

Expanding the Natural Repertoire of Protein Structure and Function

Spencer J. Anthony-Cahill* and Thomas J. Magliery†

*Department of Chemistry, Western Washington University, Bellingham, WA 98225-9150 and †Yale University, Molecular Biophysics & Biochemistry, New Haven, CT 06520-8114, USA

Abstract: This review considers chemical and genetic approaches to the modification of protein structure. The historical interest in chemical and site-directed modifications will be briefly covered. Current chemical modification strategies will be presented. Biosynthetic mutagenesis with unnatural aminoacyl-tRNAs and current synthetic peptide ligation technologies will be covered in greater detail. The application of combinatorial genetic methods (e.g. phage display, DNA shuffling) to protein engineering with unnatural amino acids will be briefly discussed, with emphasis on the *in vitro* evolution of new enzymatic function (i.e. aminoacyl-tRNA synthetases). Throughout the review, the powerful insights gained from the combined use of these technologies will be illustrated by examples that focus on the elucidation of protein-ligand interactions.

I. PROTEIN ENGINEERING AS A TOOL FOR UNDERSTANDING PROTEIN FUNCTION

Much effort in the biotechnology industry today is directed at engineering proteins with desirable properties for the commercialization of protein-based therapeutics or diagnostics. In a broad sense "protein engineering" refers to the intentional design and subsequent synthesis of proteins with desired novel properties. In the more commonly used and more limited sense, "protein engineering" refers to that field of protein research that has developed in the past twenty years as a result of advances in protein structure determination, computer graphics, macromolecular modeling, and recombinant DNA technology [1-6]. Protein engineering can be used not only to obtain proteins with "designed" functional properties (e.g. enhanced activity, specificity, or stability), but to map the functional portions of existing proteins as well [7, 8]. The principles that emerge from protein engineering research relate *structure to function* and are applicable not only to the tailoring of existing proteins, but also in drug design [9] and the design of proteins *de novo* [10].

In order to probe the relationship of protein structure to protein function it is necessary to generate proteins with novel structures. Such proteins can be obtained by chemical modification of naturally-occurring proteins, total chemical synthesis, semi-synthesis techniques, recombinant DNA techniques, or a combination of these procedures. Each of these approaches will be considered in turn, with emphasis on the methods for the generation of proteins that include "unnatural" or "non-proteinogenic" amino acids (i.e. amino acid structures beyond the 20 commonly used in protein biosynthesis). The incorporation of non-coded amino acid structures into proteins is of particular interest to the pharmaceutical industry as a means to design bioavailable,

highly specific, longer-acting drugs. Biosynthetic incorporation of unnatural amino acids *in vitro* has dramatically expanded the scope of protein structure/function studies. New developments in this field will make possible the *in vivo* generation of proteins with an expanded repertoire of amino acid structures. It is possible that this technology may allow commercial scale protein production in the near future.

II. METHODS FOR THE PRODUCTION OF NOVEL PROTEINS

Chemical Modification

Chemical modification of solvent accessible reactive sidechains has a long history in protein science [11, 12] and a number of group-specific modifying agents are known [13, 14]. In the past, typical applications of this methodology included: identification of residues involved in catalysis or binding, introduction of reporter groups such as spin-labels, determination of protein topology by chemical crosslinking, and determination of side-chain reactivities in native proteins. More recently chemical modification techniques have been used to map extracellular regions of membrane proteins, as well as introduce novel reactive or specificity-determining elements into proteins [12]. A protein modifying reagent of great interest to the pharmaceutical industry is polyethylene glycol (PEG) which is commonly used to increase the therapeutic potential of proteins. For example, PEGylation may increase protein half-life in serum, reduce protein immunogenicity, and/or reduce toxicity (see [15] and references therein).

As an example of the early use of chemical modification in protein structure/function studies, the N-terminal glycine in the A chain of insulin was chemically modified in order to study receptor/hormone interactions. Biological activity assays and crystal structure analyses of the modified proteins were used to map putative receptor-binding regions in insulin [16]. Although this study provided a model for the conformational requirements for productive insulin-receptor

*Address correspondence to this author at Department of Chemistry, Western Washington University, Bellingham, WA 98225-9150, USA; Tel: (360) 650 3152; Fax: (360) 650 2826; E-mail: sacahill@chem.wwu.edu

binding, it was not possible to separate the effects of increasing the steric bulk of the modified insulin from the effect of neutralizing the positive charge on the N-terminal amino group. This illustrates a recurring problem in the analysis of data from such chemical modification experiments. Chemical modifications typically result in loss or reduction of activity in the protein under study. Most modifications result in an increase in the size of the reactive amino acid and any loss in activity might therefore be due to conformational perturbations of the protein rather than modification of the amino acid *per se*. When studying proteins by chemical modification, efficient methods for purification of the desired modified protein must be developed in order to accurately determine the properties of the new protein, and compare them to the native protein.

An early target for side-chain conversion studies was subtilisin (an enzyme that will be discussed below in terms of current peptide ligation strategies). It was hoped that the conversion of an active site hydroxyl (serine) to a sulfhydryl (cysteine) would result in a protease with altered (and perhaps enhanced) activity rather than an inactive enzyme. The groups of Bender [17, 18] and Koshland [19, 20] both reported the conversion of subtilisin to thiol-subtilisin by the same method. Both groups reported 0.7-0.8 equivalents of thiol/mole of thiol-subtilisin with up to 3% residual activity in the final protein preparation. Thiol-subtilisin does not show any peptidase activity, but it does hydrolyze activated esters such as *p*-nitrophenyl acetate (PNPA). It was determined that the substrate binding was not affected by the serine to cysteine conversion, but that the acylation rate was dramatically reduced, and the de-acylation rate of the acyl-enzyme intermediate was moderately reduced. The observed inactivity toward hydrolysis of peptide substrates may be due to partitioning of the acyl-thiolenzyme intermediate back to thiolenzyme•substrate complex. The greater leaving-group ability of the *p*-nitrophenylate (compared to an amino group) effectively competes with the thiolate in partitioning of the acyl-thiolenzyme intermediate, resulting in the hydrolysis of these activated substrates. In light of the experimental results, both groups concluded that the subtle changes in the active-site geometry of thiol-subtilisin were responsible for the dramatic reduction in the enzymatic activity.

More recently Wu and Hilvert reported the conversion of subtilisin to seleno-subtilisin [21] (Se-subtilisin) with >0.95 equivalents of selenium/mol of subtilisin in 40-50% overall yield. Like thiol-subtilisin, Se-subtilisin has no measurable peptidase activity, but it will cleave cinnamoylimidazole. The de-acylation of cinnamoyl-Se-subtilisin at 25 °C is 500 times slower than the analogous de-acylation of subtilisin. Se-subtilisin was thus studied as a potential "peptide ligase" (*vide infra*).

Immunoglobulins have also been chemically modified to introduce new nucleophilic catalytic functionality, as well as spectroscopic labels, into antibody binding sites. Schultz and co-workers [22] used cleavable, thiol-containing affinity labels to selectively modify lysine 52 (heavy chain) or tyrosine 34 (light chain) in F_{ab} fragments from MOPC 315. In this fashion a uniquely reactive thiol was introduced into the combining site of the F_{ab}. The thiol-containing F_{ab} accelerated the hydrolysis of coumarin esters by a factor of

60,000 over the background rate. In addition, the thiol could be selectively derivatized with a fluorescein analog, enabling a direct spectrophotometric assay for ligand binding. Uniquely reactive thiols can be introduced on the surface of a protein by site-directed mutagenesis, and then modified chemically. For example, Gloss and Kirsch treated a Lys→Cys mutant of aspartate aminotransferase with bromoethylamine to generate a -thialysine at the active-site base with altered p_{ka} [23]. As an alternative to thiol derivatization, Offord *et al.* have used periodate oxidation on chemokines with N-terminal serine or threonine residues to generate a unique aldehyde functionality for conjugation with an oxime-linked fluorophore [24]. In general, these approaches are severely limited by the kinds of modifications that can be made and in the need for a unique, exposed, reactive handle (*e.g.*, a single cysteine or a specific residue at the N-terminus). Distefano and coworkers have recently used chemical modification of the hydrophobic cavities of lipid binding proteins to generate novel transaminases (reviewed in [12]). The resulting catalysts carry out transamination reactions with good enantioselectivity, albeit with relatively slow rates.

