9** is not incorporated, as seen from SDS-PAGE and the MS (7C). The insets in D-F are deisotope filtered for clarity.

4.5 References

1. Schimmel, P. (1987) *Annu Rev Biochem 56*, 125-58.
2. Carter, C. W., Jr. (1993) *Annu Rev Biochem 62*, 715-48.
3. Ibba, M., and Soll, D. (2000) *Annu Rev Biochem 69*, 617-50.
4. Fersht, A. R. (1977) *Biochemistry 16*, 1025-30.
5. Fersht, A. R., and Dingwall, C. (1979) *Biochemistry 18*, 2627-31.
6. Baldwin, A. N., and Berg, P. (1966) *J Biol Chem 241*, 839-45.
7. Fersht, A. R., and Dingwall, C. (1979) *Biochemistry 18*, 1238-45.
8. Loftfield, R. B., and Vanderjagt, D. (1972) *Biochem J 128*, 1353-6.
9. Nureki, O., Vassylyev, D. G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T. L., Schimmel, P., and Yokoyama, S. (1998) *Science 280*, 578-82.
10. Silvian, L. F., Wang, J., and Steitz, T. A. (1999) *Science 285*, 1074-7.
11. Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D. G., and Yokoyama, S. (2000) *Cell 103*, 793-803.
12. Cusack, S., Yaremchuk, A., and Tukalo, M. (2000) *Embo J 19*, 2351-61.
13. Starzyk, R. M., Webster, T. A., and Schimmel, P. (1987) *Science 237*, 1614-8.
14. Eldred, E. W., and Schimmel, P. R. (1972) *J Biol Chem 247*, 2961-4.
15. Bishop, A. C., Nomanbhoy, T. K., and Schimmel, P. (2002) *Proc Natl Acad Sci U S A 99*, 585-90.
16. Hendrickson, T. L., Nomanbhoy, T. K., and Schimmel, P. (2000) *Biochemistry 39*, 8180-6.
17. Doring, V., Mootz, H. D., Nangle, L. A., Hendrickson, T. L., de Crecy-Lagard, V., Schimmel, P., and Marliere, P. (2001) *Science 292*, 501-4.
18. Hartlein, M., and Madern, D. (1987) *Nucleic Acids Res 15*, 10199-210.
19. Englisch, S., Englisch, U., von der Haar, F., and Cramer, F. (1986) *Nucleic Acids Res 14*, 7529-39.
20. Chen, J. F., Li, T., Wang, E. D., and Wang, Y. L. (2001) *Biochemistry 40*, 1144-9.
21. Chen, J. F., Guo, N. N., Li, T., Wang, E. D., and Wang, Y. L. (2000) *Biochemistry 39*, 6726-31.
22. Mursinna, R. S., Lincecum, T. L., Jr., and Martinis, S. A. (2001) *Biochemistry 40*, 5376-81.
23. Furter, R. (1998) *Protein Sci 7*, 419-26.

24. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F., and Tirrell, D. A. (2001) *Angew Chem Int Ed Engl 40*, 1494-1496.

25. Tang, Y., and Tirrell, D. A. (2001) *J Am Chem Soc 123*, 11089-90.

26. Kiick, K. L., van Hest, J. C., and Tirrell, D. A. (2000) *Angew Chem Int Ed Engl 39*, 2148-2152.

27. van Hest, J. C., and Tirrell, D. A. (1998) *FEBS Lett 428*, 68-70.

28. Kirshenbaum, K., Carrico, I. S., and Tirrell, D. A. (2002) *Chembiochem 3*, 235-7.

29. Datta, D., Wang, P., Carrico, I. S., Mayo, S., and Tirrell, D. A. (2002) *J Am Chem Soc in press*.

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

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

32. van Hest, J. C., Kiick, K. L., and Tirrell, D. A. (2000) *J Am Chem Soc 122*, 1282-1288.

