Chapter 1

Introduction to Aminoacyl-tRNA Synthetases and Incorporation of Non-natural Amino Acids

1. Aminoacyl-tRNA Synthetases (aaRSs)

Protein biosynthesis involves two major steps: transcription (DNA directed synthesis of messager RNA) and translation (mRNA directed synthesis of protein). During transcription, DNA is strictly copied into mRNA through complementarity of nucleic acids. Translation is an mRNA templated polymerization to join the amino acids into a protein. This process is more complicated than transcription because there is no complementarity between nucleic acids and the structurally diverse amino acids. An "adapter" molecule, transfer RNA (tRNA), participates in translation and acts as an amino acid carrier (Figure 1-1). Upon aminoacyl-tRNA synthesis, the amino acids are attached to their cognate tRNAs; by pairing of a codon in mRNA with the anti-codon in tRNA the amino acid is inserted into a particular position of a growing polypeptide. Accurate joining of an amino acid to a cognate tRNA therefore largely determines the fidelity of protein biosynthesis. Synthesis of aminoacyl-tRNA is realized in the following two-step reaction:

$$aaRS + aa + ATP \leftrightarrow aaRS \bullet (aa - AMP) + PP_i \qquad (1)$$

$$aaRS \bullet (aa - AMP) + tRNA^{aa} \rightarrow aa - tRNA^{aa} + aaRS + AMP \qquad (2)$$

where aa is an amino acid and aaRS is the corresponding aminoacyl-tRNA synthetase.

Aminoacyl-tRNA synthetases (aaRSs) are a remarkable class of enzymes that catalyze aminoacyl-tRNA synthesis and safeguard the fidelity of information flow from

nucleic acid sequences into biologically functional proteins. In the first step of the synthesis of aminoacyl-tRNA, the amino acid reacts with ATP to form an aminoacyl adenylate intermediate; in the second step, the amino acid moiety is joined to an appropriate tRNA. In order to maintain such a highly accurate reaction, an aaRS has to recognize one amino acid out of a pool of twenty canonical amino acids and a few non-canonical amino acids, as well as one or more cognate tRNAs out of twenty tRNA families. tRNAs are usually 76-93 nucleotides long and fold into a L-shaped structure; the structure is stabilized by tertiary interactions between bases in the D-loop and T-loop (Figure 2) (*1, 2*). The D-stem, anticodon loop and acceptor stem form the inside surface of the L-shaped tRNA and interact with most of the aaRS (Figure 2b) (*2*). Therefore, specific recognition of tRNA molecules by aaRSs is usually not a problem since they are large enough to be presented to synthetases and form a number of specific interactions between nucleotides and amino acids (Figure 2c) (*3, 4*). Error rates for tRNA recognition are generally less than 10^{-6} (*5*). The situation is different when synthetases discriminate substrate amino acids; amino acids are much smaller and a few of them even have structurally or chemically similar side chains. In 1957 Linus Pauling reported that if aaRSs distinguish the two amino acids with only a difference of a single methylene group, the estimated differences of the binding energies could not allow the error rate to be better than 1 in 5 (*6*). The observed error rates for aaRSs to select their cognate amino acids are much lower, ranging from 10^{-4} to 10^{-5} (*7*). Additional proofreading or editing mechanisms are required to ensure the low error rate. These mechanisms will be discussed substantially in the following sections on isoleucyl-tRNA synthetase (IleRS), valyl-tRNA synthetase (ValRS) and leucyl-tRNA synthetase (LeuRS).

The majority of organisms have 20 aaRSs, one for each natural amino acid. These 20 canonical aaRSs can be divided evenly into two different classes (Table 1-1) (*8, 9*), as a result of the amazing discovery of two fundamentally distinct catalytic domains. The catalytic domains of class I enzymes have the representative dinucleotide binding Rossmann fold (*10*), consisting of a central five-stranded parallel β-sheet connected by α-helices. Class I synthetases also incorporate the short consensus HIGH and KMSKS motifs, which are involved in constituting an ATP binding site. On the other hand, the class II enzymes maintain an unusual catalytic domain, composed by a six-stranded antiparallel β-fold flanked by loops and α-helices (*11-13*). Three characteristic motifs have been identified in these class II synthetases; motif 1 constitutes an interface between two monomers in the dimeric enzymes; motifs 2 and 3 are involved in amino acid and ATP recognition and binding (*14*). In terms of quaternary structure, class I enzymes are generally monomers whereas the class II enzymes tend to form dimers (Table 1-1). Class I aaRSs also differ from the class II enzymes in the aminoacylation reactions; the class I enzymes approach their tRNA substrates from the minor groove of the acceptor stem and join the amino acids to the 2'OH group of the ribose of the 3' terminal adenosine base of the tRNA, whereas in class II families the synthetases approach the tRNA from the major groove side of the acceptor stem and attach the amino acids to 3'OH ribose group.

In last decade aaRSs have been subjected to extensive structural study. To date, except for AlaRS, representative structures for all of the synthetases have been solved. In class I aaRSs, structures of ArgRS (*15*), CysRS (*16*), IleRS (*17, 18*), LeuRS (*19*), MetRS (*20-22*), ValRS (*23*), GlnRS (*24-27*), GluRS (*28, 29*), TrpRS (*30, 31*) and TyrRS (*32-35*) have been determined; for class II enzymes, atomic resolution structures of GlyRS (*36,*

37), HisRS (*38-41*), ProRS (*42*), ThrRS (*43*), SerRS (*44-46*), AsnRS (*47*), AspRS (*48-52*), LysRSII (*53-55*), and PheRS (*14, 56, 57*) have been solved. Insights gleaned from these structures and from sequence comparisons allow one to further arrange the synthetases in one class into subclasses depending on the locations and sequence similarity of modular domains that are mostly involved in recognition of the anticodon nucleotides of the tRNA (Table 1-1) (*5, 58-60*).

Structural analysis of the active sites of enzymes and numerous biochemical data provide reasonably clear pictures of the recognition of substrate amino acids by their cognate synthetases and plausible mechanisms of amino acid activation. The features of such mechanisms are rather different for class I and class II enzymes (*58*). Class I aaRSs require an induced-fit mechanism, where a transition state is stabilized by the energy released by amino acid activation reaction to form the adenylate (*61, 62*). Recent investigation of the detailed structures of TrpRS in the ligand-free state, in a closed pre-transition complex and in a relaxed adenylate product complex further supports such a mechanism (*31*). Notably, GlnRS, ArgRS, GluRS and LysRSI require tRNA to catalyze amino acid activation. Upon complexation with their tRNAs, the activation sites of these synthetases are reorganized into particular configurations to favor binding of cognate amino acids (*27, 63, 64*). The specificity of synthetases toward amino acids will be further discussed in great detail for IleRS, ValRS and LeuRS in following sections. In contrast, the active sites of class II aaRSs form rigid templates so that the ATP and amino acid substrates adopt optimal orientations to facilitate in-line nucleophilic reactions to form the transition state (*58, 65-67*). The specificity of class II enzymes is determined by complementary contacts between the side chains of the amino acid substrates and the

active sites adapted to engulf the cognate amino acids. An induced fit mechanism imposed by ATP binding to the loop of motif 2 further refines this specificity; this effect of "induced fit" varies on different class II aaRSs; for PheRS (discussed in the following section), the effect is fairly modest (*14*).

In the following sections, I will review some of the aaRSs that are associated with this thesis work, including IleRS, ValRS, LeuRS and PheRS.

2. Isoleucyl-tRNA Synthetase (IleRS)

IleRS is a monomeric, class Ia synthetase (Table 1-1), containing approximately 800-1000 amino acid residues (*E. coli*: 937; *T. thermophilus*: 821; *S. aureus*: 917; *S. cerevisiae*: 1072) (*68*). Many of the insights on the mechanism of this family of enzymes emerge from the determination of crystal structures of IleRS in complexes with different substrates (*18, 69*), as well as with tRNA (*17*). The crystal structure of IleRS from *T. thermophilus* is an L-shaped molecule with dimensions of 100 Å by 80 Å by 45 Å (Figure 1-3) (*18*), consisting of a β-strand rich domain on the top, a Rossmann-fold domain in the middle and an α-helical rich domain at the bottom. The Rossmann domain is composed of alternating β-strands and α-helices forming a $\beta_6\alpha_4$ structure. In the active site, the hydrophobic side chain of substrate isoleucine is surrounded by Gly45, Pro46, Trp518 and Trp558 through van der Waals interactions (Figure 1-4a); Asp85 and Gln554 form hydrogen bonds with the ammonium and carboxylate of isoleucine, respectively. The site of the pocket appears to be small enough that the larger amino acids (and leucine) cannot fit in (*18*). Valine is observed to bind to the same synthetic active site when one soaks the crystal with valine (Figure 1-4b). Lacking one methylene group, the

side chain of valine has less hydrophobic contact with Pro46 and Trp518 than that of isoleucine, while the polar interactions of the two substrates are almost identical. This is in good agreement with what Pauling proposed (*6*) and with kinetic data previously reported (*70*). The structure of IleRS complexed with Ile-AMS (a non-hydrolyzable analogs of Ile-AMP) was determined with 3.0 Å resolution (*69*). Recognition of the isoleucyl moiety of Ile-AMS is nearly identical to that of isoleucine, except that hydrogen bond is formed with Asp85 in place of Pro46 in the IleRS/Ile complex (Figure 1-5) (*18*).

Sequence alignment of IleRS from different organisms consistently reveals an insertion in the Rossmann-fold domain of approximately 200 amino acids, which is called the connective polypeptide 1 (CP1) (*71*). Biochemical data showed that mutation of certain residues in the CP1 domain impairs the hydrolysis of Val-tRNA^{Ile} (*72*); the cloned CP1 domain as an independent protein exhibited hydrolytic activity toward to Val-tRNA^{Ile} (*73*). All of this evidence suggests that the CP1 domain is the catalytic center for the editing function. When Nureki and co-worker soaked the crystal of IleRS from *T. thermophilus* with valine, valine not only appeared in synthetic active site of IleRS, it was also seen in the CP1 domain; no second center of electron density was observed when crystals were soaked with isoleucine (*18*). The structure of the CP1 domain showed that the pocket constituted by Trp232 and Tyr386 can only accommodate valine; isoleucine is too large to fit into the pocket. Therefore this is the first structural evidence to support the double-sieve mechanism for editing proposed by Fersht et al. (*74-76*). The remaining question is how the substrate (either Val-tRNA^{Ile} or Val-AMP) is shuttled from the synthetic site to the editing site for proofreading since the two sites are separated by more than 30 Å (Figure 1-6) (*18*).

The atomic structure of tRNA complexed with IleRS from *S. aureus* sheds light on the translocation problem (Figure 1-6) (*17*). The overall structure of IleRS from *S. aureus* resembles that from *T. thermophilus*. The acceptor stem of bound tRNA (located at the 3' terminus) showed a helical conformation (*17*), which could be observed with uncharged tRNA (*77, 78*) or tRNAs complexed with class II synthetases (*13, 79*). Further modeling showed that the 3' end of tRNA could not reach the synthetic site of IleRS in the Rossmann-fold domain, instead it pointed toward the CP1 domain, forming potential interactions with residues His392 and Tyr394, which have been shown to be located in the center of editing site and are the two of important residues to make up the editing pocket (*18*). Therefore this tRNA complexed structure might represent an editing state. A shuttling mechanism was proposed wherein the 3' end of tRNA translocates from the synthetic active site to the editing site by changing its conformation from a hairpin structure to a helical structure. Similar shuttling of nucleotides for editing was also observed in DNA polymerization; one strand of DNA is shuttled between the polymerase active site and the editing site by processive sliding (*80-84*). Schimmel and co-workers also reported a similar shuttle mechanism for *E. coli* IleRS in editing Val-AMP, a pre-transfer editing, by a fluorescent study (*85*).

Data accumulated from these structural (*17, 18, 69*) and biochemical (*70, 74, 85-87*) studies constitute a rather completed picture of the proofreading mechanism of IleRS. In the pre-transfer editing, upon complexing with $tRNA^{Ile}$, IleRS directly hydrolyzes the misactivated Val-AMP into Val and AMP. For Val-AMP that escaped from the pre-transfer editing, a post-transfer editing can make the further corrections; Val is attached to 3'-end of $tRNA^{Ile}$ to generate Val-$tRNA^{Ile}$, which is subsequently hydrolyzed into Val

and $tRNA^{Ile}$ by the post-transfer editing of IleRS. Both editing reactions require a shuttling mechanism. In Chapter 3&4 of this thesis, I will present cases where we identified two non-natural amino acids that can escape this editing and be charged to $tRNA^{Ile}$.

3. Valyl-tRNA Synthetase (ValRS)

In addition to valine, ValRS can also activate structurally similar canonical amino acids threonine (*76*) and cysteine (*88, 89*) as well as non-canonical amino acid α-aminobutyrate (*75, 89, 90*). Fersht proposed a "double-sieve" model similar to that described for IleRS to explain these observations. Recently the structure of ValRS has been solved and it supports the "double-sieve" model firmly (*23*). ValRS from *T. thermophilus* is a monomeric synthetase with 862 amino acid residues and belongs to the class Ia family. The domain structures of ValRS, including the CP1 editing domain, Rossmann-fold domain, stem-contact (SC) domain, helix bundle domain, anticodon stem binding domain and coiled-coil domain, are shown in Figure 1-7. The recognition of Val-AMS at the active site by ValRS is illustrated in Figure 1-8. The ammonium group of aminoacyl moiety of Val-AMS forms hydrogen bonds with Pro42, Asp81 and Asn44; Pro41, Pro42, Trp456, Ile491 and Trp495 constitute a hydrophobic pocket to hold the valyl side chain. In the binding pocket of IleRS (as described in the previous section), the hydrophobic pocket for the isoleucyl side chain is composed by Gly45, Pro46, Trp518 and Trp558. Because of substitution of Gly 45 in IleRS by Pro41 in ValRS, the pocket of ValRS is shallower and narrower; in fact modeling shows that it is Pro41 in ValRS that prevents binding of isoleucine into the pocket (*23*). Therefore the active site of ValRS

can act as a first sieve to exclude any amino acids larger than valine from the pocket. Threonine, as well as cysteine and α-aminobutyrate, is slightly smaller than valine ($V_{(valine)}$=140 Å^3; $V_{(threonine)}$=116 Å^3) and could pass this sieve to fit into the pocket of ValRS. ValRS and IleRS use the similar means to recognize the adenosyl moiety (Figures 1-5; 1-8).