In summary, chemical modification has provided much valuable information regarding the importance of certain residues in protein function. It has often been impossible, however, to draw meaningful conclusions relating structure and function in chemically-modified proteins since the process of modification may introduce steric perturbation into the protein structure and/or the modification yields a heterogeneous product.

Total Synthesis

The solid-phase peptide synthesis techniques introduced by Merrifield in 1963 [25] have led to the development of rapid reliable methods for the chemical synthesis of peptides up to about 45 amino acids in length [26-28]. Difficulties with undesired side-reactions, the purification of products, and the solubility of intermediates, are often encountered in the synthesis of peptides longer than 50 residues. Kent has estimated that side reactions typically occur with a frequency of 0.1% per residue [27]. He also notes that the presence of stable, slow-exchanging folded structures complicates HPLC purification of peptides longer than 50 residues in length. The coupling of highly-purified segments to give longer peptides (segment condensation synthesis) increases the purity of the final products compared to the products of a stepwise solid-phase synthesis [29]. Aside from difficulties associated with purification, the refolding of larger peptides to give active protein can be problematic [27]. The inability to separate conformational isomers may be the primary obstacle to isolating synthetic proteins with full biological activities. Solid-phase or segment condensation synthesis has been the method of choice for the routine production of peptides 30-35 residues in length, and the methods can yield suitably pure peptides of up to 50 residues. A few expert laboratories have reported the synthesis small (<140 residues) proteins (reviewed in [28]).

The refolding of synthetic proteins with multiple disulfide bonds can be complicated by the many disulfide

bond isomers and polymers obtained during the oxidation of the synthetic product. This problem is illustrated in a series of reports on the synthesis of human insulin [30, 31]. In spite of such problems, several large peptides with full biological activities have been synthesized. These include murine epidermal growth factor [32, 33] (EGF, 53 residues, 3 disulfide bonds) and -inhibin-92 [34] (92 residues, no cysteines). In Heath and Merrifield's synthesis of murine EGF, the greatest loss of material (always <15% recovery) was incurred during the disulfide bond-forming oxidation step [33]. In 1986, Kent and co-workers reported the solid-phase synthesis of murine interleukin-3 (IL-3) [35], a protein of 140 residues with two disulfide bonds. They obtained 500 mg of protein that had the same biological activities as natural IL-3, albeit with lower specific activity.

Schneider and Kent have also synthesized a 99 residue protein corresponding to the sequence of the HIV-1 protease [36]. The synthetic protein exhibited proteolytic activity toward natural and synthetic substrates of the HIV-1 protease, and the availability of the synthetic protein proved useful in the further characterization of the protease. The crystal structure of an active synthetic variant of HIV-1 protease has been solved at 2.8 Å resolution [37] and agrees well with the structure reported for biosynthetic HIV-1 protease [38].

In addition to the rapid assembly of peptides, chemical synthesis methods can be used to incorporate unnatural amino acids [39] or structural mimics [40] into peptides. This expands the range of peptide structural analogs available in the study of structure/function questions. For example, eight α -aminoisobutyric acid residues were used in the synthesis of the highly helical antibiotic alamethicin [41]. Somatostatin analogs containing non-peptide linkages and unnatural amino acids were shown to have increased stability and activity *in vivo* [42]. A synthetic analog of leutenizing hormone-releasing hormone (LH-RH) which contains a β -lactam designed to mimic a β -turn, was shown to be 2.4-fold more active than natural LH-RH [43]. Kaiser reported that insertion of unnatural amino acids into a β -endorphin analog [44] resulted in an increase in activity of 10% compared to an optimized β -endorphin analog (which is itself 3-fold more active than natural β -endorphin). A β -endorphin analog containing an amphiphilic helical segment (residues 13- 31) made up entirely of D-amino acids has been reported [45]. The product hormone contains left-handed helical structure and is active in a number of *in vitro* and *in vivo* assays. Kaiser and co-workers have studied structure/function relationships in many peptide hormones, antigens, and toxins which are thought to contain amphiphilic β -helical structural elements [46, 47]. The biological activities of these peptides have been increased by optimizing the β -helix-forming propensity of the proposed amphiphilic helical regions in these peptides. These results, along with CD data, support the existence of amphiphilic β -helices as structural elements in these peptides, and suggest a method for attenuating the respective activities of synthetic analogs. These structural optimization experiments could, in principle, be applied to the study of larger proteins.

Total synthesis of proteins by chemical means suffers from three major drawbacks. These are: (1) increasing

complications of side-reaction and failed couplings as the length of the protein increases above 50 residues; (2) difficulty in obtaining homogenous, fully-deprotected proteins (primarily due to the inability of HPLC methods to separate failed sequences in products of over 50 residues in length); and (3) loss in yield from improperly folded synthetic proteins. The first and second limitations have recently been overcome using clever segment condensation methods [48-51] that will be discussed in the context of "semi-synthetic" methods.

Semi-Synthesis

The ability to specifically cleave a protein into fragments and then reconstitute the activity of the protein by recombining those fragments (or analogs) is central to the methods of protein semi-synthesis [52, 53].

Cleavage of proteins into a small number of easily separated fragments may be achieved by standard enzymatic or chemical means, or by site-specific incorporation of a cleavable unnatural amino acid into a protein [54]. The association of fragments to yield active protein can either be non-covalent [55] (e. g. RNase S, cytochrome c) or covalent. The covalent association of fragments is typically achieved through the formation of disulfide (e.g. insulin) or peptide (e. g. trypsin inhibitor) bonds. Because at least one of the fragments of the semi-synthetic protein is obtained by chemical synthesis, this technique allows for the site-specific incorporation of unnatural amino acids into the sequence of the semi-synthetic protein. In addition, since the majority of the new protein sequence is acquired from natural sources, many of the complications encountered in the total synthesis of proteins can be avoided (e.g. incorrect refolding, disulfide bond formation, and racemization). In those cases where the expression of cloned genes is problematic, or where the incorporation of non-coded amino acids is desired, semi-synthesis has advantages over recombinant methods.

Semi-synthesis has been used to study structure/function relationships for aprotinin (Kunitz bovine trypsin inhibitor) analogs containing coded as well as non-coded amino acids [56]. Over a dozen analogs containing substitutions for the reactive site residue Lys15 were synthesized by a combination of enzymatic and chemical steps in 1-2% overall yields. The values of K_{dissoc} for complexes of aprotinin with trypsin (bovine pancreas), chymotrypsin (bovine pancreas) and elastase (human leukocyte) are 6.0×10^{-14} M, 9.0×10^{-9} M, and 3.5×10^{-6} M, respectively. The values for K_{dissoc} for complexes of the semi-synthetic analog norvaline15-aprotinin with the same proteases are 1.3×10^{-7} M, 1.4×10^{-8} M, and 1.6×10^{-10} M, respectively. These experiments demonstrate the feasibility of modulating the binding specificity of this inhibitor by varying the reactive-site amino acid.

One method for resynthesis of peptide bonds which has been used to expand the range of semi-synthesis studies is the use of proteases in reverse. In principle one can use proteases to make amide bonds based on the principle of microscopic reversibility. For a more detailed discussion of the history and theory behind protease-catalyzed peptide

bond synthesis, the interested reader is referred to the excellent reviews of Fruton [57] and Jakubke [58, 59] (and references therein). Two limitations to the use of proteases in peptide couplings are the inherent specificity of the enzymes (which limits the choice of terminal residues on the peptides to be coupled) and the secondary hydrolysis of the product peptide and/or substrates. The latter is especially bothersome in fragment condensations since hydrolysis of the substrates leads to multiple products. In spite of these complications, proteases have been used to achieve peptide bond synthesis, and remain attractive targets for the evolution of "peptide ligases". In most cases, organic cosolvent is added to favor amide bond formation [60-62] over amide cleavage.