33. Bilgicer, B., Fichera, A., and Kumar, K. (2001) *J Am Chem Soc 123*, 4393-9.

34. Sampson, J. R., and Uhlenbeck, O. C. (1988) *Proc Natl Acad Sci U S A 85*, 1033-7.

35. Tocchini-Valentini, G., Saks, M. E., and Abelson, J. (2000) *J Mol Biol 298*, 779-93.

36. Petka, W. A., Harden, J. L., McGrath, K. P., Wirtz, D., and Tirrell, D. A. (1998) *Science 281*, 389-92.

37. Steipe, B. (2000). http://www.lmb.uni-muenchen.de.users/steipe/boris.html

38. Li, D., Wang, E. D., and Wang, Y. L. (1997) *Acta Biochimica et Biophysica Sinica 29*, 591-596.

39. Jakubowski, H., and Goldman, E. (1992) *Microbiol Rev 56*, 412-29.

40. Apostol, I., Levine, J., Lippincott, J., Leach, J., Hess, E., Glascock, C. B., Weickert, M. J., and Blackmore, R. (1997) *J Biol Chem 272*, 28980-8.

41. Bigelow, C. C. (1967) *J Theor Biol 16*, 187-211.

42. Miller, S. J., Blackwell, H. E., and Grubbs, R. H. (1996) *J Am Chem Soc 118*, 9606-9614.

43. Clark, T. D., Kobayashi, K., and Ghadiri, M. R. (1999) *Chem. Eur. J. 5*.

44. Mosharov, E., Cranford, M. R., and Banerjee, R. (2000) *Biochemistry 39*, 13005-11.

45. Abeles, R. H., and Walsh, C. T. (1973) *J Am Chem Soc 95*, 6124-5.

46. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) *Science 292*, 498-500.

47. Behrens, C., Nielsen, J. N., Fan, X. J., Doisy, X., Kim, K. H., Praetorius-Ibba, M., Nielsen, P. E., and Ibba, M. (2000) *Tetrahedron 56*, 9443-9449.

48. Jakubowski, H. (2000) *J Biol Chem 275*, 21813-6.

49. Kowal, A. K., Kohrer, C., and RajBhandary, U. L. (2001) *Proc Natl Acad Sci U S A 98*, 2268-73.

50. Saks, M. E., Sampson, J. R., Nowak, M. W., Kearney, P. C., Du, F., Abelson, J. N., Lester, H. A., and Dougherty, D. A. (1996) *J Biol Chem 271*, 23169-75.