Comparison of the structures of IleRS and ValRS (Figures 1-6; 1-7) shows similar arrangements of the CP1 domain. In ValRS, the editing site is approximately 37 Å away from the catalytic site. The CAA end of bound $tRNA^{Val}$ stretches toward the editing domain and is anchored by two β-strands (Figure 1-7). Specific interactions between A76, the 3' terminal nucleotide, with residues at the editing domain, have been identified (*23*). A76 is embedded between Phe264 and Leu269 through van der Waals contacts and also forms hydrogen bonds with Tyr337 and Thr214. This complex structure is believed to represent a configuration that is involved in post-transfer editing. Based on the editing site identified in IleRS (with a bound valine) and structure modeling between IleRS and ValRS, Fukai and co-workers were able to determine a post-transfer editing pocket in the editing domain of ValRS (*23*), which is constituted by Arg216, Thr219, Lys270, Thr272, Asp276 and Asp279. It is conceivable to argue that this hydrophilic editing pocket can favor binding of threonine; the γ-OH of threonine forms a hydrogen bond with Asp276. The lack of such a hydrogen bond interaction may lead to exclusion of the cognate amino acid valine from this post-transfer editing pocket. Employing a similar strategy, Fukai et al. postulated a pre-transfer editing pocket for ValRS to proofread Thr-AMP. In this editing pocket, the adenosyl binding site for Thr-AMP shares the same site used for binding of adenosyl of Thr-$tRNA^{Val}$ at the post-transfer step; the newly identified binding

site for threonyl moiety is composed of Asp276, Arg216 and Leu341, where Asp276 and Arg216 are believed to be involved in hydrogen bond interactions with the γ-OH of threonine. Therefore, in the both editing scenarios (pre-transfer or post-transfer), the second sieve of ValRS is based on use of the hydrophilic γ-OH to discriminate threonine from valine, contrasting sharply to that of IleRS, where discrimination is mostly based on size.

By manipulating the editing pocket of *E. coli* ValRS, Schimmel and co-workers were able to expand the amino acid set of *E. coli* (*91*). A selection was designed to screen for chromosomal mutations to favor the charging of cysteine to $tRNA^{Val}$. The authors found that all identified mutations resided in the editing domain of ValRS. One such editing disabled mutant *E. coli* strain was able to incorporate more than 20% non-canonical amino acid α-aminobutyrate into cellular proteins at valine codon sites, suggesting that editing may play an important role in restricting the genetic code to 20 amino acids and suppressing the genetic code ambiguity. Further experiments showed that cell viability is heavily reliant upon the accuracy of translation imposed by ValRS: the concentration of α-aminobutyrate to arrest the growth of *E. coli* cells inversely correlates with the impaired level of editing function (*92*). In Chapter 4 of this thesis, I will present an example where we found a non-canonical amino acid able to escape editing by ValRS and infiltrate into valine codon sites. In Chapter 7 of this thesis, I will describe an editing impaired mutant LeuRS to facilitate incorporation of a novel amino acid that otherwise could be edited by wild-type LeuRS.

4. Leucyl-tRNA Synthetase (LeuRS)

Studies on LeuRS are not as thorough as those of IleRS and ValRS. Sequences of LeuRS from many organisms have been identified (*68*). LeuRS is a monomeric enzyme with 800-1000 amino acid residues and belongs to the class I family. Sequence alignment between IleRS, ValRS and LeuRS from *T. thermophilus* indicates the high similarity of these three enzymes and high conservation of their catalytic domains (*19*). The structure of LeuRS from *T. thermophilus* has been determined to 2 Å resolution (Figure 1-9). The overall structure exhibits some characteristic features in common with IleRS and ValRS (*93*). The catalytic Rossmann-fold domain is situated at the center of the structure, followed by a connecting module and tRNA anticodon binding domain (Figure 1-9) (*19*). The CP1 domain of LeuRS extends from residue 224 to residue 417, which is comparable to the CP1 domains found in IleRS and ValRS. However, the location of the insertion in LeuRS is strikingly different from ValRS or IleRS (*19*); CP1 is set after a 50-residue long peptide in LeuRS, whereas the CP1 domain in IleRS or ValRS is located before this unique peptide. This peptide is involved in zinc binding and maintenance of the correct folding of the elongated LeuRS structure (*18, 19*). LeuRS also contains a special insertion domain that is not identifiable in IleRS and ValRS; this domain is comprised mainly of β-strands and likely involved in recognition of $tRNA^{Leu}$ (Figure 1-9).

The active site of LeuRS was identified by soaking the enzyme with a non-hydrolysable adenylate analog Leu-AMS, as shown in Figure 1-10. The ammonium and carboxyl group of leucyl moiety of Leu-AMS forms hydrogen bonds with Asp80, Phe41 and His541. A hydrophobic pocket composed of Met40, Phe41, Tyr43, Phe501, Tyr507, His541 and His545 makes van der Waals contacts with the side chain of the leucyl

moiety. A water molecule is identified near the leucine side chain and stabilized by surrounding Tyr43, Asp80 and Thr507 through hydrogen bonds, supporting the early report that LeuRS can misactivate γ–hydroxy-leucine (*94*). As expected, the adenosine moiety of Leu-AMS makes extensive hydrogen bond interactions with residues Glu540, Gln574, Val577, Met638, Tyr43 and Arg178, some of which are through the water bridges. Comparing the ligand-free and ligand-bound structure of LeuRSs, one can observe movement at the catalytic site. This conformational change is largely induced by adenosine moiety in order to form hydrogen bonds with Glu540 and Gln574. Other extended conformational changes are also observed and their roles in recognition and enzyme function remain to be explored.

Recent structures of LeuRS complexed with analogous substrates at pre- and post-transfer editing steps provide considerable insight into the mechanism of catalytic hydrolysis at the editing site (*95*). As we discussed in the ValRS section, structural modeling and comparative analysis of IleRS and ValRS led to a proposal that two distinct editing sites are responsible for pre- and post-transfer editing steps of ValRS and IleRS; the two sites approximate each other, sharing the same binding site to accommodate adenosyl moiety of AMP analogs and having two distinct binding sites for recognition of aminoacyl moieties (*18, 23*). Schimmel and co-workers were able to separate the pre- and post-transfer editing steps by introducing mutations at different sites of the CP1 domain of *E. coli* IleRS, supporting the proposal of two distinct sites involved in editing (*96*). In contrast, LeuRS utilizes the same editing site to bind the two distinct substrates for the two different editing processes. It has been shown or suggested that LeuRS edits canonical amino acids methionine and isoleucine (*97, 98*) as well as some

nonproteinogenic amino acids such as norleucine, norvaline, homocysteine and γ–hydroxy-leucine (*94, 99-101*). Cusack and coworkers synthesized 5'-O-[N-(L-norvalyl)sulphamoyl]adenosine (Nva-AMS) as a substrate for pre-transfer editing and 2'-(L-norvalyl)amino-2'-deoxyadenosine (Nva-2AA) as a substrate for post-transfer editing (*95*). According to electron density maps, Nva-AMS appeared in both synthetic and editing site of the protein. In the synthetic site, Nva-AMS exhibits an extended conformation and maintains interactions similar to those of Leu-AMS. Interestingly, a bent conformation is found when Nva-AMS resides in the editing site and the side chain of norvaline is located in a hydrophobic pocket consisting of Met338, Val340, Thr252 and the aliphatic part of Arg249. Mursinna et al. previously identified a "gatekeeper residue" T252 in the editing pocket by mutational analysis and was able to show that mutating T252 into A252 enables LeuRS to hydrolyze Leu-tRNA^{Leu}, indicating that the T252 might play a very important role in sterically excluding binding of leucine (*100*). Modeling of leucine in this pocket evidently showed that this additional methyl group would cause clashes with T252 and V340, in accord with the mutational study (*95*).

Nva-2AA, an analog likely representing the amino acid attached to the 3' end of tRNA, can only be seen in the editing site. Although the manner of recognition of adenine base and aminoacyl moiety is strikingly similar for Nva-2AA and Nva-AMS, a notable difference is observed on the orientation of ribose, indicative of the different linkage to the amino acid. The nearly identical conformation and position for the adenosyl moiety of Nva-2AA and the terminal A76 bound to ValRS (*23*) suggest that LeuRS complexed with Nva-2AA represents the structure involved in post-transfer

editing, as further confirmed by competition assay of Nva-2AA and Nva-3AA for replacing mischarged Ile-tRNA^{Leu} at the editing site (*95*).

LeuRS bound with either Nva-AMS or Nva-2AA also sheds light on a plausible mechanism of hydrolysis for editing (*95*). The hypothesis for this catalytic hydrolysis is that the substrate (Nva-AMS or Nva-2AA) is positioned in the pocket with a configuration favorable to be attacked by a water molecule. In both cases a water molecule is identified to hydrogen bond with Asp344 and the bridging amide of either substrate; it is possible that this water, brought close enough to the reaction center by hydrogen bonding, facilitates the nucleophilic reaction to hydrolyze the substrate, although the detailed structural information to support such a model needs further investigation.

Supported by structural and biochemical data, mutation of T252 to A252 decreases the constraint of the "second sieve" and leads to mis-editing of Leu-tRNA^{Leu}. Recent reports from our laboratory (*97*) and others (*102*) independently showed that mutation of T252 to bulkier residues such as leucine, phenylalanine and tyrosine blocks the editing site so that methionine, isoleucine and many non-proteinogenic amino acids are able to be stably attached to tRNA^{Leu}. In Chapter 7, I will present an example that a mutant LeuRS with attenuated editing function allows incorporation of a novel amino acid with unique chemical functionality into proteins *in vivo*.

5. Phenylalanyl-tRNA Synthetase (PheRS)

PheRS is one of the class II synthetases and probably the largest aaRS found so far. It has two subunits (α and β), constituting a $\alpha_2\beta_2$ hetero-tetrameric structure (*12*). In

E. coli the α and β subunit of PheRS are encoded by the *pheS* and *pheT* genes, respectively; these two genes are adjacent to each other with an intercistronic sequence between, allowing the re-initiation of translation of the β-subunit after the stop codon of the α subunit (*103*). Affinity labeling studies showed that the β-subunit of *E. coli* PheRS has a tRNA binding site and is likely involved in tRNA recognition, whereas the α subunit is responsible for the catalytic function (*104*). Early experiments showed that the wild-type *E. coli* PheRS can attach *para*-fluorophenylalanine to $tRNA^{Phe}$, but not *para*-chlorophenylalanine or *para*-bromophenylalanine (*105*). Hennecke and co-workers developed a genetic method to identify the catalytic active site of *E. coli* PheRS. A *para*-fluorophenylalanine resistant strain was selected and the PheRS was isolated and sequenced. This mutant strain contains an Ala294Ser mutation to confer the resistance (*106*). Further mutational analysis confirmed the important role of Ala294 in recognition of amino acid substrate and one of mutants, A294G, exhibited a high sensitivity toward *para*-chlorophenylalanine or *para*-bromophenylalanine, implicating that this mutation could enlarge the binding site of PheRS (*106*).

Recently the crystal structures of *Thermus thermophilus* PheRS complexed with $tRNA^{Phe}$, phenylalanine and phenylalanyl-adenylate analog have been determined to 2.5-3.28 Å resolution (Figure 1-11) (*14, 56, 57*). Structures confirmed that the α-subunit is the actual catalytic unit, whereas the function of the β-subunit is recognition and binding of tRNA. The β-subunit consists of 8 topologically different domains; one such domain displays the same topology as α-subunit of PheRS, but lacks the catalytic activity. The complexed structures with either phenylalanine (Phe) or phenylalanyl-adenylate analog (PheOH-AMP) elucidate the detailed interactions between ligand and protein. At the

amino acid binding site, hydrogen bonding and van der Waals are the major interactions that control the binding and orientation of Phe (Figure 1-12). The polar parts of Phe are anchored by hydrogen bonding with S180, Q218, E220, W149, H178 and R204. The side chain of the substrate is embedded in a distinct hydrophobic pocket, consisting of E220, F258, F260, V261, G282, A283, G284, A314, F315 and G316. The phenyl ring is sandwiched between the phenyl side chains of F258 and F260; V261 and A314 construct a back-wall to prevent further penetration of the Phe side chain into the interior of the protein. Biochemical studies showed that mutation of A294 to G294 or S294 in *E. coli*, which is structurally equivalent to A314 in *T. thermophilus*, could decrease the affinity binding of Phe to PheRS (*107, 108*). However, the A294G mutation can relax the substrate specificity to allow binding of several *para*-halogenated Phe analogs (*107, 108*). Removing the side chain of A314 in the crystal structure can create extra room in the substrate binding cleft, consistent with the biochemical experiments. In the structure complexed with PheOH-AMP, the phenylalanyl moiety has nearly identical position and interactions as Phe in PheRS. Extensive hydrogen bonds are involved in interactions between AMP and PheRS in the active site (*14*).

The active-site pockets of PheRS bound to PheOH-AMP and to Phe are almost super-imposable, indicating the rigid-template nature of the catalytic site of this enzyme family. The rigidity suggests that PheRS may be a good candidate for rational re-design of synthetase activity. In Chapter 2, I will describe an attempt to use computational method to screen ligand candidates to be recognizable by PheRS. In Chapter 5 and 6, I will present some computationally designed mutants of PheRS exhibiting relaxed

substrate specificity and allowing incorporation of several amino acid analogs into proteins.

6. Site-specific Incorporation of Non-natural Amino Acids into Proteins by Chemically Misacylated tRNA

Site-directed mutagensis has been the most powerful tool for protein engineering (*109*) and it allows one or several amino acids in proteins to be substituted by any of the other 20 natural amino acids. But the chemical invariance of the 20 natural amino acids limits our ability to manipulate the structure and function of proteins. During the last decade, the efforts have been directed to expand the set of amino acids as building blocks for biomacromolecules. One of the methods is to use chemically misacylated tRNA as the non-natural amino acid supplier and an amber codon (nonsense codon, stop codon) as the sense codon for this non-coded amino acid to site-specifically insert this amino acid to the programmed site of proteins (Figure 1-13) (*110-149*). In 1984 Hecht and co-workers reported a method to prepare chemically misacylated tRNA (*150*). Both the Schultz and the Chamberlin groups adopted this method and combined it with the cell-free translation technology to accomplish early successes of site-specific incorporation of non-natural amino acids (*110, 111*). Briefly, the non-natural amino acid of interest is chemically attached to the suppressor tRNA, which contains an anticodon corresponding to the UAG amber codon; site-directed mutagenesis introduces an amber codon at the intended site of the target protein. This chemically charged tRNA is added into a cell-free translation system and the amber codon is decoded by the suppressor tRNA and the non-natural amino acid is inserted in the intended site of the protein. One of critical

components for this method is choosing the appropriate suppressor tRNA; the charged suppressor tRNA should not be acylated or deacylated by any of the aminoacyl-tRNA synthetases presented in the translation system. Schultz and co-workers initially chose a yeast $tRNA^{Phe}$ with an amber anticodon as the suppressor for use in *E. coli in vitro* translation system because it is known that yeast $tRNA^{Phe}$ is unrecognizable by any of *E. coli* aminoacyl-tRNA synthetases (*110*). Chamberlin and co-workers used *E. coli* $tRNA^{Gly}$ as the suppressor and developed this incorporation method in a rabbit reticulocyte expression system (*111*). Further development by the Schultz laboratory showed that an *E. coli* $tRNA^{Asp}$ exhibits more efficient suppression in the *E. coli* expression system (*151*). To date almost 100 non-natural amino acids have been incorporated into proteins and several applications emerged to use these amino acids as structure probes (*120, 122-125, 152*), functional probes (*153-156*), isotopic labeling (*157-161*), chemical handles (*130*) and fluorescent labeling (*124, 138, 162*). Although this method has been well developed and certain applications have been demonstrated, its labor-intensive nature (large quantities of aminoacylated tRNA) and intrinsic low protein production yield via *in vitro* synthesis (milligram or less) limit its applications (*144*).