One way to overcome the length limitation for peptide synthesis and facilitate chemical synthesis of proteins is by the ligation of smaller, chemically synthesized peptides that can be obtained in high purity. A general enzymic strategy for condensation of peptides involves the use of "subtiligase", a mutant of the serine protease subtilisin BPN' shown to ligate one peptide to a second peptide that contains a C-terminal ester bond [63]. The success of this approach is based on the fact that the mutant protease has impaired amidolytic activity while retaining good esterolytic activity (as mentioned above for thiol- and seleno- subtilisins). Wells and coworkers defined the sequence specificity of subtiligase on both sides of the incipient amide bond (including the preferred types of C-terminal esters) and demonstrated its use by synthesizing RNase A and installing a binding ligand (biotin) or heavy atoms (mercury) into proteins such as human growth hormone (hGH) and Met-hGH displayed on phage [64, 65]. Wells has selected for modified subtiligases with improved stability and activity and has optimized the methodology for efficient ligation [49, 66]. Nevertheless,

this method is limited by the solubility of the peptide fragments and the sequence requirements of the enzyme.

A related chemical approach, dubbed "native chemical ligation", has been developed by Kent and colleagues [51]. Initial work involved the attack of a sulfur nucleophile from one peptide on an alkyl bromide in a second peptide to generate a full-length protein mimic with a single unnatural mainchain bond (a thioester) [48]. This approach was improved by the direct reaction of a peptide bearing an N-terminal cysteine with a second peptide bearing a C-terminal thioester. During the coupling reaction the cysteine thiolate attacks the thioester on the second peptide, and a spontaneous rearrangement follows to generate the native amide bond. By this method, Dawson *et al.* produced full-length, 72-amino acid human interleukin 8 from two unprotected peptide fragments in high yields with a short reaction time (about 5 min) [67]. Two important improvements in the method have been reported recently. By elaborating the N-terminus of a peptide with HSCH₂CH₂O- the rearrangement can be used to generate X-Gly and Gly-X bonds (in addition to X-Cys). The oxyethanethiol group can be removed afterwards by mild treatment with Zn and acid [68]. In order to improve the efficiency of multiple ligations needed to build large proteins, a solid-phase version of native chemical ligation was developed. Variations can be used to assemble unprotected peptides either N-to-C or C-to-N. Canne *et al.* assembled human group V secretory phospholipase A2, a 118 amino acid protein, from four segments 25 to 33 amino acids each without purification between ligations [69]. The method is still limited by the need to synthesize all of the component peptides, but has allowed synthetic access to proteins (monomers) of approximately 15 kD.

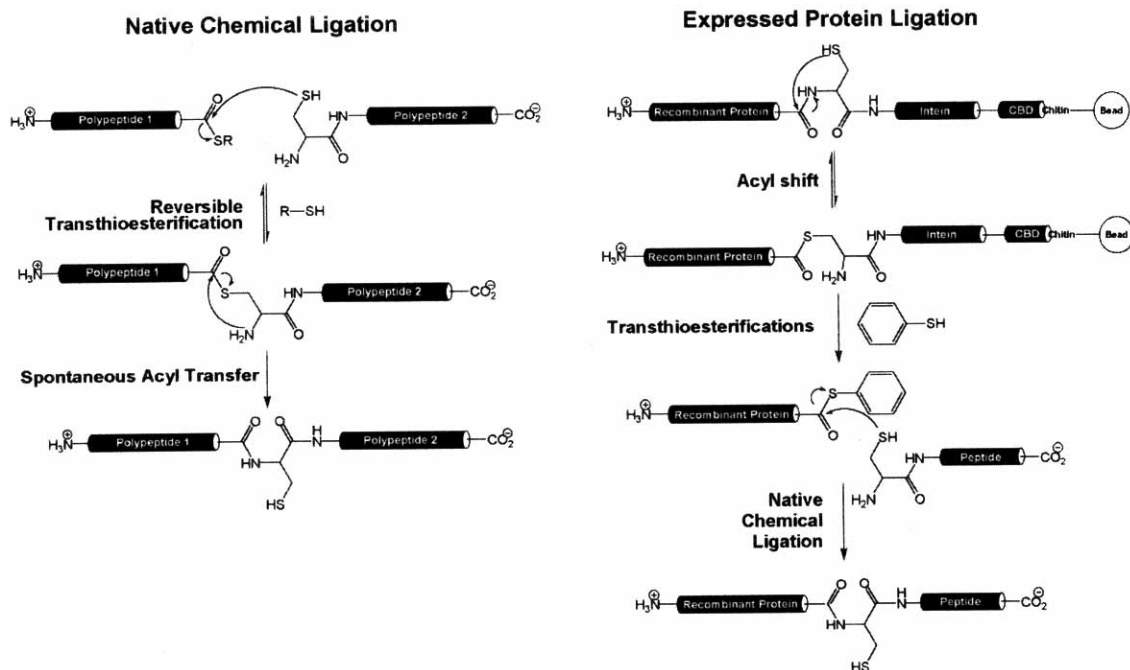


Fig. (1). Native chemical ligation and expressed protein ligation. Schematic representation of techniques for assembling synthetic and semi-synthetic proteins from peptide fragments.

A key observation that allowed the extension of this technology to much larger proteins was that protein splicing involves the excision of an intein with the intermediacy of a thioester [50]. A mutant version of the splicing domain has been generated that traps the thioester, and this has been exploited commercially as a means of protein purification, wherein the intein is linked to a chitin-binding domain and the recombinant protein is purified over chitin and released with DTT [70, 71]. Muir and coworkers instead released the trapped thioester from the resin with a synthetic peptide bearing an N-terminal cysteine, which resulted in rearrangement to yield the native amide bond. This process, called "expressed protein ligation", is similar to the solid-phase ligation developed by Kent except that the immobilized N-terminal portion of the protein is produced by recombinant methods, which allows access to proteins of virtually any size [72, 73]. Expressed protein ligation has been used to insert probes into proteins that can act as biosensors. Insertion of an environmentally sensitive 5-(dimethylamino)-naphthylene-1-sulfonamide fluorophore into a peptide between Src homology domains SH2 and SH3 allowed sensing of interaction with a bidentate peptide ligand [74]. Also, a Crk-II adapter protein was labeled with tetramethylrhodamine at the N-terminus and fluorescein at the C-terminus to generate a phosphorylation-sensitive protein by examination of FRET efficiency [75]. The method is obviously most amenable to introduction of unnatural amino acids near the C-terminus of the protein, but it can be combined with synthetic approaches to label internal regions, as well [76].

Genetic Methods

Without question the most powerful techniques for the design of novel protein structures and functions are those methods derived from recombinant DNA technology. Indeed, the term "protein engineering" entered the vocabulary of protein scientists as a result of the ability to incorporate well-defined changes in protein structures using recombinant DNA techniques. This topic is reviewed in greater detail in this volume (see previous chapter by T. Graddis and J. McGrew). The major limitation of these methods is that changes in protein structure are limited to the twenty biosynthetically incorporated amino acids (note: under certain circumstances it is possible to site-specifically incorporate selenocysteine into a protein sequence [77, 78]).

The advent of combinatorial methods such as phage display [79], DNA shuffling [80] and *in vitro* evolution [81, 82] have extended the power of recombinant technology and it is clear that such methods have wide application in the search for proteins with novel functions. With respect to expanding the natural repertoire of protein structure and function, a few specific developments deserve mention. Recently a method for site-specific incorporation of selenocysteine (Sec) into phage libraries has been described [83]. Site-specific chemical modification of the Sec residues was achieved using electrophilic modifying agents at pH values of 2.5, 5.2 or 8.0. The modification reactions were specific for the Sec residue and the phage remained infective after the modification. This technology will allow for phage

library screening with an expanded set of chemical structures.