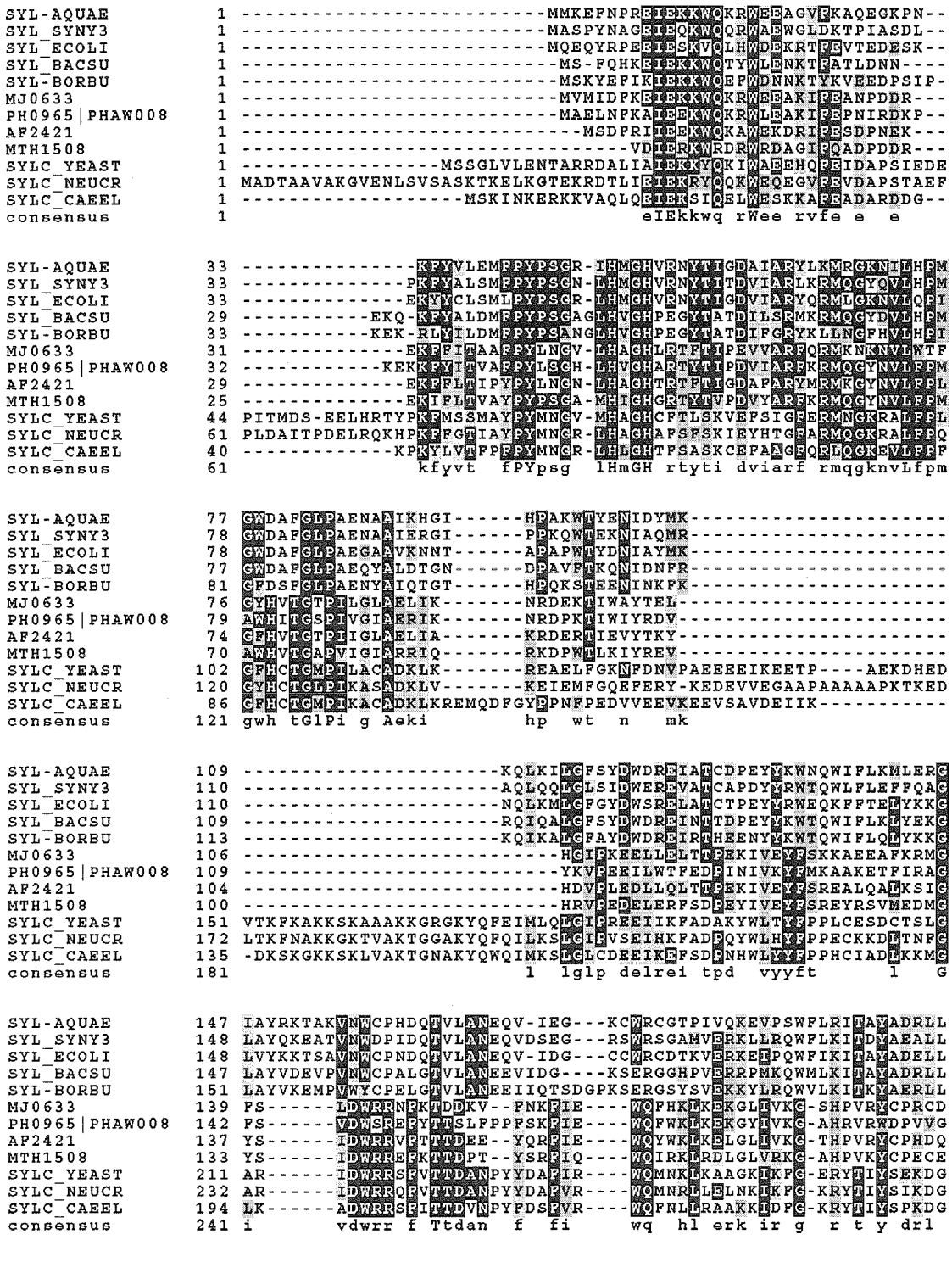
APPENDIX A