Dougherty, Lester, and co-workers recognized such limitations and found a rather unique way around this problem. Combining with this *in vitro* suppression mutagensis method and *in vivo* injection techniques, they devised a method to *in vivo* incorporate non-natural amino acids site-specifically into an intact mammalian cell system (Figure 1-13). By co-injection of charged suppressor tRNA and amber-encoded mRNA into *Xenopus* oocytes, they were able to introduce non-natural amino acids into an α-subunit of the nicotinic acetylcholine receptor (nAchR) (*127, 128, 131-136, 139-143,*

145-149, 163-169). Initially they rationally designed a suppressor tRNA (denoted as MN3) and found it to be effective in oocyte system (*128*). Starting with naturally occurring amber suppressor tRNA from *Tetrahymena thermophila*, they further adopted this suppressor (denoted as THG73) and showed an enhanced suppression efficiency in the oocyte system (*164*). In addition, they were able to circumvent the low quantity protein issue by developing a highly sensitive assay; electrophysiological methods have been employed to measure the protein function and very small amounts of protein are required. Equipped with this incorporation method, the Lester and Dougherty groups have collaborated to address issues in neuroscience, particularly in the ion channels and neuroreceptors that mediate synaptic transmission (*128, 131-133, 140, 141, 148, 166, 167, 169*). This technique has also been successfully applied to mechanistic probes of the structure and function of proteins, particularly in evaluating the role of π-cation interactions in ligand-receptor recognitions and maintaining tertiary protein folding (*170-172*).

In addition to amber suppression, frame-shift suppression (or four codon suppression) (*173-183*), use of an unassigned codon (*184-187*) and non-natural base paring (*188, 189*) have been exploited in this chemically misacylated tRNA method to incorporate non-natural amino acids with the aim to improve translation efficiency and achieve the possibility of introducing multiple non-proteinogenic amino acids into proteins *in vitro*. Nevertheless, the generality of the chemically misacylated tRNA methodology for incorporation of non-canonical amino acids has been staunchly established. The limitations, particularly the quantity limitation, remain to be addressed. Recent advances in the development of site-specific and residue-specific incorporation of

non-natural amino acids into proteins *in vivo* could circumvent some of these limitations, which will be extensively discussed in the following two sections.

7. Site-specific Incorporation of Non-natural Amino Acids into Proteins *In Vivo*

Building on amber suppression mutagenesis *in vitro* described above, Furter reported an early success of site-specific incorporation of novel amino acids *in vivo* in *E. coli* expression system (*190*). The method relied on introduction of a heterogeneous aaRS/tRNA pair (often referred to a "21st pair") from different species (other than *E. coli*). One of the most critical requirements for this method to work is orthogonality (*191-194*); in order to prevent misincorporation of natural amino acids into the amber site and misincorporation of the 21st amino acid into sites other than the intended amber site, the introduced aaRS can only recognize the non-natural amino acid and catalyze attachment of this amino acid exclusively to the introduced suppressor tRNA. aaRS/tRNA pairs from different organisms potentially contain distinct recognition elements and thus could be orthogonal (*195, 196*). It was known that some of the identity elements of yeast $tRNA^{Phe}$ that enable recognition by yeast PheRS are different from those of *E. coli* (*195*). In fact this was the primary reason that Schultz and co-workers chose the suppressor tRNA derived from yeast $tRNA^{Phe}$ for their early attempt to incorporate non-natural amino acids into proteins *in vitro* in an *E. coli* translation system (*110*). *In vitro* (*197*) and *in vivo* (*190*) assays showed that yeast PheRS indeed has negligible reactivity toward any of the *E. coli* tRNAs. Taken together, this pair exhibits the characteristic of orthogonality and could serve as a potential "21st pair" in *E. coli*. By using yeast PheRS, its cognate suppressor $tRNA^{Phe}$ and an auxotrophic strain resistant to

the non-natural amino acid of interest, Furter was able to demonstrate efficient incorporation of *para*-fluorophenylalaine (pFphe) *in vivo* in response to an amber codon, without significant leakage of pFphe into sites encoded by phenylalanine codons (*190*).

Building on their successful experience with *in vitro* systems, Schultz and co-workers recently have also devoted their efforts to development of a similar approach toward the *in vivo* site-specific incorporation of aromatic amino acids into proteins (*198-207*). They identified a TyrRS/$tRNA^{Tyr}$ pair from the archaebacterium *Methanococcus jannaschii* (*M. jaanaschii*), and found that *M. jaanaschii* TyrRS is orthogonal to the pool of *E. coli* tRNAs, whereas *E. coli* TyrRS shows very low reactivity toward *M. jannaschii* $tRNA^{Tyr}$ (*194*). The orthogonality is further improved by replacing eleven nucleotides in the *M. jannaschii* $tRNA^{Tyr}$ region (*208*). Then they devised a powerful section method to identify mutant forms of *M. jannaschii* TyrRS exhibiting high specificity toward the non-natural amino acid of interest. Five key residues in the hydrophobic portion of the active site of *M. jannaschii* TyrRS were all mutated into alanines in order to generate a clean background for screening. Oligonucleotide-encoded DNA shuffling was employed to create a library of mutant forms of *M. jannaschii* TyrRS with mutations on these five key residues. A two-step selection scheme was developed to screen for mutants exhibiting high activity for the non-natural amino acid. In first step (positive selection), cells transformed with the library of mutant *M. jannaschii* TyrRSs, engineered $tRNA^{Tyr}$(CUA), and chloramphenicol acetyltransferase (CAT) gene with one amber codon placed in the permissive position, grow in minimal media supplemented with all 20 natural amino acids and one non-natural amino acid, as well as the antibiotic chloramphenicol. Surviving cells contain the mutant *M. jannaschii* TyrRSs capable of

attaching either non-natural amino acid or one of the natural amino acids to the cognate suppressor. In a second step (negative selection), cells containing the exogenous synthetase/tRNA pair and a lethal gene (Barnase) with 3 amber codons in permissive positions grow in the same media without the non-natural amino acid. Surviving cells would be those incapable of making the full length of Barnase, indicating that their contained *M. jannaschii* TyrRSs have the low capability to charge tyrosine or any other natural amino acid to the suppressor tRNA. After several iterations, cells that survive the first step are selected and analyzed. From the same library, Schultz and co-workers were able to incorporate several amino acids into proteins, including *O*-methyl-tyrosine(*198*), 3-(2-naphthyl)alanine (*203*), *p*-azidophenylalanine (*200*), *p*-benzoylphenylalanine (*201*) and *p*-acetylphenylalaine (*206*).

In Chapter 8, I will describe a rationally identified mutant yeast PheRS, displaying a rather broad spectrum of substrate specificity. Cells transformed with such a mutant allow site-specific incorporation of several non-natural amino acids into proteins *in vivo*.

8. Residue-specific Incorporation of Non-natural Amino Acids into Proteins *In Vivo*

There is another class of investigation in our laboratory directed toward residue-specific incorporation of non-natural amino acids into proteins *in vivo*, in which one attempts to replace one of the 20 natural amino acids in a protein with a desired non-natural amino acid in living bacterial cells, resulting in a protein constituted by 19 natural amino acids and one non-natural amino acid. This work is inspired by Tirrell and co-

workers' early focus on use of DNA templates to direct the synthesis of novel biopolymers in *E. coli* cells (*209-222*). By *de novo* design of artificial genes, several artificial proteins have been prepared, including proteins with controllable crystallization properties (*209*), proteins capable of forming liquid crystal phases and ordered films (*215, 219, 220*), proteins with reversible hydrogel properties (*216*), and matrix proteins potentially applicable to reconstruction of diseased vascular tissues (*217, 218, 221*).

Although the breadth and utility of the *de novo* design of artificial proteins are firmly demonstrated, we run into a dilemma: on the one hand, we would like to design proteins or protein-like polymers with novel side chains because we know from polymer science these side chains could provide particular properties; on the other hand, we are limited by the available set of 20 amino acid building blocks for protein biosynthesis *in vivo* yet only *in vivo* methods can provide enough materials for further characterization and applications. Therefore, one of our objectives has been to develop general methods to incorporate chemically or physically interesting non-natural amino acids into artificial proteins. The method we devised is dependent on the depletion of the cellular pool of one natural amino acid in an auxotrophic bacterial strain. Upon induction, cells are forced to take up the supplemented non-natural amino acid as the natural amino acid surrogate. The cellular aminoacyl-tRNA synthetase must then attach the surrogate to the cognate tRNA. Through *in vivo* translation by cellular ribosomal machinery the mRNA encoding the target protein is then translated and the non-natural amino acid is assigned to the specific codons that normally encode the depleted amino acid. To date we have been successfully introduced alkenes (*223, 224*), alkynes (*225*), cyclobutene (*226*), azido (*227*) and fluorinated side chains (*228-230*) into recombinant proteins expressed in *E.*

coli cells. Further studies showed that the engineered *E. coli* hosts with over-expression of an aminoacyl-tRNA synthetase from *E. coli* could enhance this kind of "re-assignment" of codons to incorporate otherwise sluggish amino acids into proteins (*231-234*). The methods described so far require the *E. coli* aminoacyl-tRNA synthetases to recognize the non-natural amino acids and catalyze the aminoacylation to join them to the appropriate tRNAs.

If one wants to incorporate amino acids not recognizable by any of the *E. coli* aaRSs, re-design of new synthetase activity is likely required. Tirrell and co-workers have successfully adopted a previously identified mutant of *E. coli* PheRS to facilitate assignment of *p*-bromo-, *p*-iodo-, *p*-ethynyl-, *p*-cyano- and *p*-azidophenylalanines to phenylalanine codon positions (*235, 236*). Tang and Tirrell were able to show that the LeuRS mutants with disabled editing function allow incorporation of novel amino acids that are normally edited by the wild-type synthetase (*97*). Part of this thesis will further extend the utility of both methods to incorporate a wide range of amino acids.

Tirrell and co-workers addressed several issues that require this multi-site incorporation of novel amino acid side chains. For example, stabilization of coiled coil proteins was achieved by nearly quantitative replacement of leucine by trifluoro- or hexafluoroleucine (*228, 234*); photochemical crosslinking of elastomeric protein films was accomplished by replacement of multiple phenylalanine by *p*-azidophenylalaine (*237*); post-translational modification of recombinant proteins can be fulfilled by Staudinger ligation with alkyl azide side chains (*227*) and by biotinylation of protein containing ketone side chains (*238*). This thesis will address some of these potential applications as well. Taken together, these results demonstrate vast possibilities to utilize

these novel side chain amino acids to engineer and design unique biomacromolecules for various purposes. Generally this multi-site incorporation is likely to be required if one wishes to change the global properties of a natural protein or a protein-like polymer. Therefore this provides a complementary method to site-specific incorporation of novel amino acids, where one can use the single copy of non-natural amino acid as a local probe, or a chemical handle, as was discussed in the previous two sections.

Table 1-1: Classes of Aminoacyl-tRNA Synthetases

Subclass	Class I		Class II	
	Synthetase	Quaternary Structure	Synthetase	Quaternary Structure
Group a				
	ArgRS	α	GlyRS	α_2
	CysRS	α	HisRS	α_2
	IleRS	α	ProRS	α_2
	LeuRS	α	ThrRS	α_2
	MetRS	α, α_2	SerRS	α_2
	ValRS	α		
	LysRS I[a,b]	α		
Group b				
	GlnRS[a]	α	AsnRS	α_2
	GluRS[a]	α	AspRS	α_2
			LysRS II[b]	α_2
Group c				
	TrpRS	α_2	AlaRS	α, α_4
	TyrRS	α_2	GlyRS	$\alpha_2\beta_2$
			PheRS	$\alpha_2\beta_2$, α

[a]require the presence of tRNA for amino acid activation.
[b]found as both a class I and class II aaRS.

Figure 1-1

The role of aminoacyl-tRNA synthetase in DNA-templated protein biosynthesis. An uncharged tRNA is aminoacylated with cognate amino acid to form aminoacylated tRNA. The fate of this amino acid is determined. The charged tRNA is shuttled into the A site of the ribosome with the assistance of elongation factor. Messager RNA (mRNA) is transcribed from the DNA template and associated with the ribosome. Based on complementarity between codon and anticodon, the amino acid is delivered into the growing polypeptide chain (*58*).

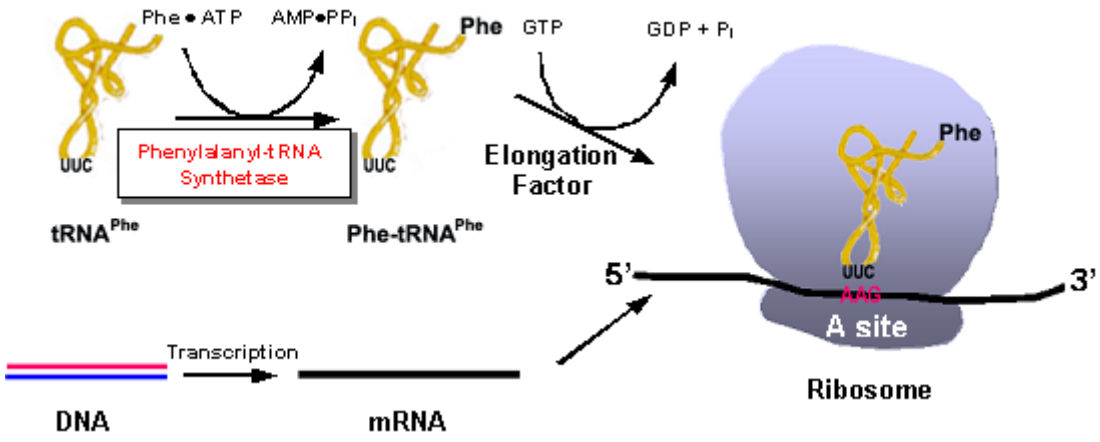

Phe • ATP
AMP•PPi
Phe
GTP
GDP + Pi
UUC
Phenylalanyl-tRNA Synthetase
UUC
Elongation Factor
Phe
tRNAPhe
Phe-tRNAPhe
5'
UUC
AAG
3'
A site
Transcription
DNA
mRNA
Ribosome

Figure 1-2

(a) Primary and secondary structures of tRNA. tRNAs contain many post-transcriptionally modified bases. In particular, adenosine (A) in first position of the anticodon, which corresponds to the third position of the codon or mRNA triplet, is always modified to inosine (I). Inosine can pair with A, U or C and therefore constitutes much of the degeneracy of the genetic code ("Wobble Theory"). The four base-paired stems maintain the "cloverleaf" secondary structure. (b) Tertiary structure of tRNA with an overall compact L shape. The amino acid is attached to an OH group of ribose A located at the upper right. The site for joining the amino acid and the anticodon loop are maximally separated. The tertiary structure is maintained by base-pairing interactions and nucleotide stacking. (c) Interactions between tRNA and aaRS. tRNA is shown in red backbone and yellow bases; aaRS is shown as a space-filling model in blue. The recognition of tRNA by aaRS is sustained by large contact area and interactions between nucleotides of tRNA and amino acid residues of synthetase.