The ability to identify the minimal functional sequence [84] for a bioactive peptide/protein is an exciting development that opens the door for the synthesis of potentially more active peptide analogs that include unnatural amino acids within the peptide sequence. Of note are successful iterative structure-based [85] and computationally determined [86] sequences that each mimic a small, stably-folded zinc finger domain. Phage display technology allows for *functional selection* of minimized peptides and is thought to be critical for the recovery of binding affinity of minimized ligands toward their receptors [87]. Using a phage display approach the 28 residue atrial natriuretic peptide (ANP) was reduced to a 15 residue analog that possesses high affinity for the natriuretic peptide receptor. The affinity of the analog is only 7.6-fold lower than the wild-type ANP affinity [87]. Phage display has also been successfully employed to increase the DNA binding affinity of a 35-residue "mini protein" by nearly 100-fold [88]. The selected minimized peptide offers a more attractive starting point for structural studies of the pharmacophore, as well as QSAR studies of synthetic analogs containing unnatural sidechain structures. For example, a series of 12-mer peptides was synthesized to probe the effects of N-substituted residues on binding to four different SH3 domains [89]. The substitution of an N-phenylethylglycine for proline in the 12-mer increased binding to the Grb2 SH3 domain by 100-fold. Significantly, the binding selectivity for the Grb2 domain increased by 100-fold as well.

In summary, genetic methods remain, for the present time, the most powerful and practical methods for protein engineering in the research lab and on industrial scale, and chemical synthetic methods allow the greatest flexibility for the introduction of novel pharmacophore structures. We now turn our attention to a hybrid strategy that attempts to take advantage of the cell's protein synthesis machinery to incorporate chemically synthesized amino acids into proteins.

Expanding the Genetic Code: A biosynthetic site-selective method.

In Vitro Methods

The cell's ability to synthesize functional proteins is unrivaled. Even in a reconstituted *in vitro* format, transcription-translation is extremely robust and efficient for producing folded soluble proteins of virtually any size. This fact inspired an *in vitro* biosynthetic approach for the site-selective insertion of unnatural amino acids (Fig. 2) [90, 91]. The method has three key requirements: (1) a genetic signal for site-selective insertion (i.e., a codon); (2) a translationally-competent tRNA to read the codon that does not interact with endogenous aminoacyl-tRNA synthetases; and (3) a method of acylating that tRNA with an amino acid of choice.

The redundancy of the genetic code provides the insertion signal: there are three stop codons (UAG, amber,

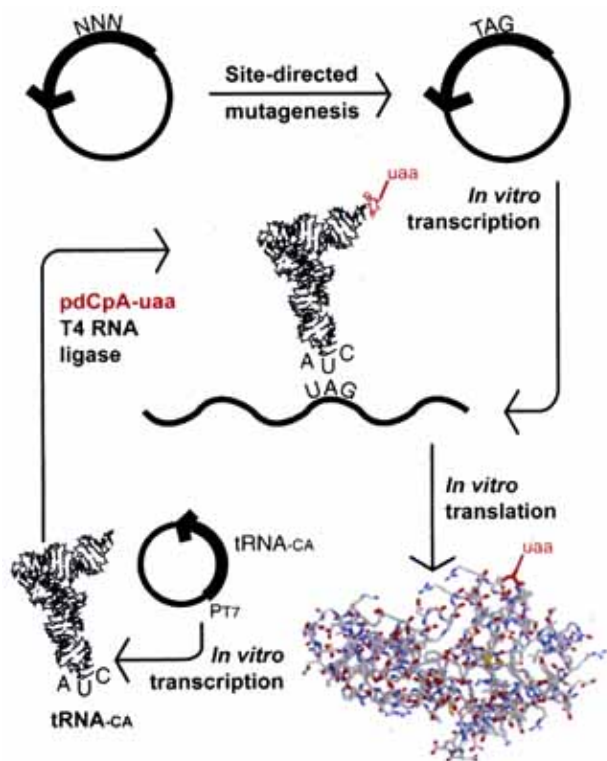


Fig. (2). A biosynthetic approach to site-specific unnatural protein mutagenesis. Site-directed mutagenesis is first used to mutate the codon for the residue of choice to the *amber* stop codon. This is added to an *in vitro* transcription-translation mixture with *amber* suppressor tRNA chemically aminoacylated with unnatural amino acid to generate full-length protein bearing the unnatural amino acid. The *amber* suppressor tRNA is *in vitro* transcribed without its final pCpA-3', and a synthetic pdCpA acylated with unnatural amino acid is ligated to the tRNA_{-CA} with T4 RNA ligase.

UGA, opal; UAA, ochre), and only one is required to terminate any given protein. Moreover, suppressor tRNAs that insert one of the natural amino acids in response to stop codons are known, and *amber* suppressors are especially robust, partially because the *amber* stop codon is the least used in *E. coli* [92]. Schultz *et al.* developed the first *amber* suppressor tRNA for this purpose that meets the two key criteria: it is not acylated (or deacylated) by aminoacyl-tRNA synthetases (aaRSs) in the *in vitro* transcription-translation reaction (which is *Escherichia coli* derived), but it is accepted by the translational machinery, including EF-Tu and the ribosome [90]. Today, the state-of-the-art uses *in vitro* transcribed yeast tRNA^{Phe}(CUA) or *E. coli* tRNA^{Asn}(CUA) (the latter especially for small, polar amino acids) in an optimized *E. coli* transcription-translation mixture generated from an RF1-deficient strain of *E. coli* to increase suppression efficiency [93-97]. Although the chemistry and suppression reactions are widely-applicable and robust (*vide infra*), this method has four major drawbacks: (1) the protein yields are poor, on the scale of 50 $\mu\text{g mL}^{-1}$ of suppression reaction; (2) the method requires

sophisticated synthetic chemistry out of the reach of many biochemistry groups who would most benefit from it; (3) the *amber* suppression method only permits site-selective insertion of a single amino acid; and (4) the modified proteins can only be examined *in vitro* because of the need to add acylated suppressor tRNA to an *in vitro* protein synthesis reaction.

Essentially, two strategies have been examined to expand the number of possible ways to specify sites for biosynthetic insertion of unnatural amino acids: (1) use of natural codons and expanded natural codons (four-base codons), and (2) development of unnatural codons containing unnatural bases. This is a very active area of research that has been reviewed in detail elsewhere [98]. Though the development of unnatural base pairs is potentially a very powerful means for extending unnatural amino acid mutagenesis technology, it remains to be seen if a single pair can be found that is stable, efficiently incorporated and extended by DNA and RNA polymerases and, ultimately, is accepted *in vivo* [99, 100].

Several recent and exhaustive reviews of unnatural amino acid mutagenesis technology have been published and interested readers will find greater detail there [101, 102]. Here we will focus on a few notable and recent examples of the generation and characterization of unnatural proteins, and the forthcoming enhancements of this technology. Since the methods are limited in various ways, and all are especially limited in the yield of protein produced, devising experiments that require small amounts of protein has been most successful. Techniques of extraordinary sensitivity, like fluorescence, measurements of catalysis and single-cell voltage clamping methods are of special note.

Schultz and colleagues have examined the effects of unnatural residues on the stability of proteins. Ala82 of T4 lysozyme (T4L), a surface residue between helices, was altered to various residues with unnatural backbone structures (see Fig. 3 for structures of the amino acids discussed in this review), such as lactic acid, proline, pipercolic acid, *N*-methylalanine, cyclopropylglycine and β -aminoisobutyric acid (Aib). It had already been shown that proline stabilized T4L at this position. Interestingly, pipercolic acid slightly destabilized the protein while Aib stabilized it, presumably due to angles different from or similar to proline, respectively. The fact that lactic acid destabilized the protein much more than *N*-methylalanine or cyclopropylglycine suggests that electronic properties have a greater effect on stability than conformational restriction [103]. This and other studies demonstrated the remarkable ability of the protein biosynthetic machinery to incorporate bulky, conformationally restricted and β -amino acids [104]. At Leu133, a buried hydrophobic residue, slightly larger amino acids predicted by molecular modeling to fill the cavity better than Leu (*S,S*-2-amino-4-methylhexanoic acid and *S*-2-amino-3-cyclopentylpropanoic acid) stabilized the protein slightly (0.6 and 1.2 kcal mol⁻¹), while successively smaller amino acids (norvaline and ethylglycine) destabilized the protein (1.1 and 3.3 kcal mol⁻¹) [105].