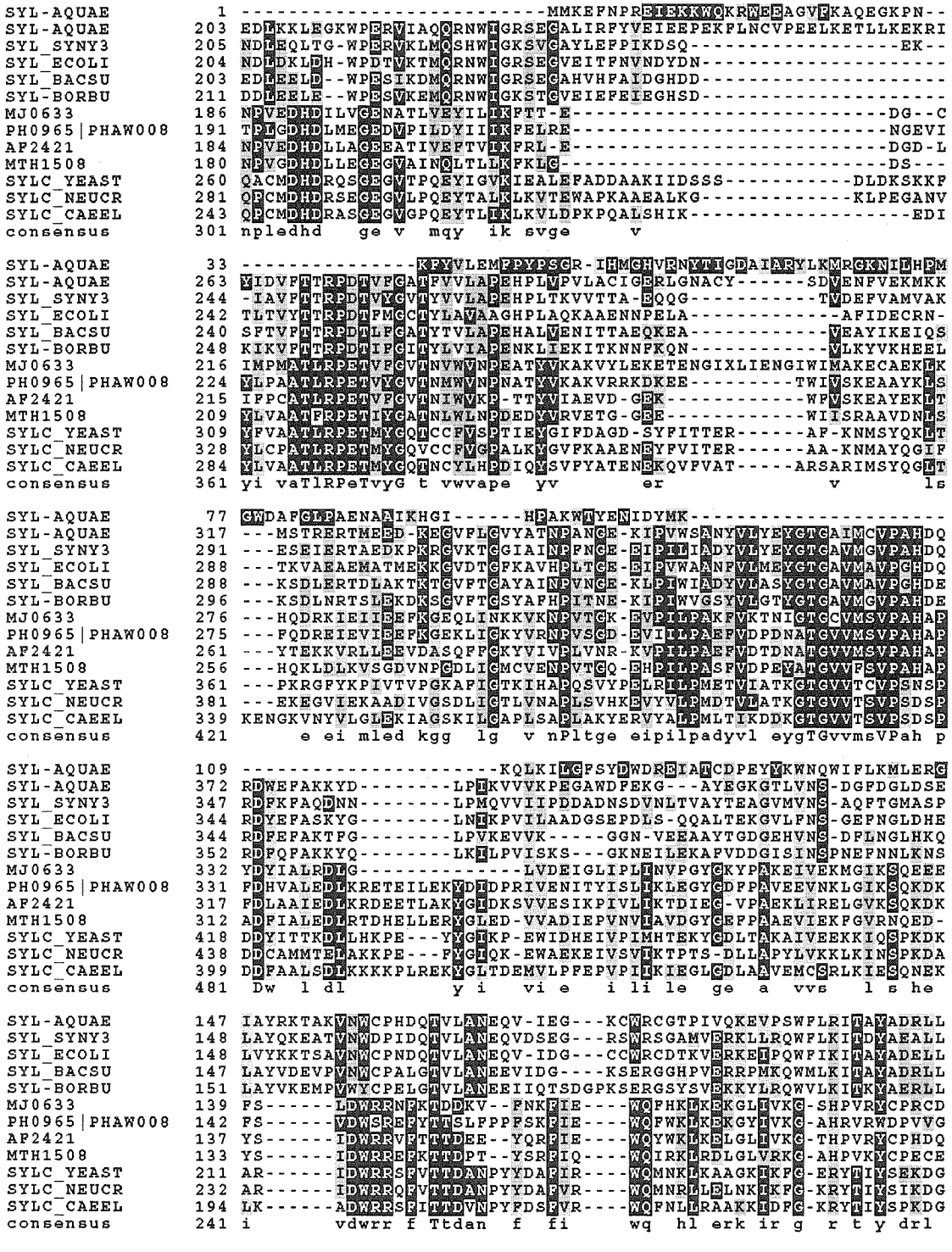
Alignment of Leucyl-tRNA Synthetase Sequences