(Picture source: http://anx12.bio.uci.edu/~hudel/bs99a/lecture21/lecture2_2.html)

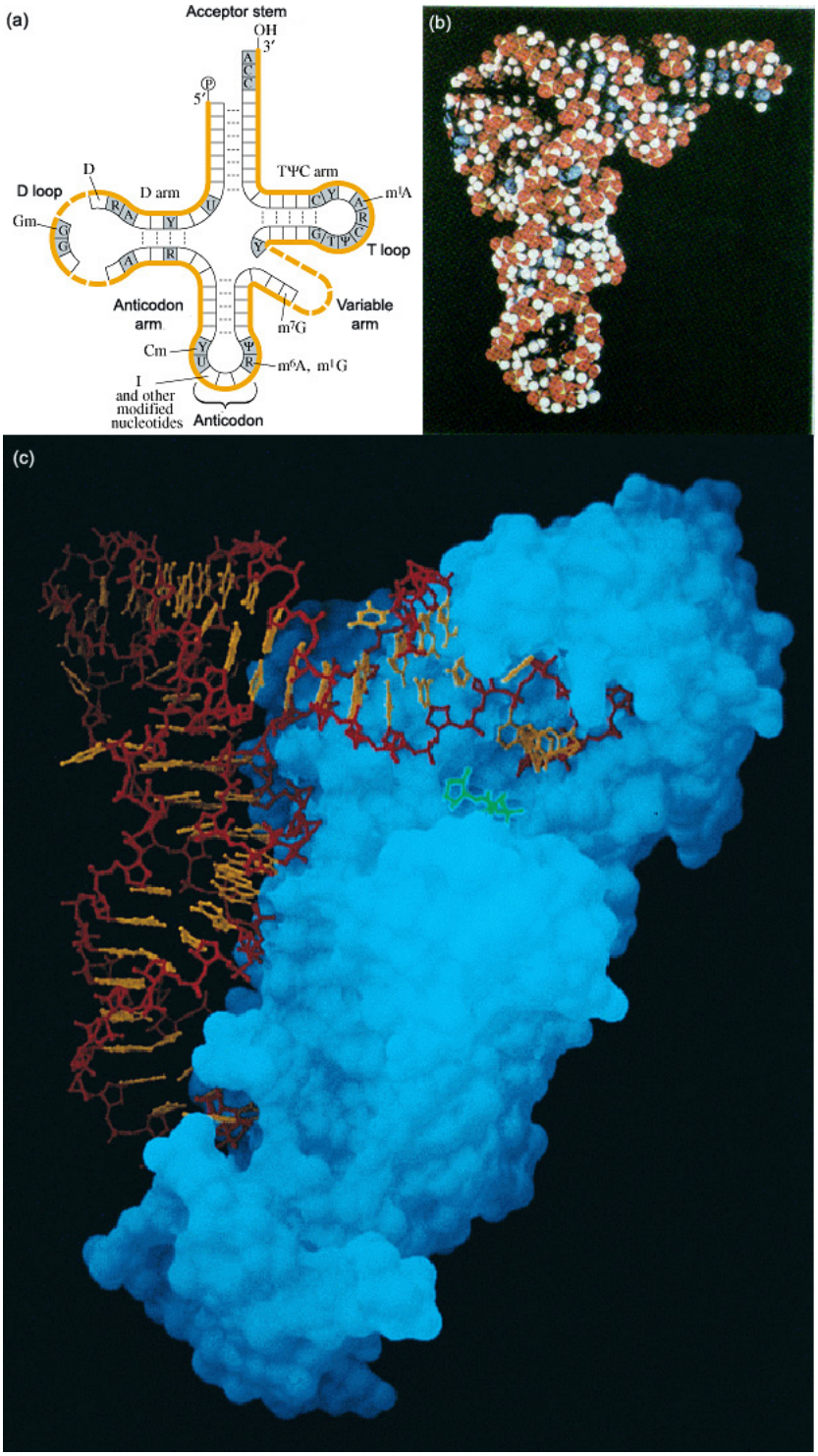

(a)
Acceptor stem
OH
3′
5′
P
D
D loop
D arm
TΨC arm
m¹A
Gm
T loop
Anticodon arm
Variable arm
m⁷G
Cm
m⁶A, m¹G
I and other modified nucleotides
Anticodon
(b)
(c)

Figure 1-3

Ribbon representation of crystal structure of IleRS from *T. thermophilus*. It is a monomeric class I enzyme. The inserted CP1 editing domain and characteristic Rossmann-fold domain are labeled. PDB code for the structure is 1ILE.

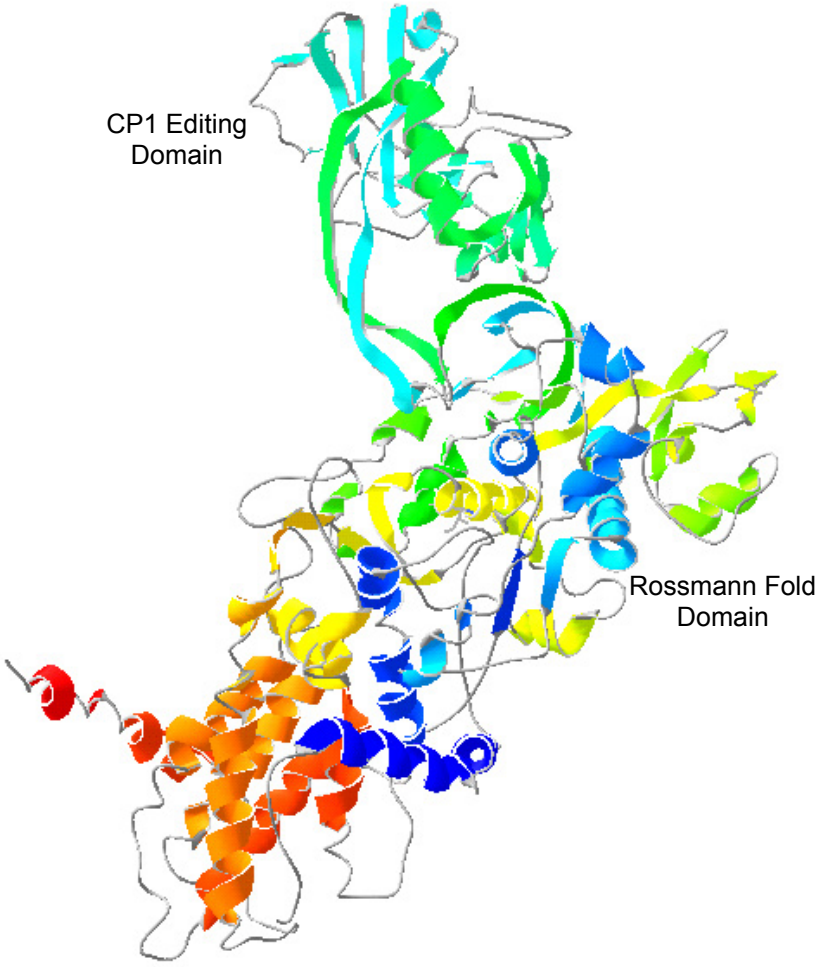

CP1 Editing
Domain
Rossmann Fold
Domain

Figure 1-4

(a) Stereoview of isoleucine in the synthetic site of IleRS *T. thermophilus* (PDB code: 1ILE). The substrate isoleucine is shown in yellow. The surrounding residues are labeled (carbon, nitrogen and oxygen atoms are shown in gray, blue and red respectively). (b) The non-cognate valine moleucule bound to the synthetic active site of IleRS. Presentation is as in (a).

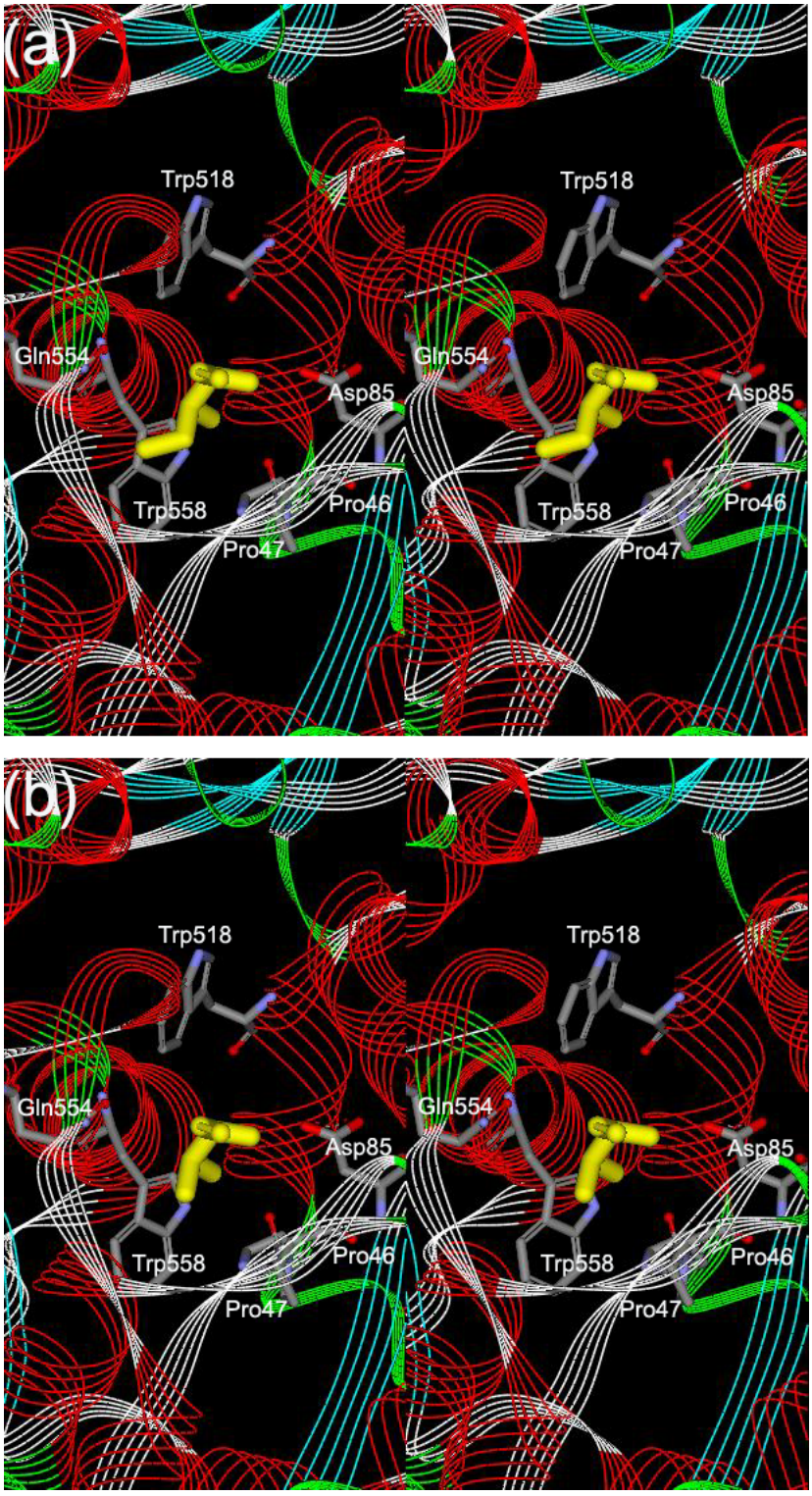

(a)
Trp518
Trp518
Gln554
Gln554
Asp85
Asp85
Trp558
Trp558
Pro46
Pro46
Pro47
Pro47
(b)
Trp518
Trp518
Gln554
Gln554
Asp85
Asp85
Trp558
Trp558
Pro46
Pro46
Pro47
Pro47

Figure 1-5

Stereoview of Ile-AMS (a non-hydrolyzable Ile-AMP analog) bound to the aminoacylation site of IleRS from *T. thermophilus.* The Ile-AMS molecule is shown as thick stick model with different atomic color (carbon, nitrogen, oxygen and sulfur atoms are shown in gray, blue red and yellow respectively). The residues surrounding Ile-AMS are shown in pink and labeled.

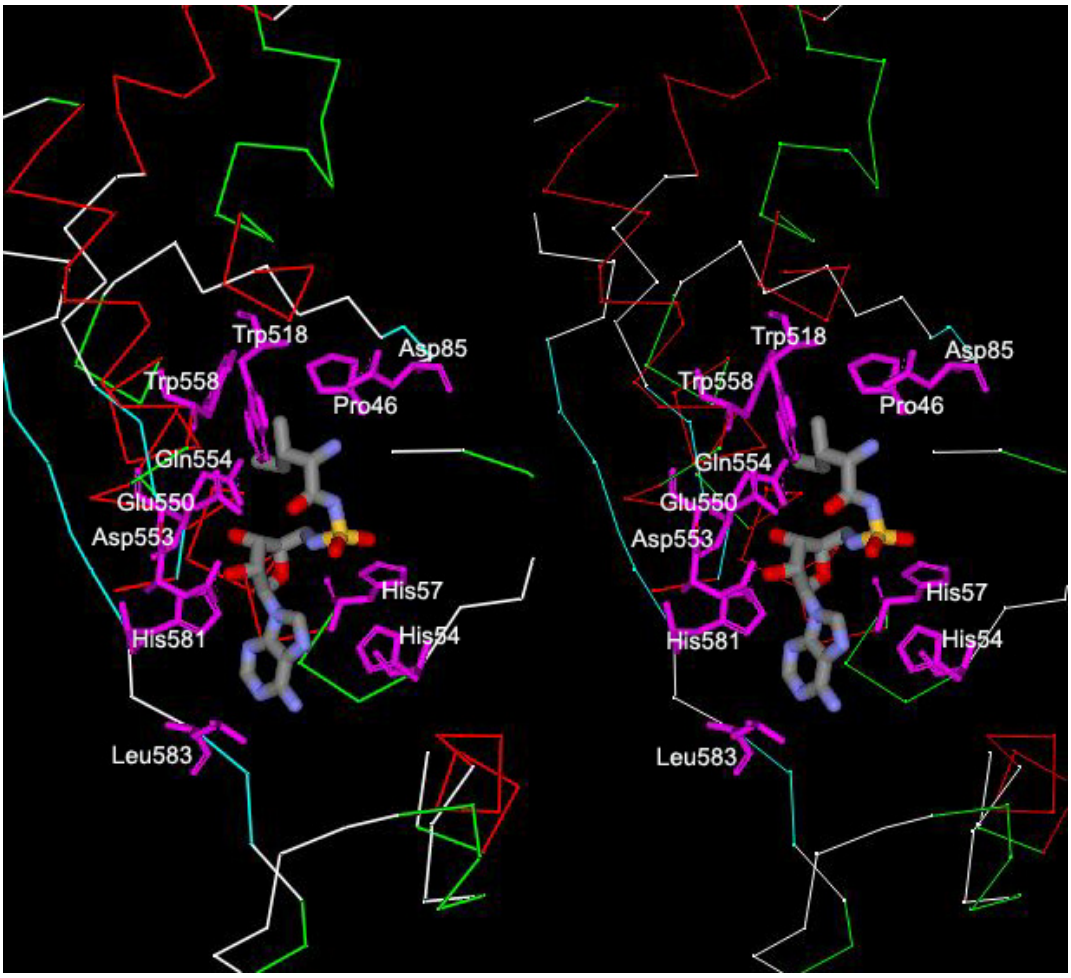

Trp518
Asp85
Trp558
Pro46
Gln554
Glu550
Asp553
His57
His581
His54
Leu583
Trp518
Asp85
Trp558
Pro46
Gln554
Glu550
Asp553
His57
His581
His54
Leu583

Figure 1-6

tRNA bound to IleRS from *S. aureus*. The tRNA is shown in pink and 3' terminal nucleotides are labeled. IleRS is represented by a ribbon model and colored by secondary structure; major domains are labeled. The synthetic and editing sites of IleRS are labeled by arrows. PDB code for this structure is 1FFY.