Thorson *et al.* examined the role of hydrogen bonding in protein stability by altering the hydrogen bonding pair

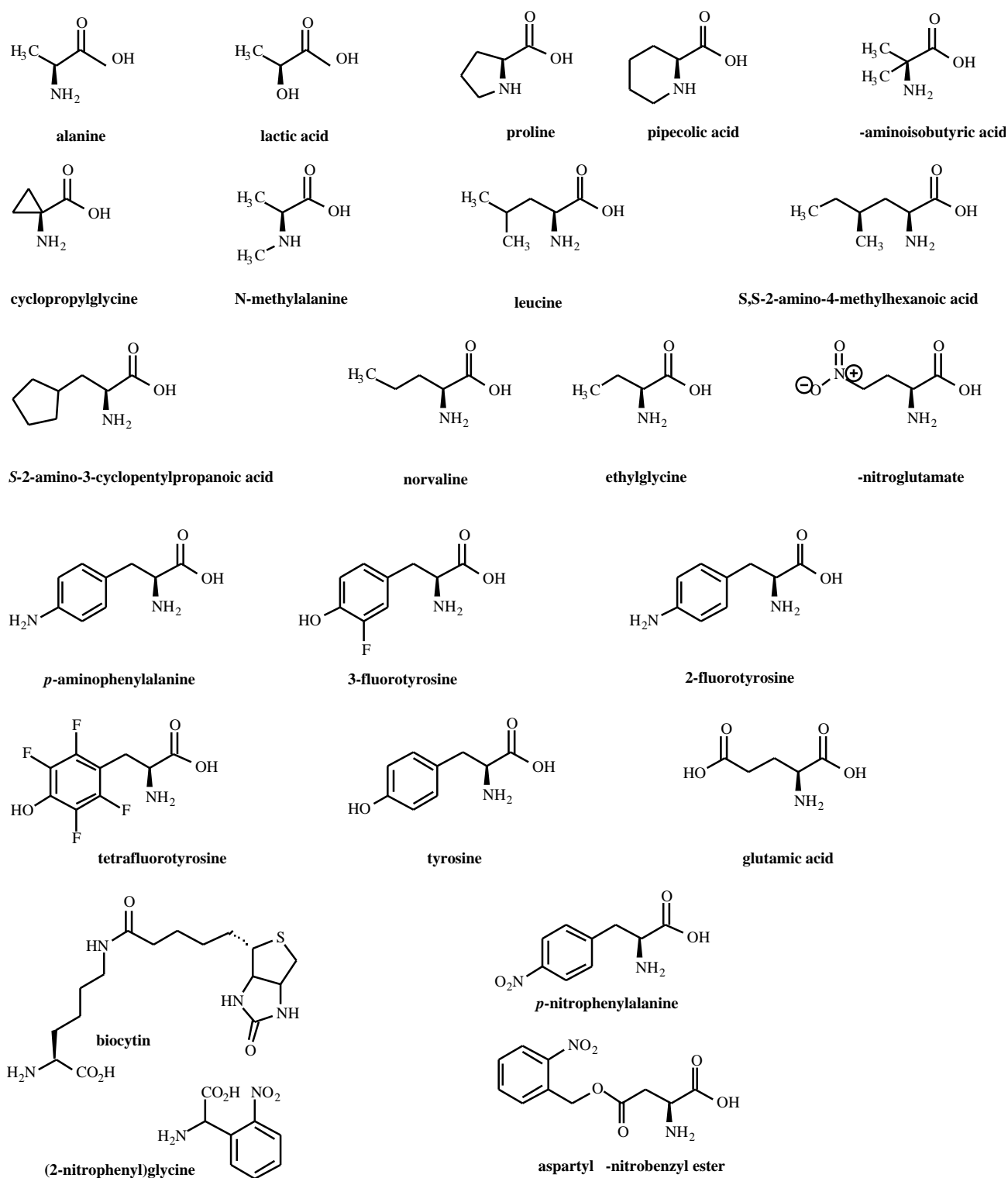


Fig. (3). Structures of some amino acids inserted into proteins with *in vitro* methods.

Tyr27-Glu10 in staphylococcal nuclease (SNase). Replacing either residue with an isosteric residue of weak hydrogen bonding ability (*p*-amino-L-phenylalanine or -nitro-L-glutamate) decreased protein stability 2.7 and 1.8 kcal mol⁻¹, respectively. Unnatural amino acid mutagenesis made

possible here a change that depends entirely on the hydrogen bond strength, since side-chain packing and solubility were very similar to the natural residues [106]. By substituting the 2-fluoro- (*pK_a* 9.3), 3-fluoro- (*pK_a* 8.8) and tetrafluoro- (*pK_a* 5.3) derivatives of Tyr (*pK_a* 10.0) at position 27, a linear free

energy analysis was applied to hydrogen bond strength. These changes increased the stability of the protein about 0.5, 1 and 2 kcal mol⁻¹, as expected for stronger hydrogen bonds on closer matching of the pK_a to the acidity of Glu10. The derived value of $\alpha = 0.35$ for the linear free energy relationship $\log K_{app} = \alpha(pK_a) + C$ suggests a nearly equal sharing of the proton between donor and acceptor. Moreover, this is strong evidence that hydrogen bonds, in addition to specifying secondary and tertiary structure, are important for protein stability [107]. A similar analysis of the importance of cation- interactions for protein stability in SNase placed the interaction strength at about 2.6 kcal mol⁻¹, similar to hydrogen bonds [108]. Dougherty's group similarly demonstrated the role of a cation- interaction in ligand binding of acetylcholine (which has a quaternary amine) to a tryptophan in the nicotinic acetylcholine receptor by measuring the effects of altered aromatic groups at the Trp site [109].

A number of residues for biophysical studies have been inserted into proteins, including isotopically labeled residues for NMR [110]; spin-labeled amino acids, fluorophores and photoaffinity labels [95, 111]; and uniquely modifiable residues bearing ketones [112]. Incorporation of biocytin, a biotin-containing amino acid, into various sites in ion channels expressed in *Xenopus* oocytes was used to determine the transmembrane topology of the channels [113]. Introduction of *p*-nitrophenylalanine into streptavidin, which bound to *N*-biotinyl-L-1-pyrenylalanine, allowed measurement of a distance decay constant (τ) for photoinduced electron transfer in proteins [114].

An especially useful kind of unnatural amino acid for biochemical studies has been the caged amino acid, which makes time-resolved studies possible upon photoinduced activation. A light-activated form of T4 lysozyme was generated by replacing the active-site Asp20, which stabilizes the carbocation generated in degradation of the β -linked NAM-NAG cell wall. The nitrobenzyl ester of the Asp20 side-chain carboxylate was efficiently removed by irradiation with a Hg-Xe arc lamp to produce active enzyme [115]. The nitrobenzyl ether of tyrosine was used to cage residues in the nicotinic acetylcholine receptor (nAChR) in *Xenopus* oocytes, and pulse irradiation was used to confirm that at sites critical for ligand binding, two tyrosines had to be decaged in order to see receptor activation, corresponding to the two such subunits in the receptor [116]. An interesting related method involves site-specific photochemical proteolysis upon irradiation of protein containing (2-nitrophenyl)glycine. The method was used to demonstrate the functional necessity of a loop formed as a result of a disulfide bond in nAChR [117]. Caged groups have also been used to control protein-protein interactions, as in the dimerization of HIV-1 protease on decaging of Asp25 on the dimer interface [118], or the interaction of ras with effector p120-GAP on decaging of Asp38 of ras [119].