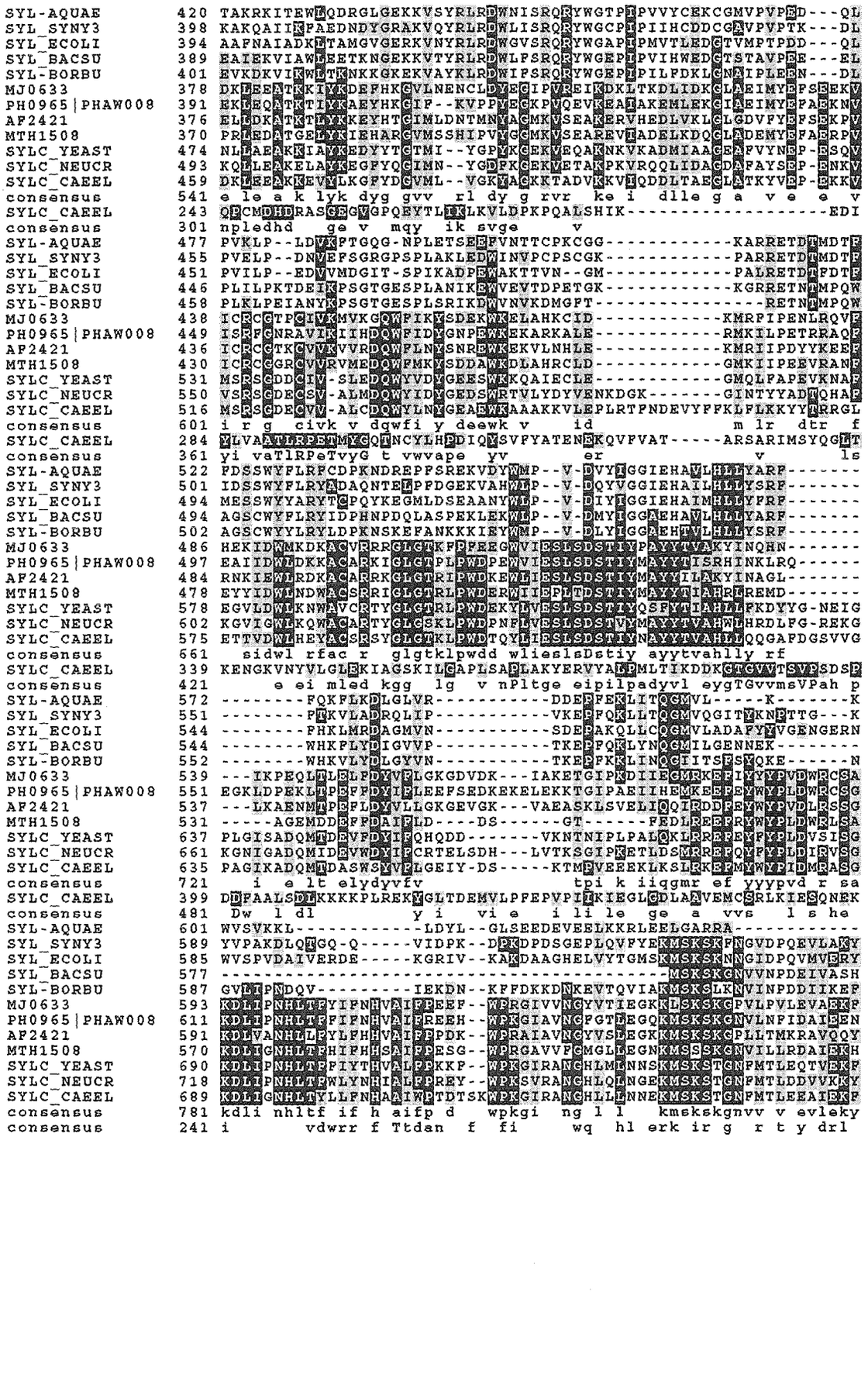
Source: http://www.expasy.ch

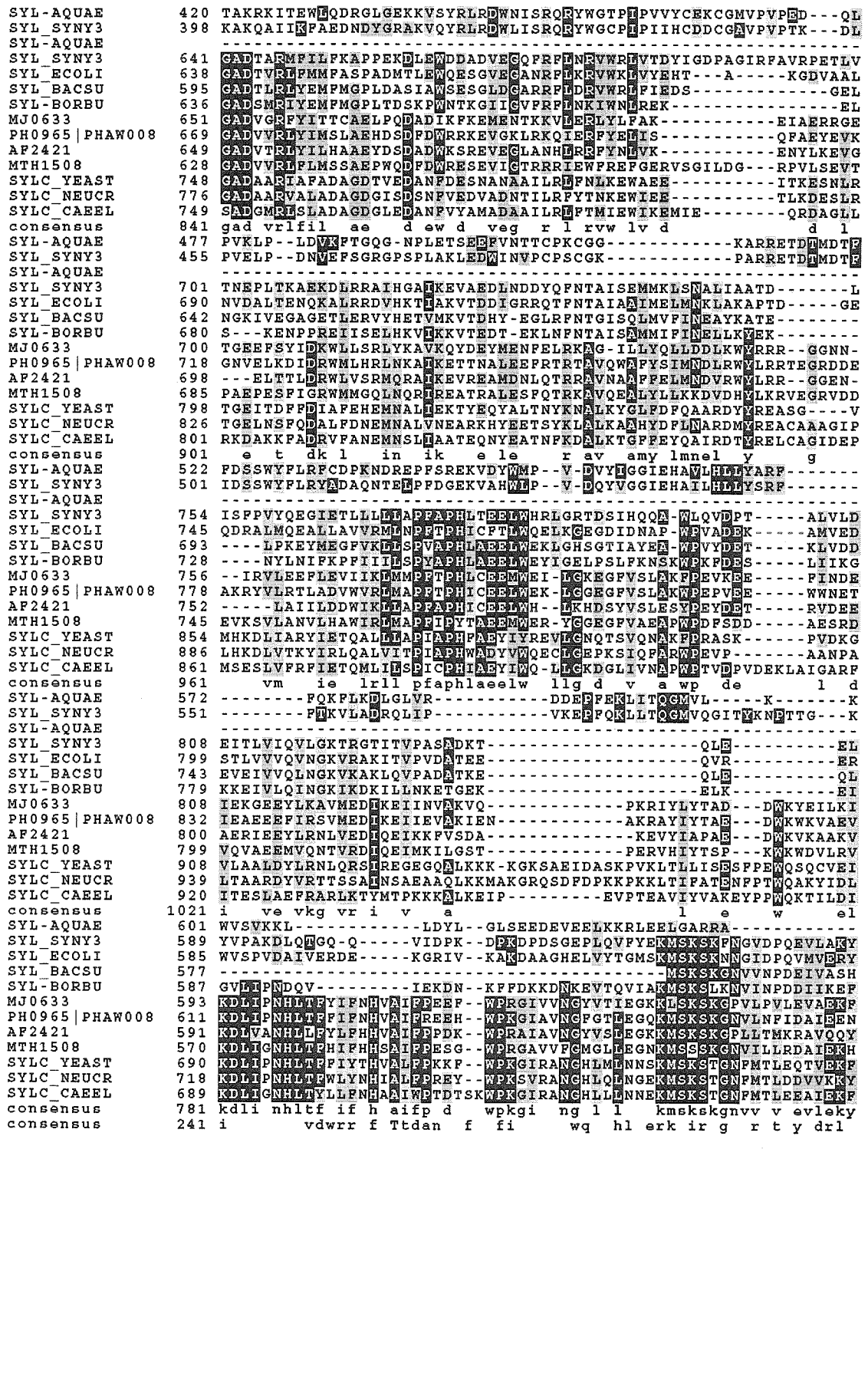
Organism key:

SYL_AQUAE: *Aquifex aeolicus*
SYL_SYNY3: *Synechocystis sp.* (strain PCC 6803)
SYL_ECOLI: *Escherichia coli*
SYL_BACSU: *Bacillus subtilis*
SYL_BORBU: *Borrelia burgdorferi*
MJ0633: *Methanococcus jannaschii*
PH0965|PHAW0008: *Pyrococcus korikoshii*
AF2421: *Archaeoglobus fulgidus*
MTH1508: *Methanobacterium thermoautotrophicum*
SYLC_YEAST: *Saccharomyces cerevisiae*
SYLC_NEUCR: *Neurospora crassa*
SYLC_CAEEL: *Caenorhabditis elegans*


SYL-AQUAE
SYL_SYNY3
SYL_ECOLI
SYL_BACSU
SYL-BORBU
MJ0633
PH0965|PHAW008
AF2421
MTH1508
SYLC_YEAST
SYLC_NEUCR
SYLC_CAEEL
consensus


SYL-AQUAE
SYL-AQUAE
SYL_SYNY3
SYL_ECOLI
SYL_BACSU
SYL-BORBU
MJ0633
PH0965|PHAW008
AP2421
MTH1508
SYLC_YEAST
SYLC_NEUCR
SYLC_CAEEL
consensus