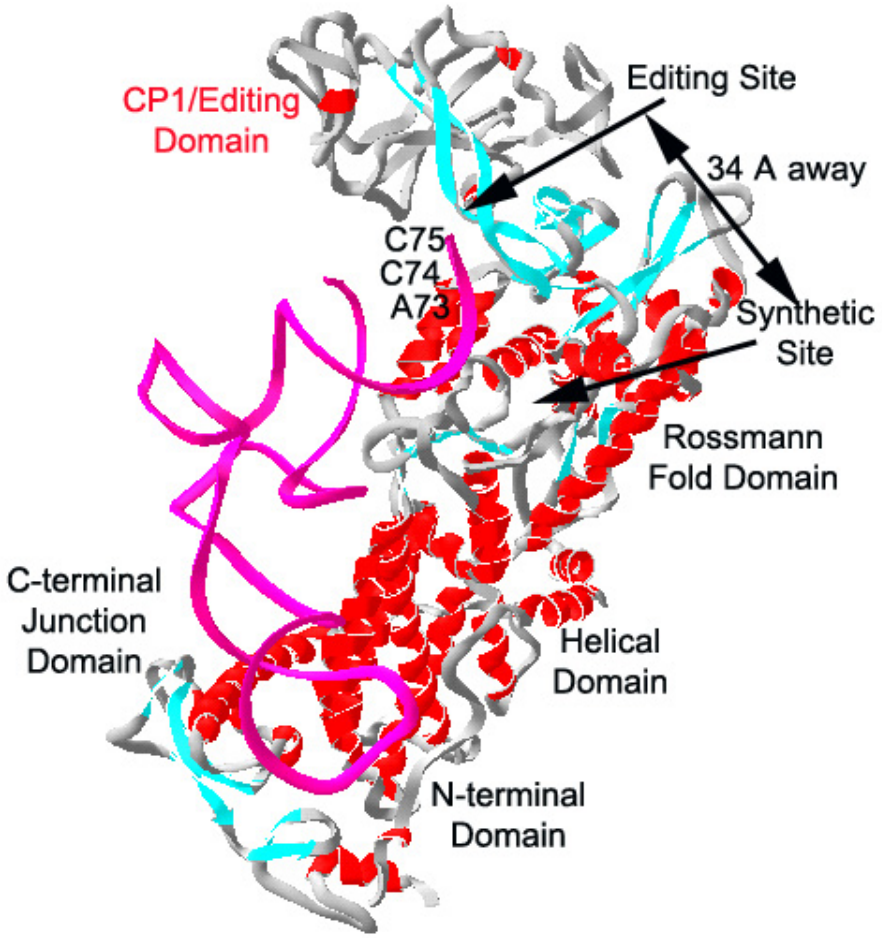

CP1/Editing
Domain
Editing Site
34 A away
C75
C74
A73
Synthetic
Site
Rossmann
Fold Domain
C-terminal
Junction
Domain
Helical
Domain
N-terminal
Domain

Figure 1-7

Crystal structure of ValRS complexed with $tRNA^{Val}$ from *T. thermophilus*. ValRS has 6 domains, including the CP1 editing domain, Rossmann-fold domain, stem-contact (SC) domain, helix bundle domain, anticodon stem binding domain and coiled-coiled domain, as labeled. The bound $tRNA^{Val}$ is shown in pink. The features of the tertiary structure of $tRNA^{Val}$ are labeled.

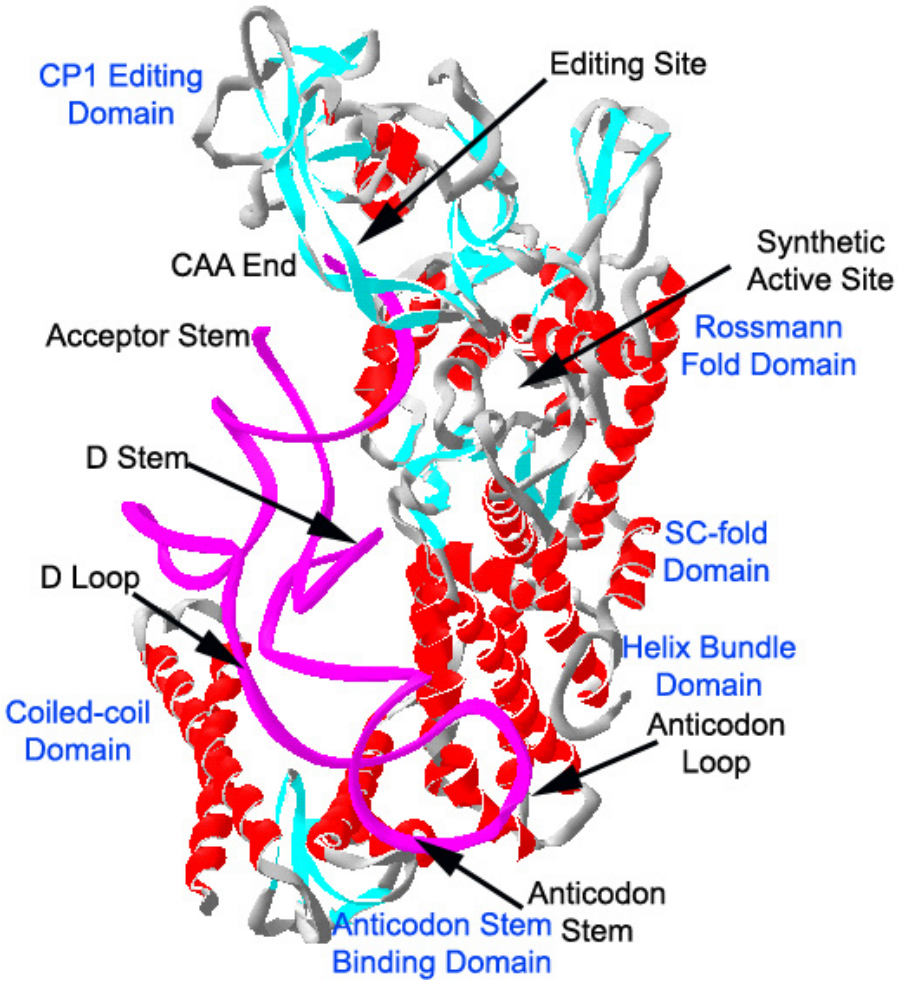

CP1 Editing Domain
Editing Site
CAA End
Synthetic Active Site
Acceptor Stem
Rossmann Fold Domain
D Stem
SC-fold Domain
D Loop
Helix Bundle Domain
Coiled-coil Domain
Anticodon Loop
Anticodon Stem
Anticodon Stem Binding Domain

Figure 1-8

Aminoacyl adenylate analog Val-AMS bound to the active site of ValRS. The substrate is shown as thick stick model (carbon, nitrogen, oxygen and sulfur atoms are shown in gray, blue, red and orange respectively). The surrounding residues that make major interactions with Val-AMS are shown in yellow and labeled.

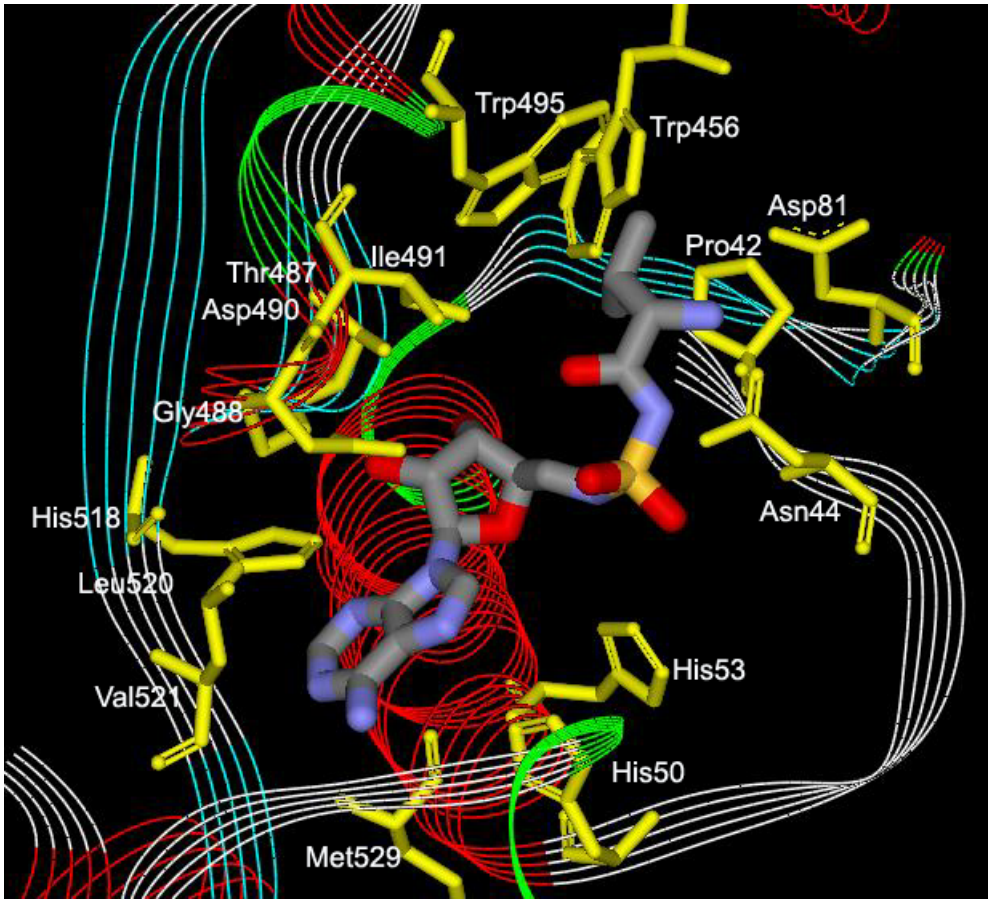

Trp495
Trp456
Asp81
Pro42
Ile491
Thr487
Asp490
Gly488
His518
Asn44
Leu520
His53
Val521
His50
Met529

Figure 1-9

Crystal structure of LeuRS from *T. thermophilus*. The CP1 editing domain is shown in yellow and a unique insertion domain is shown in pink; other domains are labeled. The synthetic active site and the editing site are separated by 37 Å. PDB code for this structure is 1H3N.

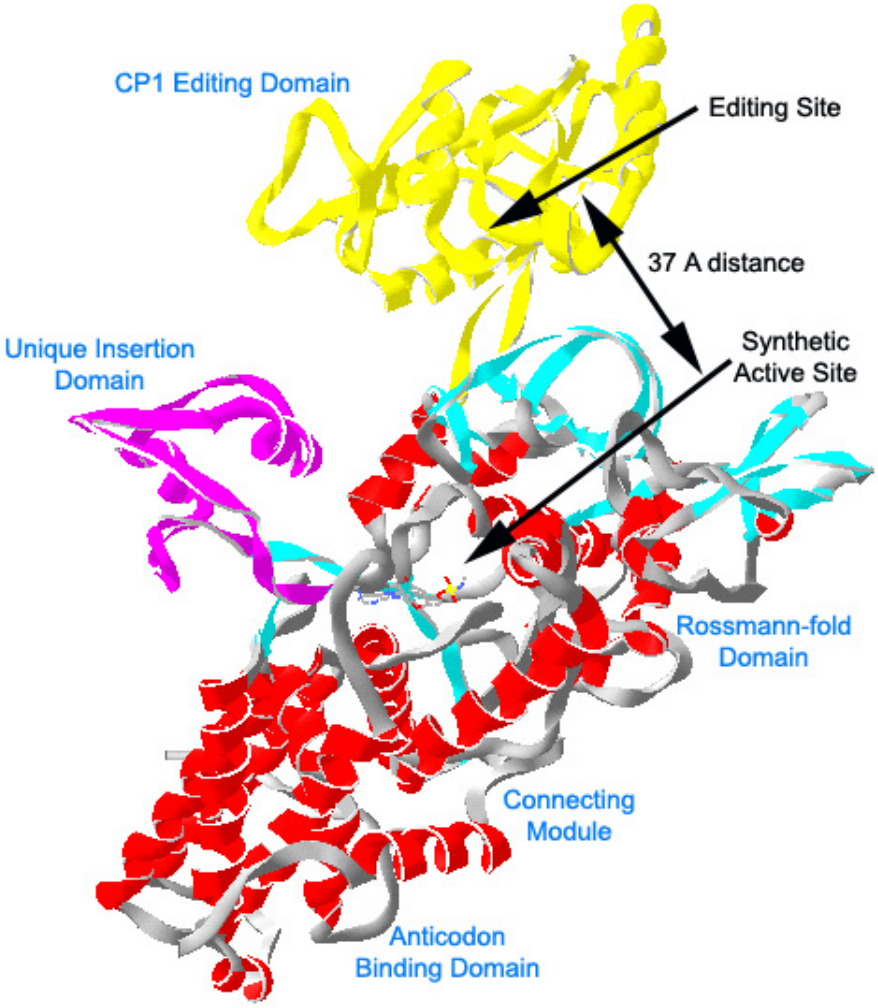

CP1 Editing Domain
Editing Site
37 A distance
Synthetic
Active Site
Unique Insertion
Domain
Rossmann-fold
Domain
Connecting
Module
Anticodon
Binding Domain

Figure 1-10

Active site of LeuRS from *T. thermophilus* complexed with Leu-AMS. Substrate Leu-AMS is shown as stick model; the surrounding residues involved the key interactions with substrate are shown as ball-stick models. A water molecule is shown as a sphere in pink. Color code: carbon, nitrogen, oxygen and sulfur atoms are shown in gray, blue, red and orange respectively. PDB code for this structure is 1H3N.