Progress Toward In Vivo Systems

Extension of this technology to *in vivo* protein expression systems is required if it is to be practical for the production of commercial proteins containing unnatural amino acids. Along these lines, Dougherty and coworkers devised a way

to make use of very sensitive voltage clamping techniques to examine single cells producing proteins containing unnatural amino acids through microinjection of chemically *in vitro* acylated tRNA. This technique allows the *in vivo* production (in *Xenopus* oocytes) of proteins with unnatural residues, and is especially useful for proteins in the membrane that could not be produced *in vitro* due to folding and expression problems [120]. It was found that the initial tRNA designed for this approach, while an improvement over tRNA^{Phe} in protein yield, was a substrate for *Xenopus* aaRSs. A modified version of *Tetrahymena thermophila* tRNA^{Gln}(CUA), a tRNA that naturally inserts glutamine in response to UAG (which is not a stop codon in *Tetrahymena*) was both efficient and not a substrate for the endogenous aaRSs of the oocyte [121].

When devising strategies for *in vivo* systems for unnatural amino acid mutagenesis, it is instructive to consider well-known cases of misincorporation of non-proteinogenic amino acids, some of which act as growth inhibitors on different kinds of cells. The significant finding here is that naturally-occurring enzymes can be used to incorporate non-proteinogenic amino acids into proteins. For example, growth of *E. coli* is inhibited by L-canavanine (an analog of arginine) because of misincorporation into proteins. Simply, cells cannot possibly be equipped to discriminate the natural amino acids from all of the amino acid analogs it is possible to synthesize. In the absence of some common environmental assault, in fact, there is little reason for cells to have generated mechanisms to distinguish natural substrates from unnatural ones. Presumably, this is why L-canavanine is taken up and inserted into yeast and bacterial proteins, whereas ornithine and citrulline (non-proteinogenic amino acids that are intermediates in arginine biosynthesis) are discriminated against [122]. This has led to the development of strains and culture conditions for producing proteins with unnatural amino acids that are non-specifically incorporated.

For example, the advent of an *E. coli* expression system that allowed high-level replacement of methionine with selenomethionine (SeMet) has made possible the widespread use of multiwavelength anomalous diffraction (MAD) for solving X-ray crystal structures. Some auxotrophic strains of *E. coli* can be used to express proteins with 100 % SeMet incorporation with careful Met-free culturing conditions [123, 124]. Many other Met analogs can be inserted this way, including 2-aminohexanoic acid (norleucine), ethionine, telluromethionine and *S*-nitrosomethionine [125, 126]. Budisa and coworkers have used these analogs to examine subtle effects on protein folding from these "atomic" mutations [127]. NMR structural characterization has been aided by the introduction of 5-fluorotryptophan (5-F-Trp) into proteins using strains auxotrophic for tryptophan [128-130]. Incorporation of 5-F-Trp was used to assess the symmetry of the large (380 kD) F₁-ATPase from *E. coli*, since Trp is rare and ¹⁹F is NMR-active, and at 100 % natural abundance. In smaller proteins like the *E. coli* glucose/galactose receptor (33 kD), all five of the individual 5-F-Trp residues could be distinguished, especially in combination with site-directed mutagenesis [131, 132]. Aided in part by the high yield of protein that can be produced *in vivo*, Armstrong and coworkers have solved a

crystal structure of a 5-F-Trp labeled glutathione transferase which has enhanced catalytic activity by virtue of its unnatural amino acids [130]. In these studies the 5-F-Trp is inserted at all the Trp positions in the protein. More reliable insertion methods have recently been developed, and the rarity of Trp residues makes it possible in some cases to make single-site changes, as in replacement of the lone Trp187 in human annexin V with 4-fluoro-, 5-fluoro- and 6-fluoro-Trp to examine changes in enzyme stability and activity [133].

Evolution of New Enzymatic Function.

In vivo methods would have the significant advantages of (1) high yield of protein and easy scale-up (more cells yield more protein); (2) technical ease (simple addition of the

unnatural amino acid to the medium); and (3) the potential to observe the altered proteins in the living cell using caged proteins, affinity labeled proteins or proteins bearing biophysical probes or moieties to expand their functionality. In addition to the requirements described above for *in vitro* mutagenesis with unnatural amino acids, there are at least three additional design considerations for an *in vivo* method. First, the suppressor tRNA must be acylated *in vivo*, an aminoacyl-tRNA synthetase is required that uniquely acylates the suppressor tRNA (i.e. the synthetase must itself be highly selective, acylating only the suppressor tRNA and not endogenous tRNAs). Second, the synthetase must be capable of acylating the suppressor tRNA with an unnatural amino acid but also incapable of using any other naturally occurring amino acid (proteinogenic or not) as a substrate. Third, the unnatural amino acid must be taken up by the cell