SYL-AQUAE 420 TAKRKITEWLQDRGLGEKKVSYRLRDWNISRQRYWGTPIPVVYCEKCGMVPVPED---QL
SYL_SYNY3 398 KAKQAIIKFAEDNDYGRAKVQYRLRDWLISRQRYWGCPIPIIHCDDCGAVPVPTK---DL
SYL-AQUAE
SYL_SYNY3 641 GADTARMFILFKAPPEKDLEWDDADVEGQFRFLNRVWRLVTDYIGDPAGIRFAVRPETLV
SYL_ECOLI 638 GADTVRLFMMFASPADMTLEWQESGVEGANRPLKRVWKLVYEHT---A-----KGDVAAL
SYL_BACSU 595 GADTLRLYEMFMGPLDASIAWSESGLDGARRFLDRVWRLFIEDS-----------GEL
SYL-BORBU 636 GADSMRIYEMFMGPLTDSKPWNTKGIIGVFRFLNKIWNLREK---------------EL
MJ0633 651 GADVGRFYITTCAELPQDADIKFKEMENTKKVLERLYLFAK----------EIAERRGE
PH0965|PHAW008 669 GADVVRLYIMSLAEHDSDFDWRRKEVGKLRKQIERFYELIS----------QFAEYEVK
AF2421 649 GADVTRLYILHAAEYDSDADWKSREVEGLANHLRRFYNLVK----------ENYLKEVG
MTH1508 628 GADVVRLFLMSSAEPWQDFDWRESEVIGTRRRIEWFREFGERVSGILDG---RPVLSEVT
SYLC_YEAST 748 GADAARIAFADAGDTVEDANFDESNANAAILRLFNLKEWAEE----------ITKESNLR
SYLC_NEUCR 776 GADAARVALADAGDGISDSNFVEDVADNTILRFYTNKEWIEE----------TLKDESLR
SYLC_CAEEL 749 SADGMRLSLADAGDGLEDANFVYAMADAAILRLFTMIEWIKEMIE--------QRDAGLL
consensus 841 gad vrlfil ae d ew d veg r l rvw lv d d l
SYL-AQUAE 477 PVKLP--LDVKFTGQG-NPLETSEEFVNTTCPKCGG----------KARRETDTMDTF
SYL_SYNY3 455 PVELP--DNVEFSGRGPSPLAKLEDWINVPCPSCGK----------PARRETDTMDTF
SYL-AQUAE
SYL_SYNY3 701 TNEPLTKAEKDLRRAIHGAIKEVAEDLNDDYQFNTAISEMMKLSNALIAATD-------L
SYL_ECOLI 690 NVDALTENQKALRRDVHKTIAKVTDDIGRRQTFNTAIAAIMELMNKLAKAPTD-----GE
SYL_BACSU 642 NGKIVEGAGETLERVYHETVMKVTDHY-EGLRFNTGISQLMVFINEAYKATE--------
SYL-BORBU 680 S---KENPPREIISELHKVIKKVTEDT-EKLNFNTAISAMMIFINELLKYEK--------
MJ0633 700 TGEEFSYIDKWLLSRLYKAVKQYDEYMENFELRKAG-ILLYQLLDDLKWYRRR--GGNN-
PH0965|PHAW008 718 GNVELKDIDRWMLHRLNKAIKETTNALEEFRTRTAVQWAFYSIMNDLRWYLRRTEGRDDE
AF2421 698 ---ELTTLDRWLVSRMQRAIKEVREAMDNLQTRRAVNAAFFELMNDVRWYLRR--GGEN-
MTH1508 685 PAEPESFIGRWMMGQLNQRIREATRALESFQTRKAVQEALYLLKKDVDHYLKRVEGRVDD
SYLC_YEAST 798 TGEITDFFDIAFEHEMNALIEKTYEQYALTNYKNALKYGLFDFQAARDYYREASG----V
SYLC_NEUCR 826 TGELNSFQDALFDNEMNALVNEARKHYEETSYKLALKAAHYDFLNARDMYREACAAAGIP
SYLC_CAEEL 801 RKDAKKFADRVFANEMNSLIAATEQNYEATNFKDALKTGFFEYQAIRDTYRELCAGIDEP
consensus 901 e t dk l in ik e le r av amy lmnel y g
SYL-AQUAE 522 FDSSWYFLRFCDPKNDREPFSREKVDYWMP--V-DVYIGGIEHAVLHLLYARF-------
SYL_SYNY3 501 IDSSWYFLRYADAQNTEIPFDGEKVAHWLP--V-DQYVGGIEHAILHLLYSRF-------
SYL-AQUAE
SYL_SYNY3 754 ISPPVYQEGIETLLLIIAPFAPHLTEELWHRLGRTDSIHQQA-WLQVDPT-----ALVLD
SYL_ECOLI 745 QDRALMQEALLAVVRMLNPFTPHICFTLWQELKGEGDIDNAP-WPVADEK-----AMVED
SYL_BACSU 693 ----LPKEYMEGFVKLISPVAPHLAEELWEKLGHSGTIAYEA-WPVYDET-----KLVDD
SYL-BORBU 728 ----NYLNIFKPFIIILSPYAPHLAEELWEYIGELPSLFKNSKWPKFDES-----LIIKG