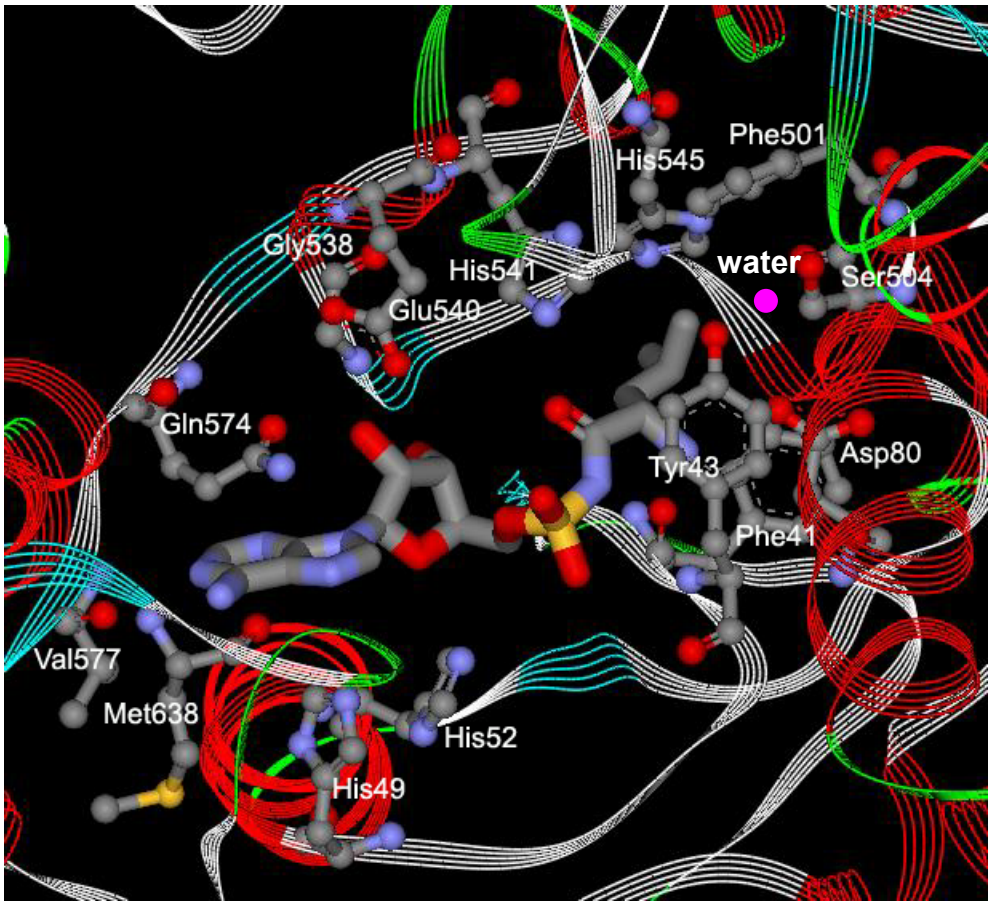

Phe501
His545
Gly538
His541
water
Ser504
Glu540
Gln574
Tyr43
Asp80
Phe41
Val577
Met638
His52
His49

Figure 1-11

(a) Overall crystal structure of PheRS from *T. thermophilus.* The α-subunit is shown in yellow and the β-subunit is shown in blue. (b) The larger view of α-subunit of PheRS. Phenylalanine bound to active site of α-subunit is shown as a space-filling model. PDB code for the structure is 1B70.

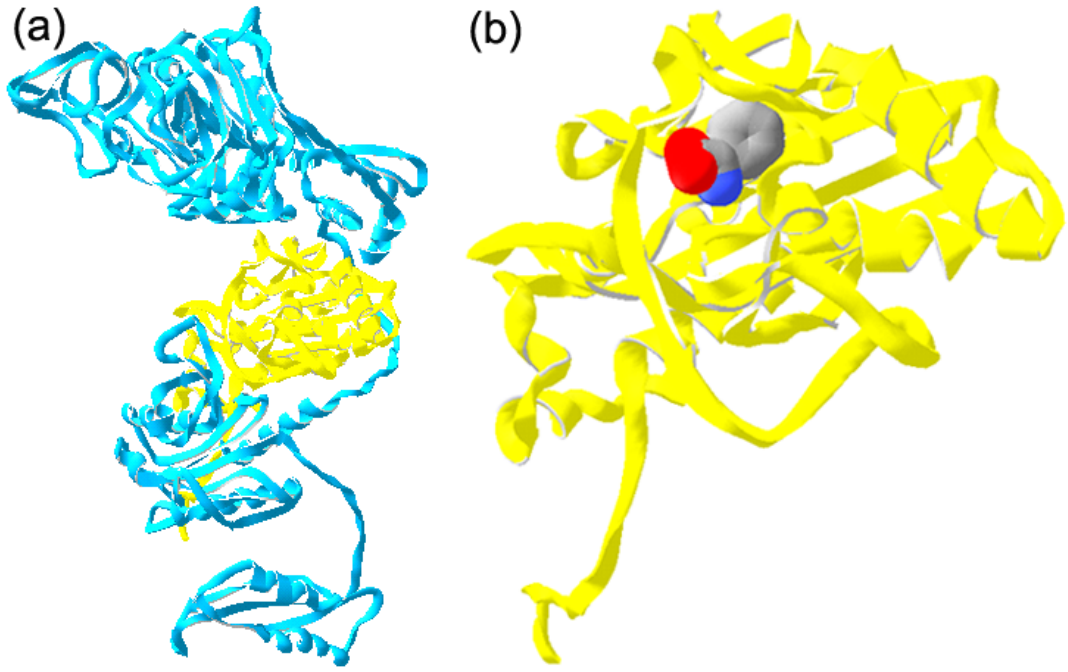

(a)
(b)

Figure 1-12

Active site of PheRS from *T. thermophilus* complexed with PheOH-AMP. Substrate PheOH-AMP is shown as a stick model; the surrounding residues involved the key interactions with substrate are shown as ball-stick model in light yellow. Color code for PheOH-AMP: carbon, nitrogen, oxygen and sulfur atoms are shown in gray, blue, red and orange respectively. PDB code for this structure is 1B7Y.

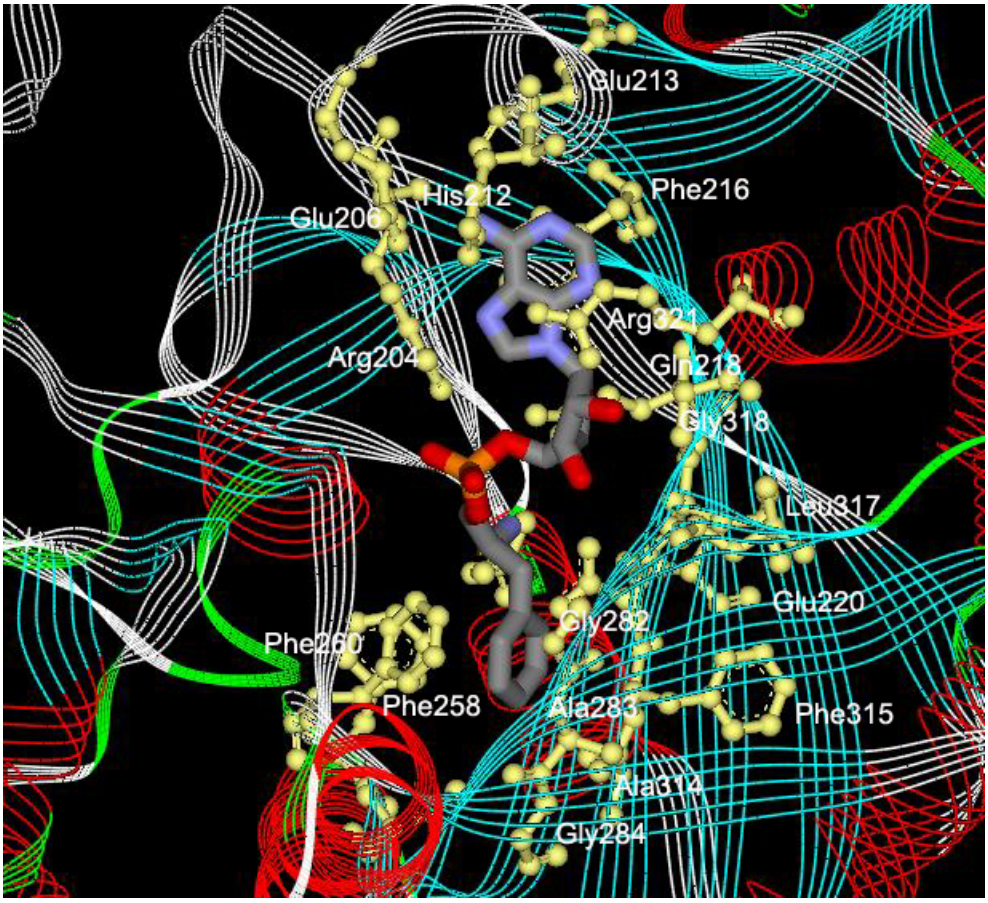
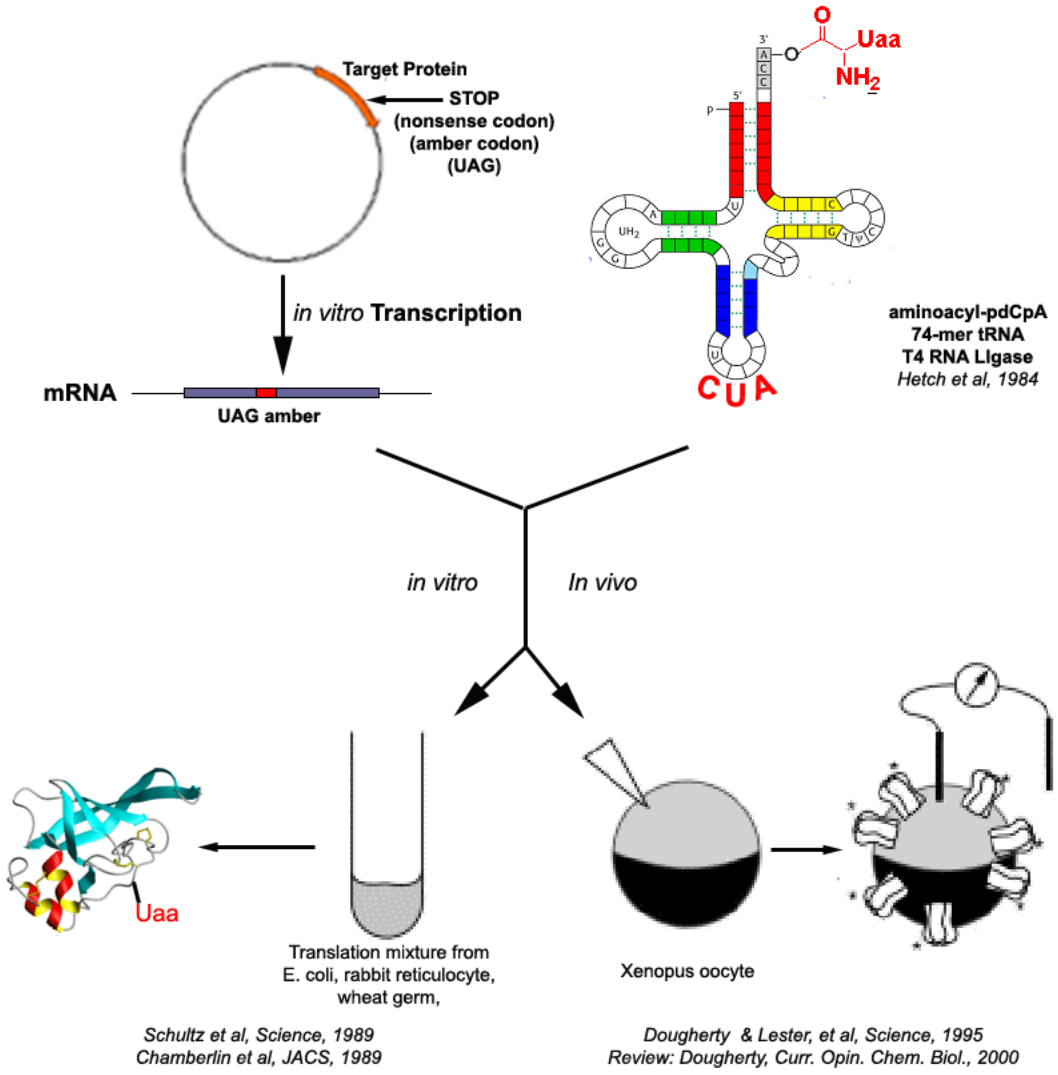

Glu213
His212
Phe216
Glu206
Arg321
Arg204
Gln218
Gly318
Leu317
Glu220
Gly282
Phe260
Phe258
Ala283
Phe315
Ala314
Gly284

Figure 1-13

Site-specific incorporation of non-natural amino acids into proteins through chemically misacylated tRNA and nonsense suppression. This method has been shown both *in vitro* and *in vitro*. The figure is modified from Dougherty's review paper (*Curr. Opin. Chem. Biol*., 2000, **4**, 645-652).


Target Protein
STOP
(nonsense codon)
(amber codon)
(UAG)
in vitro Transcription
mRNA
UAG amber
Uaa
NH2
CUA
aminoacyl-pdCpA
74-mer tRNA
T4 RNA Ligase
Hetch et al, 1984
in vitro
In vivo
Uaa
Translation mixture from
E. coli, rabbit reticulocyte,
wheat germ,
Xenopus oocyte
Schultz et al, Science, 1989
Chamberlin et al, JACS, 1989
Dougherty & Lester, et al, Science, 1995
Review: Dougherty, Curr. Opin. Chem. Biol., 2000

9. References

1. Rich, A., and Schimmel, P. R. (1977) *Accounts Chem. Res. 10*, 385-387.

2. Rich, A., and Schimmel, P. R. (1977) *Nucleic Acids Res. 4*, 1649-1665.

3. Giege, R., Sissler, M., and Florentz, C. (1998) *Nucleic Acids Res. 26*, 5017-5035.

4. Sissler, M., Giege, R., and Florentz, C. (1998) *RNA 4*, 647-657.

5. Sankaranarayanan, R., and Moras, D. (2001) *Acta Biochimica Polonica 48*, 323-335.

6. Pauling, L. (1957) in *Festschrift fur Pr. Dr. Arthur Stoll* pp 597-602, Birkhauser-Verlag, Basel, Switzerland.

7. Loftfield, R. B., and Vanderjagt, M. A. (1972) *Biochem. J. 128*, 1353-1356.

8. Eriani, G., Dealrue, M., Poch, O., Gangloff, J., and Moras, D. (1990) *Nature 347*, 203-206.

9. Cusack, S. (1995) *Nat. Struct. Biol. 2*, 824-831.

10. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) *Nature 250*, 194-199.

11. Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989) *Science 246*, 1135-1142.

12. Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N., and Leberman, R. (1990) *Nature 347*, 249-255.

13. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., and Moras, D. (1991) *Science 252*, 1682-1689.