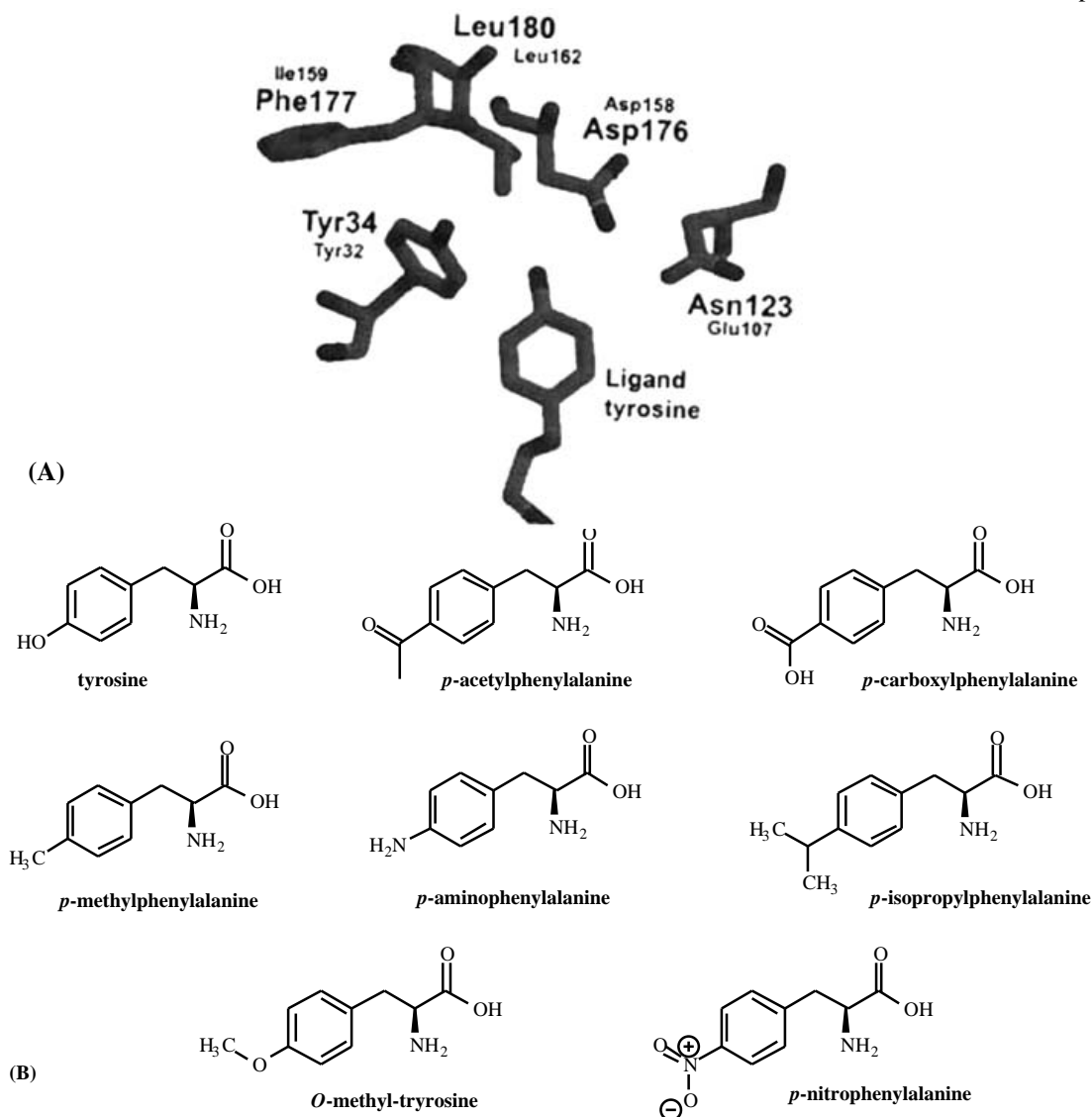


Fig. (4). TyrRS *para*-targeted library and tyrosine *para*-substituted analogs. (A) The indicated residues (smaller font) were randomized in *M. jannaschii* TyrRS due to proximity to the phenolic oxygen of the ligand tyrosine as evidenced in the co-crystal structure of *B. stearothermophilus* TyrRS with an analog of Tyr-AMP. The homologous residues identified from sequence alignment are indicated (larger font). (B) *Para*-substituted analogs of tyrosine used in the selection with the library. (A) was created with RasMol 2.6 from the PDB file 3ts1.

or produced by it, and it cannot be inserted by any other synthetase or, more generally, be markedly toxic to the cell. Moreover, since a great deal of protein engineering will be required to achieve these goals, a well-understood organism that can be transformed with large libraries (*i.e.*, with high transformation efficiency) is desired. Similarly, since the required novel synthetase is likely to be derived from a naturally-occurring enzyme, a well-characterized protein is desired as a starting point, especially one for which an X-ray crystal structure exists [134].

In the remainder of this review we describe current efforts to develop the technology for *in vivo* unnatural amino acid mutagenesis. *E. coli* is the logical choice for the host organism given the existing technologies for engineering proteins in bacteria. The *amber* stop codon was selected initially as the insertion signal due to the excellent suppression that is possible with known suppressor tRNAs isolated from *E. coli* strains; however, attempts to make orthogonal tRNA/synthetase pairs using four-base codons are underway, in light of identification of easily suppressible extended codons (J.C. Anderson, T.J. Magliery and P.G. Schultz, unpublished work). Initially, an engineering approach was adopted to generate the suppressor tRNA/synthetase pair, starting with the extremely well-characterized *E. coli* glutamine pair. Since that time, the Schultz group and others have found it more advantageous to import *amber* suppressor tRNAs and aminoacyl-tRNA synthetases from other organisms (“heterologous pairs”, see Table 1). Of course, these heterologous pairs insert a natural amino acid, and so robust screens and selections have been developed to find variants of the synthetases from carefully designed libraries of mutagenized enzymes capable of acylating suppressor tRNA with unnatural amino acids. Using these methods, the Schultz group and others have seen the first successes in engineering living cells that are capable

of site-selectively inserting unnatural amino acids into proteins.

Work from many labs directed at the generation of the requisite suppressor tRNAs has been recently reviewed [98] and will not be discussed in detail here. We will mention a few significant developments and then focus on the engineering of the tRNA synthetases. These suppressor/synthetase pairs have been called “orthogonal” pairs. “Orthogonal” is a term borrowed from the protecting group nomenclature of synthetic chemists. In this context “orthogonal” means that the suppressor tRNA is only recognized by its cognate synthetase (*i.e.* it is neither acylated nor deacylated by one of the host organism enzymes). Likewise, the orthogonal synthetase only charges the suppressor with the desired unnatural amino acid (*i.e.* it will not charge an endogenous tRNA with the unnatural amino acid, nor will it charge the suppressor with an amino acid other than the desired unnatural one).

The first orthogonal tRNA developed for the purpose of *in vivo* site selective delivery of unnatural amino acids was derived from *E. coli* tRNA^{Gln}. Glutamyl-tRNA synthetase (GlnRS) was known to acylate its *amber*-suppressing derivative (anticodon 5'-CUA-3') and biochemical and X-ray crystal structural information defined the nature of the interaction between tRNA and synthetase (see Magliery [98] and references therein). Three sites (so-called “knobs”) at which mutations were expected to modulate the ability of GlnRS to acylate the tRNA were selected, and tRNAs bearing mutations at each site (and in all possible combinations) were generated. These mutations turned out to interact in complicated, non-additive ways both with respect to aminoacylation by GlnRS and performance as tRNAs for delivery of amino acids at the level of translation. The *amber* suppressor tRNA with all three knob mutations,

Table 1. Orthogonal tRNA/aaRS Pairs for Delivering Unnatural Amino Acids

For use in	tRNA	Synthetase	Notes	Ref.
<i>E. coli</i>	O-tRNA ^{Gln} (CUA), mutant of <i>E. coli</i> tRNA ₂ ^{Gln} with three “knob” mutations	selected mutant of <i>E. coli</i> GlnRS	tRNA is not as orthogonal as O-ScRNA ^{Gln} (CUA); synthetase still acylates w.t. tRNA ^{Gln}	[136]
<i>E. coli</i>	O-ScRNA ^{Gln} (CUA)	yeast GlnRS	tRNA is highly orthogonal but ScGlnRS activity is weak	[140]
<i>E. coli</i>	O-MjtRNA ^{Tyr} (CUA)	<i>M. jannaschii</i> TyrRS	tRNA is not as orthogonal as O-ScRNA ^{Gln} (CUA) but MjTyrRS is very active	[153]
<i>E. coli</i>	O-MjtRNA ^{Tyr} (CUA)*, mutant of O-MjtRNA ^{Tyr} (CUA)	<i>M. jannaschii</i> TyrRS	mutations in tRNA reduce recognition by <i>E. coli</i> aaRSs while maintaining recognition by MjTyrRS	[153]
<i>E. coli</i>	O-EctRNA ^{Met} , mutant of the <i>E. coli</i> initiator tRNA	Selected mutant of yeast TyrRS	Mutant ScTyrRS selected not to acylate EctRNA ^{Pro}	[154]
<i>E. coli</i>	O-ScRNA ^{Asp} (CUA)	D188K mutant of yeast AspRS	AspRS(D188K) has very weak activity	[155]
yeast	O-HstRNA ^{Met} (CUA), mutant of human initiator tRNA	<i>E. coli</i> GlnRS	First pair for use in eukaryotic cells	[154]

O-tRNA^{Gln}(CUA), met the criteria for an orthogonal tRNA in *E. coli*: it was not a substrate for endogenous synthetases but was competent to act in translation [135].

This tRNA was also characterized *in vivo* by transforming a plasmid encoding the O-tRNA^{Gln}(CUA) into an *E. coli* strain with an *amber* mutation in the gene for β -galactosidase (*lacZ*). The cells were incapable of surviving on lactose minimal media due to the fact that the O-tRNA^{Gln}(CUA) was not appreciably acylated and thus full-length, functional LacZ could not be produced. This observation was used as the basis for a selection for a GlnRS mutant capable of aminoacylating the O-tRNA^{Gln}(CUA). The gene for GlnRS (*glnS*) was randomly mutagenized by the method of DNA shuffling [80] and co-transformed on a compatible plasmid into the *E. coli* strain bearing the O-tRNA^{Gln}(CUA) expression plasmid and the *lacZ_{am}* mutation. Cells that survived on lactose minimal media contained library members with increased ability to acylate the O-tRNA^{Gln}(CUA) (presumably with glutamine). These were expressed, purified, examined *in vitro* for ability to acylate O-tRNA^{Gln}(CUA) and tRNA^{Gln}, pooled and resubmitted to mutagenesis and selection. After seven rounds of mutagenesis and selection, a mutant GlnRS was found that acylated the wild-type tRNA^{Gln} substrate only 9-fold better than the O-tRNA^{Gln}(CUA) and was down only 250-fold with respect to acylation of tRNA^{Gln} by wild-type GlnRS. This enzyme, with an overall 1,500-fold change in specificity, was capable of acylating the O-tRNA^{Gln}(CUA) as determined by Western blot detection of full-length protein produced from an *amber* mutant of the gene for *E. coli* surface protein LamB. No full-length protein was observed with the O-tRNA^{Gln}(CUA) only. Despite the remarkable change in activity, this mutant GlnRS was still not ideal, since it acylated the wild-type tRNA^{Gln} about as well as the orthogonal suppressor [136]. This would cause toxicity through insertion of an unnatural amino acid at other Gln codons in *E. coli* proteins [137].