MJ0633 756 --IRVLEEFLEVIIKLMMPFTPHLCEEMWEI-LGKEGFVSLAKFPEVKEE-----FINDE
PH0965|PHAW008 778 AKRYVLRTLADVWVRLMAPFTPHICEELWEK-LGGEGFVSLAKWPEPVEE-----WWNET
AF2421 752 -----LAIILDDWIKLIAPFAPHICEELWH--LKHDSYVSLESYPEYDET-----RVDEE
MTH1508 745 EVKSVLANVLHAWIRLMAPFIPYTAEEMWER-YGGEGFVAEAPWPDFSDD-----AESRD
SYLC_YEAST 854 MHKDLIARYIETQALLIAPIAPHFAEYIYREVLGNQTSVQNAKFPRASK------PVDKG
SYLC_NEUCR 886 LHKDLVTKYIRLQALVITPIAPHWADYVWQECLGEPKSIQFARWPEVP-------AANPA
SYLC_CAEEL 861 MSESLVFRFIETQMLILSPICPHIAEYIWQ-LLGKDGLIVNAPWPTVDPVDEKLAIGARF
consensus 961 vm ie lrll pfaphlaeelw llg d v a wp de l d
SYL-AQUAE 572 --------FQKFLKDLGLVR----------DDEPFEKLITQGMVL-----K-------K
SYL_SYNY3 551 --------FTKVLADRQLIP----------VKEPFQKLLTQGMVQGITYKNPTTG---K
SYL-AQUAE
SYL_SYNY3 808 EITLVIQVLGKTRGTITVPASADKT--------------------QLE----------EL
SYL_ECOLI 799 STLVVVQVNGKVRAKITVPVDATEE--------------------QVR----------ER
SYL_BACSU 743 EVEIVVQLNGKVKAKLQVPADATKE--------------------QLE----------QL
SYL-BORBU 779 KKEIVLQINGKIKDKILLNKETGEK--------------------ELK----------EI
MJ0633 808 IEKGEEYLKAVMEDIKEIINVAKVQ----------KRIYLYTAD---DWKYEILKI
PH0965|PHAW008 832 IEAEEEFIRSVMEDIKEIIEVAKIEN----------AKRAYIYTAE---DWKWKVAEV
AF2421 800 AERIEEYLRNLVEDIQEIKKFVSDA----------KEVYIAPAE---DWKVKAAKV
MTH1508 799 VQVAEEMVQNTVRDIQEIMKILGST----------PERVHIYTSP---KWKWDVLRV
SYLC_YEAST 908 VLAALDYLRNLQRSIREGEGQALKKK-KGKSAEIDASKPVKLTLLISESPPEWQSQCVEI
SYLC_NEUCR 939 LTAARDYVRTTSSAINSAEAAQLKKMAKGRQSDFDPKKPKKLTIPATENFPTWQAKYIDL
SYLC_CAEEL 920 ITESLAEFRARLKTYMTPKKKALKEIP---------EVPTEAVIYVAKEYPPWQKTILDI
consensus 1021 i ve vkg vr i v a l e w el
SYL-AQUAE 601 WVSVKKL----------LDYL--GLSEEDEVEELKKRLEELGARRA----------
SYL_SYNY3 589 YVPAKDLQTGQ-Q-----VIDPK--DPKDPDSGEPLQVFYEKMSKSKFNGVDPQEVLAKY
SYL_ECOLI 585 WVSPVDAIVERDE-----KGRIV--KAKDAAGHELVYTGMSKMSKSKNNGIDPQVMVERY
SYL_BACSU 577 ----------MSKSKGNVVNPDEIVASH
SYL-BORBU 587 GVLIPNDQV---------IEKDN--KFFDKKDNKEVTQVIAKMSKSLKNVINPDDIIKEF
MJ0633 593 KDLIPNHLTFYIFNHVAIFPEEF--WPRGIVVNGYVTIEGKKLSKSKGPVLPVLEVAEKF
PH0965|PHAW008 611 KDLIPNHLTFFIFNHVAIFREEH--WPKGIAVNGFGTLEGQKMSKSKGNVLNFIDAIEEN
AF2421 591 KDLVANHLLFYLFHHVAIFPPDK--WPRAIAVNGYVSLEGKKMSKSKGPLLTMKRAVQQY
MTH1508 570 KDLIGNHLTFHIFHHSAIFPESG--WPRGAVVFCMGLLEGNKMSSSKGNVILLRDAIEKH
SYLC_YEAST 690 KDLIPNHLTFFIYTHVALFPKKF--WPKGIRANGHLMLNNSKMSKSTGNFMTLEQTVEKF
SYLC_NEUCR 718 KDLIPNHLTFWLYNHIALFPREY--WPKSVRANGHLQLNGEKMSKSTGNFMTLDDVVKKY
SYLC_CAEEL 689 KDLIGNHLTYLLFNHAAIWPTDTSKWPKGIRANGHLLLNNEKMSKSTGNFMTLEEAIEKF
consensus 781 kdli nhltf if h aifp d wpkgi ng l l kmskskgnvv v evleky
consensus 241 i vdwrr f Ttdan f fi wq hl erk ir g r t y drl