14. Reshetnikova, L., Moor, N., Lavrik, O., and Vassylyev, D. G. (1999) *J. Mol. Biol. 287*, 555-568.

15. Cavarelli, J., Delagoutte, B., Eriani, G., Gangloff, J., and Moras, D. (1998) *EMBO J. 18*, 5438-5448.

16. Newberry, K. J., Hou, Y. M., and Perona, J. J. (2002) *EMBO J. 21*, 2778-2787.

17. Silvian, L. F., Wang, J. M., and Steitz, T. A. (1999) *Science 285*, 1074-1077.

18. 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*, 578-582.

19. Cusack, S., Yaremchuk, A., and Tukalo, M. (2000) *EMBO J. 19*, 2351-2361.

20. Brunie, S., Zelwer, C., and Risler, J. L. (1990) *J. Mol. Biol. 216*, 411-424.

21. Fourmy, D., Dardel, F., and Blanquet, S. (1993) *J. Mol. Biol. 231*, 1078-1089.

22. Mechulam, Y., Schmitt, E., Maveyraud, L., Zelwer, C., Nureki, O., Yokoyama, S., Konno, M., and Blanquet, S. (1999) *J. Mol. Biol. 294*, 1287-1297.

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

24. Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989) *Science 246*, 1135-1142.

25. Rould, M. A., Perona, J. J., and Steitz, T. A. (1991) *Nature 352*, 213-218.

26. Arnez, J. G., and Steitz, T. A. (1996) *Biochemistry 35*, 14725-14733.

27. Rath, V. L., Silvian, L. F., Beijer, B., Sproat, B. S., and Steitz, T. A. (1998) *Struct. Fold. Des. 6*, 439-449.

28. Nureki, O., Vassylyev, D. G., Katayanagi, K., Shimizu, T., Sekine, S., Kigawa, T., Miyazawa, T., Yokoyama, S., and Morikawa, K. (1995) *Science 267*, 1958-1965.

29. Sekine, S., Nureki, O., Dubois, D. Y., Bernier, S., Chenevert, R., Lapointe, J., Vassylyev, D. G., and Yokoyama, S. (2003) *EMBO J. 22*, 676-688.

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

31. Retailleau, P., Huang, X., Yin, Y. H., Hu, M., Weinreb, V., Vachette, P., Vonrhein, C., Bricogne, G., Roversi, P., Ilyin, V., and Carter, C. W. (2003) *J. Mol. Biol. 325*, 39-63.

32. Bhat, T. N., Blow, D. M., Brick, P., and Nyborg, J. (1982) *J. Mol. Biol. 158*, 699-709.

33. Brick, P., and Blow, D. M. (1987) *J. Mol. Biol. 194*, 287-297.

34. Brown, K. A., Brick, P., and Blow, D. M. (1987) *Nature 326*, 416-418.

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

36. Logan, D. T., Mazauric, M. H., Kern, D., and Moras, D. (1995) *EMBO J. 14*, 4156-4167.

37. Arnez, J. G., Dock-Bregeon, A. C., and Moras, D. (1999) *J. Mol. Biol. 286*, 1449-1459.

38. Arnez, J. G., Harris, D. C., Mitschler, A., Rees, B., Francklyn, C. S., and Moras, D. (1995) *EMBO J. 14*, 4143-4155.

39. Aberg, A., Yaremchuk, A., Tukalo, M., Rasmussen, B., and Cusack, S. (1997) *Biochemistry 36*, 3084-3094.

40. Arnez, J. G., Augustine, J. G., Moras, D., and Francklyn, C. S. (1997) *Proc. Natl. Acad. Sci. USA 94*, 7144-7149.

41. Qiu, X. Y., Janson, C. A., Blackburn, M. N., Chhohan, I. K., Hibbs, M., and Abdel-Meguid, S. S. (1999) *Biochemistry 38*, 12296-12304.

42. Cusack, S., Yaremchuk, A., Krikliviy, I., and Tukalo, M. (1998) *Structure 6*, 101-108.

43. Sankaranarayanan, R., Dock-Bregeon, A. C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B., and Moras, D. (1999) *Cell 97*, 371-381.

44. Belrhali, H., Yaremchuk, A., Tukalo, M., Larsen, K., Berthetcolominas, C., Leberman, R., Beijer, B., Sproat, B., Alsnielsen, J., Grubel, G., Legrand, J. F., Lehmann, M., and Cusack, S. (1994) *Science 263*, 1432-1436.

45. Belrhali, H., Yaremchuk, A., Tukalo, M., Berthetcolominas, C., Rasmussen, B., Bosecke, P., Diat, O., and Cusack, S. (1995) *Structure 3*, 341-352.

46. Cusack, S., Yaremchuk, A., and Tukalo, M. (1996) *EMBO J. 15*, 2834-2842.

47. Berthet-Colominas, C., Seignovert, L., Hartlein, M., Grotli, M., Cusack, S., and Leberman, R. (1998) *EMBO J. 17*, 2947-2960.

48. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., and Moras, D. (1991) *Science 252*, 1682-1689.

49. Cavarelli, J., Eriani, G., Rees, B., Ruff, M., Boeglin, M., Mitschler, A., Martin, F., Gangloff, J., Thierry, J. C., and Moras, D. (1994) *EMBO J. 13*, 327-337.

50. Delarue, M., Poterszman, A., Nikonov, S., Garber, M., Moras, D., and Thierry, J. C. (1994) *EMBO J. 13*, 3219-3229.

51. Schmitt, E., Moulinier, L., Fujiwara, S., Imanaka, T., Thierry, J. C., and Moras, D. (1998) *EMBO J. 17*, 5227-5237.

52. Eiler, S., Dock-Bregeon, A. C., Moulinier, L., Thierry, J. C., and Moras, D. (1999) *EMBO J. 18*, 6532-6541.

53. Commans, S., Plateau, P., Blanquet, S., and Dardel, F. (1995) *J. Mol. Biol. 253*, 100-113.

54. Onesti, S., Miller, A. D., and Brick, P. (1995) *Structure 3*, 163-176.

55. Cusack, S., Yaremchuk, A., and Tukalo, M. (1996) *EMBO J. 15*, 6321-6334.

56. Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M., and Safro, M. G. (1995) *Nat. Struct. Biol. 2*, 537-547.

57. Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S., and Safro, M. (1997) *Structure 5*, 59-68.

58. Ibba, M., and Soll, D. (2000) *Annu. Rev. Biochem. 69*, 617-650.

59. Dealrue, M., and Moras, D. (1993) *Bioassays 15*, 675-687.

60. Sankaranarayanan, R., and Moras, D. (1999) in *Perspectives in Structure Biology (A volume in honour of Prof. G. N. Ramachandran)* (M. Vijayan, N. Y., and Kolaskar, A. S., Eds.) pp 128-138, Universities Press (India) Limited, Hyderabad.

61. Schmitt, E., Meinnel, T., Blanquet, S., and Mechulam, Y. (1994) *J. Mol. Biol. 242*, 566-577.

62. First, E. A., and Fersht, A. R. (1995) *Biochemistry 34*, 5030-5043.

63. Hong, K. W., Ibba, M., WeygandDurasevic, I., Rogers, M. J., Thomann, H. U., and Soll, D. (1996) *EMBO J. 15*, 1983-1991.

64. Ibba, M., Hong, K. W., Sherman, J. M., Sever, S., and Soll, D. (1996) *Proc. Natl. Acad. Sci. USA 93*, 6953-6958.

65. Cusack, S. (1997) *Curr. Opin. Struct. Biol. 7*, 881-889.

66. Arnez, J. G., and Moras, D. (1997) *Trends Biochem.Sci. 22*, 211-216.

67. Arnez, J. G., and Moras, D. (1998) in *RNA structure and Function* (Simons, R. W., and Grunberg-Manago, Eds.) pp 469-94, Cold Spring Harbor Lab, New York.

68. Szymanski, M., Deniziak, M. A., and Barciszewski, J. (2001) *Nucleic Acids Res. 29*, 288-290.

69. Nakama, T., Nureki, O., and Yokoyama, S. (2001) *J. Biol. Chem. 276*, 47387-47393.

70. Freist, W., Pardowitz, I., and Cramer, F. (1985) *Biochemistry 24*, 7014-7023.

71. Starzyk, R. M., Webster, T. A., and Schimmel, P. (1987) *Science 237*, 1614-1618.

72. Schmidt, E., and Schimmel, P. (1995) *Biochemistry 34*, 11204-11210.

73. Lin, L., Hale, S. P., and Schimmel, P. (1996) *Nature 384*, 33-34.

74. Fersht, A. R. (1977) *Biochemistry 16*, 1025-1030.

75. Fersht, A. R., and Dingwall, C. (1979) *Biochemistry 18*, 2627-2631.

76. Fersht, A. R. (1985) *Enzyme structure and mechanism*, Freeman, New York.

77. Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C., and Klug, A. (1974) *Nature 250*, 546-551.

78. Suddath, F. L., Quigley, G. J., McPherso.A, Sneden, D., Kim, J. J., Kim, S. H., and Rich, A. (1974) *Nature 248*, 20-24.

79. Biou, V., Yaremchuk, A., Tukalo, M., and Cusack, S. (1994) *Science 263*, 1404-1410.

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

81. Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., and Steitz, T. A. (1988) *Proc. Natl. Acad. Sci. USA 85*, 8924-8928.

82. Steitz, T. A., Friedman, J. M., Beese, L. S., and Freemont, P. S. (1988) *Faseb J. 2*, A1855-A1855.

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

84. Joyce, C. M., and Steitz, T. A. (1994) *Annu. Rev. Biochem. 63*, 777-822.

85. Nomanbhoy, T. K., Hendrickson, T. L., and Schimmel, P. (1999) *Mol. Cell 4*, 519-528.

86. Eldred, E. W., and Schimmel, P. R. (1972) *J. Biol. Chem. 247*, 2961-2964.

87. Hale, S. P., and Schimmel, P. R. (1996) *Proc. Natl. Acad. Sci. USA 93*, 2755-2758.

88. Jakubowski, H., and Fersht, A. R. (1981) *Nucleic Acids Res. 9*, 3105-3117.

89. Jakubowski, H. (1980) *Biochemistry 19*, 5071-5078.

90. Igloi, G. L., von der Haar, F., and Cramer, F. (1977) *Biochemistry 16*, 1696-1702.

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

92. Nangle, L. A., de Crecy-Lagard, V., Doring, V., and Schimmel, P. (2002) *J. Biol. Chem. 277*, 45729-45733.

93. Sugiura, I. e. a. (2000) *Structure 8*, 197-208.

94. Englisch, S., Englisch, U., von der Haar, F., and Cramer, F. (1986) *Nucleic Acids Res. 14*, 7529-7539.

95. Lincecum, T. L. J., Tukalo, M., Yaremchuk, A., Mursinna, R. S., Williams, A. M., Sproat, B., Van Den Eynde, W., Link, A., Van Calenbergh, S., Grotli, M., Martinis, S. A., and Cusack, S. (2003) *Mol. Cell 11*, 951-963.

96. Hendrickson, T. L., Nomanbhoy, T. K., de Crecy-Lagard, V., Fukai, S., Nureki, O., Yokoyama, S., and Schimmel, P. (2002) *Mol. Cell 9*, 352-362.

97. Tang, Y., and Tirrell, D. A. (2002) *Biochemistry 41*, 10635-10645.

98. Chen, J. F., Guo, N. N., Li, T., Wang, E. D., and Wang, Y. L. (2000) *Biochemistry 39*, 6726-6731.

99. Martinis, S. A., and Fox, G. E. (1997) *Nucleic Acids Symp. Ser. 36*, 125-128.

100. Mursinna, R. S., Lincecum, T. L., and Martinis, S. A. (2001) *Biochemistry 40*, 5376-5381.

101. Apostol, I., Levine, J., Lippincott, J., Leach, J., Hess, E., Glascock, C. B., Weickert, M. J., and Blackmore, R. (1997) *J. Biol. Chem. 272*, 28980-28988.

102. Mursinna, R. S., and Martinis, S. A. (2002) *J. Am. Chem. Soc. 124*.

103. Hennecke, H., Springer, M., and Bock, A. (1977) *Mol. Genet. 152*, 205-210.

104. Khrodyreva, S. N., Moor, N. A., Ankilova, V. N., and Lavrik, O. I. (1985) *Biochim. Biophys. Acta. 830*, 206-212.

105. Santi, D. V., and Danenberg, P. V. (1971) *Biochemistry 10*, 4813-4820.

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

107. Ibba, M., Kast, P., and Hennecke, H. (1994) *Biochemistry 33*, 7107-7112.

108. Ibba, M., and Hennecke, H. (1995) *FEBS Lett. 364*, 272-275.

109. Brannigan, J. A., and Wilkinson, A. J. (2002) *Nature Rev. Mol. Cell Biol. 3*, 964-970.

110. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989) *Science 244*, 182-188.

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

112. Anthony-Cahill, S. J., Griffith, M. C., Noren, C. J., Suich, D. J., and Schultz, P. G. (1989) *Trends Biochem. Sci. 14*, 400-403.

113. Noren, C. J., Anthonycahill, S. J., Suich, D. J., Noren, K. A., Griffith, M. C., and Schultz, P. G. (1990) *Nucleic Acids Res. 18*, 83-88.

114. Bain, J. D., Diala, E. S., Glabe, C. G., Wacker, D. A., Lyttle, M. H., Dix, T. A., and Chamberlin, A. R. (1991) *Biochemistry 30*, 5411-5421.

115. Bain, J. D., Wacker, D. A., Kuo, E. E., and Chamberlin, A. R. (1991) *Tetrahedron 47*, 2389-2400.

116. Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C. J., and Schultz, P. G. (1991) *Method Enzymol. 202*, 301-336.

117. Mendel, D., Ellman, J. A., and Schultz, P. G. (1991) *J. Am. Chem. Soc. 113*, 2758-2760.

118. Bain, J. D., Switzer, C., Chamberlin, A. R., and Benner, S. A. (1992) *Nature 356*, 537-539.

119. Ellman, J. A., Mendel, D., and Schultz, P. G. (1992) *Science 255*, 197-200.

120. Mendel, D., Ellman, J. A., Chang, Z. Y., Veenstra, D. L., Kollman, P. A., and Schultz, P. G. (1992) *Science 256*, 1798-1802.

121. Chung, H. H., Benson, D. R., and Schultz, P. G. (1993) *Science 259*, 806-809.

122. Cornish, V. W., and Schultz, P. G. (1994) *Curr. Opin. Struct. Biol. 4*, 601-607.

123. Cornish, V. W., Kaplan, M. I., Veenstra, D. L., Kollman, P. A., and Schultz, P. G. (1994) *Biochemistry 33*, 12022-12031.

124. Cornish, V. W., Benson, D. R., Altenbach, C. A., Hideg, K., Hubbell, W. L., and Schultz, P. G. (1994) *Proc. Natl. Acad. Sci. USA 91*, 2910-2914.

125. Cornish, V. W., Mendel, D., and Schultz, P. G. (1995) *Angew. Chem., Int. Edit. Engl. 34*, 621-633.

126. Mendel, D., Cornish, V. W., and Schultz, P. G. (1995) *Annu. Rev. Biophys. Biomolec. Struct. 24*, 435-462.

127. Nowak, M. W., Kearney, P. C., Silverman, S. K., Sampson, J. R., Saks, M. E., Labarca, C. G., Zhong, W., Thorson, I., Chapman, E., Abelson, J. N., Davidson, N., Schultz, P. G., Dougherty, D. A., and Lester, H. A. (1995) *Abstr. Pap. Am. Chem. Soc. 209*, 132-ORGN.

128. Nowak, M. W., Kearney, P. C., Sampson, J. R., Saks, M. E., Labarca, C. G., Silverman, S. K., Zhong, W., Thorson, J., Abelson, J. N., Davidson, N., Schultz, P. G., Dougherty, D. A., and Lester, H. A. (1995) *Science 268*, 439-442.

129. Carlson, J. E., Steward, L. E., Rusinova, R., Chamberlin, A. R., and Ross, J. B. A. (1996) *Biophys. J. 70*, MP330-MP330.

130. Cornish, V. W., Hahn, K. M., and Schultz, P. G. (1996) *J. Am. Chem. Soc. 118*, 8150-8151.

131. Kearney, P. C., Zhang, H. Y., Zhong, W., Dougherty, D. A., and Lester, H. A. (1996) *Neuron 17*, 1221-1229.

132. Kearney, P. C., Nowak, M. W., Zhong, W., Silverman, S. K., Lester, H. A., and Dougherty, D. A. (1996) *Mol. Pharmacol. 50*, 1401-1412.

133. Nowak, M. W., Kearney, P. C., Zhong, W., Silverman, S. K., Labarca, C. G., Saks, M. E., Sampson, J. R., Abelson, J., Davidson, N., Dougherty, D. A., and Lester, H. A. (1996) *Biophys. J. 70*, SU253-SU253.

134. England, P. M., Lester, H. A., Davidson, N., and Dougherty, D. A. (1997) *Proc. Natl. Acad. Sci. USA 94*, 11025-11030.

135. England, P. M., Dougherty, D. A., and Lester, H. A. (1997) *Biophys. J. 72*, TUAM5-TUAM5.

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

137. Park, Y. W., Luo, J. K., Schultz, P. G., and Kirsch, J. F. (1997) *Biochemistry 36*, 10517-10525.

138. Steward, L. E., Collins, C. S., Gilmore, M. A., Carlson, J. E., Ross, J. B. A., and Chamberlin, A. R. (1997) *J. Am. Chem. Soc. 119*, 6-11.

139. England, P. M., Brandt, G. S., Dougherty, D. A., and Lester, H. A. (1998) *Biophys. J. 74*, A252-A252.

140. Miller, J. C., Silverman, S. K., England, P. M., Dougherty, D. A., and Lester, H. A. (1998) *Biophys. J. 74*, A251-A251.

141. Miller, J. C., Silverman, S. K., England, P. M., Dougherty, D. A., and Lester, H. A. (1998) *Neuron 20*, 619-624.

142. Nowak, M. W., Gallivan, J. P., Silverman, S. K., Labarca, C. G., Dougherty, D. A., and Lester, H. A. (1998) in *Ion Channels, Pt B* pp 504-529.

143. England, P. M., Zhang, Y. N., Dougherty, D. A., and Lester, H. A. (1999) *Cell 96*, 89-98.

144. Gilmore, M. A., Steward, L. E., and Chamberlin, A. R. (1999) in *Implementation and Redesign of Catalytic Function in Biopolymers* pp 77-99.

145. Dougherty, D. A. (2000) *Curr. Opin. Chem. Biol. 4*, 645-652.

146. Monahan, S. L., Dougherty, D. A., and Lester, H. A. (2001) *Biophys. J. 80*, 564A-564A.

147. Philipson, K. D., Gallivan, J. P., Brandt, G. S., Dougherty, D. A., and Lester, H. A. (2001) *Am. J. Physiol., Cell Physiol. 281*, C195-C206.

148. Tong, Y. H., Brandt, G. S., Li, M., Shapovalov, G., Slimko, E., Karschin, A., Dougherty, D. A., and Lester, H. A. (2001) *J. Gen. Physiol. 117*, 103-118.

149. Petersson, E. J., Brandt, G. S., Zacharias, N. M., Dougherty, D. A., and Lester, H. A. (2003) in *Biophotonics, Pt A* pp 258-273.

150. Heckler, T. G., Chang. L, H., Zama, Y., Naka, T., Chorghade, M. S., and Hecht, S. M. (1984) *Biochemistry 23*, 1468-1476.

151. Cload, S. T., Liu, D. R., Forland, W. A., and Schultz, P. G. (1998) *Chem. Biol. 3*, 1033-1038.

152. Schultz, P. G. (1995) *FASEB J. 9*, A1261-A1261.

153. Arslan, T., Mamaev, S. V., Mamaeva, N. V., and Hecht, S. M. (1997) *J. Am. Chem. Soc. 119*, 10877-10887.

154. Karginov, V. A., Mamaev, S. V., and Hecht, S. M. (1997) *Nucleic Acids Res. 25*, 3912-3916.

155. Karginov, V. A., Mamaev, S. V., An, H. Y., VanCleve, M. D., Hecht, S. M., Komatsoulis, G. A., and Abelson, J. N. (1997) *J. Am. Chem. Soc. 119*, 8166-8176.

156. Killian, J. A., Van Cleve, M. D., Shayo, Y. F., and Hecht, S. M. (1998) *J. Am. Chem. Soc. 120*, 3032-3042.

157. Ellman, J. A., Volkman, B. F., Mendel, D., Schultz, P. G., and Wemmer, D. E. (1992) *J. Am. Chem. Soc. 114*, 7959-7961.

158. Sonar, S., Lee, C. P., Coleman, M., Patel, N., Liu, X. M., Marti, T., Khorana, H. G., Rajbhandary, U. L., and Rothschild, K. J. (1994) *Nat. Struct. Biol. 1*, 512-517.

159. Liu, X. M., Sonar, S., Lee, C. P., Coleman, M., Rajbhandary, U. L., and Rothschild, K. J. (1995) *Biophys. Chem. 56*, 63-70.

160. Ludlam, C. F. C., Sonar, S., Lee, C. P., Coleman, M., Herzfeld, J., Rajbhandary, U. L., and Rothschild, K. J. (1995) *Biochemistry 34*, 2-6.

161. Sonar, S., Liu, X. M., Lee, C. P., Coleman, M., He, Y. W., Pelletier, S., Herzfeld, J., Rajbhandary, U. L., and Rothschild, K. J. (1995) *J. Am. Chem. Soc. 117*, 11614-11615.

162. Turcatti, G., Nemeth, K., Edgerton, M. D., Meseth, U., Talabot, F., Peitsch, M., Knowles, J., Vogel, H., and Chollet, A. (1996) *J. Biol. Chem. 271*, 19991-19998.

163. Kearney, P. C., Zhong, W., Nowak, M. W., Silverman, S. K., Zhang, H., Labarca, C., Saks, M. E., Sampson, J. R., Gallivan, J. P., Abelson, J., Davidson, N., Dougherty, D. A., and Lester, H. A. (1996) *Biophys. J. 70*, TUAM5-TUAM5.

164. Saks, M. E., Sampson, J. R., Nowak, M. W., Kearney, P. C., Du, F. Y., Abelson, J. N., Lester, H. A., and Dougherty, D. A. (1996) *J. Biol. Chem. 271*, 23169-23175.

165. England, P. M., Lester, H. A., and Dougherty, D. A. (1999) *Tetrahedron Lett. 40*, 6189-6192.

166. Dang, H., England, P. M., Farivar, S. S., Dougherty, D. A., and Lester, H. A. (2000) *Mol. Pharmacol. 57*, 1114-1122.

167. Brandt, G. S., Zhong, W., Zacharias, N. M., Gibbs, C. M., Dougherty, D. A., and Lester, H. A. (2001) *Biophys. J. 80*, 103A-104A.

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

169. Petersson, E. J., Lester, H. A., and Dougherty, D. A. (2002) *Biophys. J. 82*, 212A-213A.

170. Zhong, W., Gallivan, J. P., Zhang, H., Li, L. T., Lester, H. A., and Dougherty, D. A. (1998) *Proc. Natl. Acad. Sci. USA 95*, 12088-12093.

171. Dougherty, D. A. (1996) *Science 271*, 163-168.

172. Ma, J. C., and Dougherty, D. A. (1997) *Chem. Rev. 97*, 1303-1324.

173. Taki, M., Hohsaka, T., Murakami, H., Taira, K., and Sisido, M. (2002) *J. Am. Chem. Soc. 124*, 14586-14590.

174. Hohsaka, T., and Sisido, M. (2002) *Curr. Opin. Chem. Biol. 6*, 809-815.

175. Sisido, M., and Hohsaka, T. (2001) *Appl. Microbiol. Biotechnol. 57*, 274-281.

176. Hohsaka, T., Ashizuka, Y., Murakami, H., and Sisido, M. (2001) *Nucleic Acids Res. 29*, 3646-3651.

177. Hohsaka, T., Ashizuka, Y., Taira, H., Murakami, H., and Sisido, M. (2001) *Biochemistry 40*, 11060-11064.

178. Murakami, H., Hohsaka, T., Ashizuka, Y., Hashimoto, K., and Sisido, M. (2000) *Biomacromolecules 1*, 118-125.

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

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

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

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

183. Ma, C. H., Kudlicki, W., Odom, O. W., Kramer, G., and Hardesty, B. (1993) *Biochemistry 32*, 7939-7945.

184. Ohama, T., Muto, A., and Osawa, S. (1990) *Nucleic Acids Res. 18*, 1565-1569.

185. Kano, A., Andachi, Y., Ohama, T., and Osawa, S. (1991) *J. Mol. Biol. 221*, 387-401.

186. Kano, A., Ohama, T., Abe, R., and Osawa, S. (1993) *J. Mol. Biol. 230*, 51-56.

187. Kowal, A. K., and Oliver, J. S. (1997) *Nucleic Acids Res. 25*, 4685-4689.

188. 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*, 177-182.

189. Ohtsuki, T., Kimoto, M., Ishikawa, M., Mitsui, T., Hirao, I., and Yokoyama, S. (2001) *Proc. Natl. Acad. Sci. USA 98*, 4922-4925.

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

191. Liu, D. R., Magliery, T. J., and Schultz, P. G. (1997) *Chem. Biol. 4*, 685-691.

192. Liu, D. R., Magliery, T. J., Pasternak, M., and Schultz, P. G. (1997) *Proc. Natl. Acad. Sci. USA 94*, 10092-10097.

193. Liu, D. R., and Schultz, P. G. (1999) *Proc. Natl. Acad. Sci. USA 96*, 4780-4785.

194. Wang, L., Magliery, T. J., Liu, D. R., and Schultz, P. G. (2000) *J. Am. Chem. Soc. 122*, 5010-5011.

195. Sampson, J. R., Direnzo, A. B., Behlen, L. S., and Uhlenbeck, O. C. (1989) *Science 243*, 1363-1366.

196. Soma, A., Uchiyama, K., Sakamoto, T., Maeda, M., and Himeno, H. (1999) *J. Mol. Biol. 293*, 1029-1038.

197. Roe, B., Sirover, M., and Dudock, B. (1973) *Biochemistry 12*, 4146-4154.

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

199. Chin, J. W., and Schultz, P. G. (2002) *ChemBioChem 3*, 1135-1137.

200. Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) *J. Am. Chem. Soc. 124*, 9026-9027.

201. Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) *Proc. Natl. Acad. Sci. USA 99*, 11020-11024.

202. Santoro, S. W., Wang, L., Herberich, B., King, D. S., and Schultz, P. G. (2002) *Nat. Biotechnol. 20*, 1044-1048.

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

204. Wang, L., and Schultz, P. G. (2002) *Chem. Commun.*, 1-11.

205. Zhang, Z. W., Wang, L., Brock, A., and Schultz, P. G. (2002) *Angew. Chem., Int. Edit. 41*, 2840-+.

206. Wang, L., Zhang, Z. W., Brock, A., and Schultz, P. G. (2003) *Proc. Natl. Acad. Sci. USA 100*, 56-61.

207. Wang, L., Xie, J. M., Deniz, A. A., and Schultz, P. G. (2003) *J. Org. Chem. 68*, 174-176.

208. Wang, L., and Schultz, P. G. (2001) *Chem. Biol. 8*, 883-890.

209. Krejchi, M. T., Atkins, E. D. T., Waddon, A. J., Fournier, M. J., Mason, T. L., and Tirrell, D. A. (1994) *Science 265*, 1427-1432.

210. Parkhe, A. D., Fournier, M. J., Mason, T. L., and Tirrell, D. A. (1994) *Abstr. Pap. Am. Chem. Soc. 207*, 166-IEC.

211. Tirrell, D. A. (1994) *J. Macromol. Sci.-Pure Appl. Chem. A31*, 1023-1028.

212. Deming, T. J., Fournier, M. J., Mason, T. L., and Tirrell, D. A. (1996) *Macromolecules 29*, 1442-1444.

213. Krejchi, M. T., Cooper, S. J., Deguchi, Y., Atkins, E. D. T., Fournier, M. J., Mason, T. L., and Tirrell, D. A. (1997) *Macromolecules 30*, 5012-5024.

214. He, S. J., Lee, C., Gido, S. P., Yu, S. J. M., and Tirrell, D. A. (1998) *Macromolecules 31*, 9387-9389.

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

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

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

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

219. Yu, S. J. M., Soto, C. M., and Tirrell, D. A. (2000) *J. Am. Chem. Soc. 122*, 6552-6559.

220. Yu, S. J. M., and Tirrell, D. A. (2000) *Biomacromolecules 1*, 310-312.

221. Di Zio, K., and Tirrell, D. A. (2003) *Macromolecules 36*, 1553-1558.

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

223. Deming, T. J., Fournier, M. J., Mason, T. L., and Tirrell, D. A. (1997) *J. Macromol. Sci.-Pure Appl. Chem. A34*, 2143-2150.

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

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

226. Carrico, I. S., and Tirrell, D. A. (2002) *Unpublished results*.

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

228. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F., and Tirrell, D. A. (2001) *Angew. Chem., Int. Edit. 40*, 1494-+.

229. Dessipri, E., Tirrell, D. A., and Atkins, E. D. T. (1996) *Macromolecules 29*, 3545-3551.

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

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

232. Kiick, K. L., van Hest, J. C. M., and Tirrell, D. A. (2000) *Angew. Chem., Int. Edit. 39*, 2148-+.

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

234. Tang, Y., and Tirrell, D. A. (2001) *J. Am. Chem. Soc. 123*, 11089-11090.

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

236. Kirshenbaum, K., Carrico, I. S., and Tirrell, D. A. (2002) *ChemBioChem 3*, 235-237.

237. Carrico, I. S., Mock, M. L., Sharma, N., Nielson, K., and Tirrell, D. A. (2003) *Manuscript in preparation*.

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