Recently, Schimmel *et al.* showed that *E. coli* GlnRS (*EcGlnRS*) does not acylate *Saccharomyces cerevisiae* tRNA^{Gln} (*SctRNA^{Gln}*) due to the lack of an N-terminal RNA-binding domain that *S. cerevisiae* GlnRS (*ScGlnRS*) possesses [138, 139]. Liu and Schultz [140] showed that the *amber* suppressing derivative of *SctRNA^{Gln}* (O-*SctRNA^{Gln}*(CUA)) and *ScGlnRS* constitute an orthogonal tRNA/synthetase pair in *E. coli*. The O-*SctRNA^{Gln}*(CUA) was neither acylated by purified *EcGlnRS*, nor did it mediate suppression of an *amber* mutation *in vitro*; however, when chemically acylated with valine, O-*SctRNA^{Gln}*(CUA) caused efficient *amber* suppression, indicating that it is orthogonal to *E. coli* aaRSs and translationally competent. *ScGlnRS* does not appreciably acylate *E. coli* tRNAs *in vitro*, but it does acylate O-*SctRNA^{Gln}*(CUA). This pair was also characterized *in vivo* by co-transforming *E. coli* with a plasmid encoding O-*SctRNA^{Gln}*(CUA) and a compatible plasmid encoding an *amber* mutant of the gene for β -lactamase (*amp*) and the gene for *ScGlnRS* or a mutant thereof. This *amber* mutation occurs at a permissive site (Ala184), so that insertion of virtually any amino acid confers resistance to ampicillin [141]. With an inactive mutant of *ScGlnRS*, these cells exhibited an IC₅₀ of about 20 $\mu\text{g ml}^{-1}$ ampicillin, indicating virtually no acylation by

endogenous synthetases. With an active *ScGlnRS*, cells exhibit an IC₅₀ of about 500 $\mu\text{g ml}^{-1}$ ampicillin, indicating that the *ScGlnRS* acylates the O-*SctRNA^{Gln}*(CUA) in *E. coli* [140].

There have been limited attempts to alter the substrate specificity of aminoacyl-tRNA synthetases with respect to the amino acid. This has proven to be a more difficult problem than tRNA specificity, probably because of the very large interaction surface with the tRNA (25 kD) in contrast to the amino acid (120 Daltons). One simple modification that has been useful is the Ala294 \rightarrow Gly mutant of *E. coli* PheRS, which allows insertion of *p*-Cl-Phe and *p*-Br-Phe in addition to *p*-F-Phe and Phe (which are both accepted by the wild-type PheRS). The mutation was discovered as a result of the observation that Ala294 \rightarrow Ser conferred resistance to *p*-F-Phe, suggesting that this residue restricts the amino acid binding cavity at the *para* ring position [142-144]. Two groups have demonstrated that mutants of GlnRS can be made to weakly accept glutamate [145, 146]. Other experiments have shown that amino acid specificity is dependent upon tRNA recognition; for example, tRNA^{Gln} mutants with changes at U35 (the middle of the anticodon) cause approximately 15-fold increases in the K_M for glutamine (and 20-fold decreases in k_{cat}). For this to occur, recognition of elements far from the active site of the enzyme (which binds the acceptor end of the tRNA) must be transduced back to the active site [146, 147]. These studies suggest that altering amino acid specificity will likely require more than simple alteration of residues that bind the natural substrate amino acid.

Given these observations, Schultz *et al.* have taken a semi-rational approach to the design of libraries for altering amino acid specificity. For example, one set of libraries of *ScGlnRS* variants was created by randomization of seven residues that either bind the substrate glutamine, bind a water molecule that is hydrogen bonded to the substrate glutamine, or position these residues [148]. Saturation mutagenesis of seven residues produces about one billion unique molecules, and DNA subcloning and transformation efficiency place an upper-limit of about 10⁹ on library size in *E. coli*. Since virtually all of these *ScGlnRS* variants will differ dramatically from wild-type at multiple residues proximal to the substrate amino acid, very few are active even weakly (about 1 in 5,000). These types of libraries are then amenable to positive selection or screening, followed by iteration of random mutagenesis and recombination by DNA shuffling and further selection (T.J. Magliery, S.W. Santoro and P.G. Schultz, unpublished results). These libraries were designed for use with very near analogs of glutamine, including *N*-methyl-, *N*-ethyl-, *N*-hydroxyethylene-, *N*-benzyl- and *N*-phenylglutamine, guanidinoalanine, homoglutamine and methionine sulfoxide. We have found this approach superior to relying purely on random mutagenesis, which requires a great deal of both luck (i.e., that suitable enzymes are fairly nearby and dense in sequence space) and performance from a negative selection to remove the many active, near-variants of the wild-type synthetase. This approach has recently been applied to *MjTyrRS*, as well (see below).

A selection for the insertion of an unnatural amino acid requires that the survival of the cell depends on something it

fundamentally does not need (*i.e.*, is “unnatural”). A way around this difficulty is the use of a general, double-sieve scheme that in two steps demands (1) active variants of the aaRS and (2) rejection of natural amino acids as substrates. For example, the first selection is from a pool of variants of an orthogonal aaRS in *E. coli* bearing an orthogonal *amber* suppressor tRNA and an antibiotic resistance gene with an *amber* mutation corresponding to a permissive position in the enzyme. In the presence of unnatural amino acids and antibiotic, survivors of the selection must contain synthetases capable of acylating the orthogonal tRNA(CUA) with some amino acid, natural or unnatural. The selected aaRSs are then transformed into a second strain of *E. coli* with the O-tRNA(CUA) and the gene for a toxic protein bearing an *amber* mutation at a permissive site. Survivors of this selection, grown in the *absence* of unnatural amino acids, must therefore contain an active synthetase that is capable of rejecting all endogenous amino acids—overall, an enzyme capable of uniquely acylating with an unnatural amino acid.

The initial screen involved the yeast glutamine orthogonal pair, positive selection with ampicillin in the context of β -lactamase bearing an Ala184→*amber* mutation, and negative selection with the toxic enzyme barnase bearing two or three *amber* mutations (Gln2, Asp44 and, optionally, Gly65). For the positive selection step, enrichment factors as high as 200,000 could be achieved for cells containing wild-type *ScGlnRS* in a high dilution of cells containing an inactive form of the synthetase, *ScGlnRS* 500 (1:10⁷), using 500 $\mu\text{g mL}^{-1}$ ampicillin. For the negative selection step, enrichment factors as high as 3 million for the three-*amber* mutant barnase and 3 x 10⁷ for the two-*amber* mutant barnase could be achieved for selection of inactive *ScGlnRS* 500 diluted 1:10⁷ into *ScGlnRS* [140].

Among the factors that allowed for substantially higher enrichment from the negative selection than from the positive, ampicillin-based selection is the fact that β -lactamase acts from the periplasm to hydrolyze the ampicillin in the media. While this clearly confers a growth advantage upon the cells that possess active β -lactamase, it also allows rescue *in trans*, since nearby cells are rescued once a sufficient amount of ampicillin has been hydrolyzed. A second problem associated with ampicillin selection is that ampicillin is bacteriocidal, and it is difficult to know what dosage is applicable when one is selecting for potentially weak synthetases from a pool of virtually inactive synthetases. For an improved positive selection, an *amber* mutant of chloramphenicol acetyltransferase (CAT) was generated at a site known to be fairly permissive [149]. In contrast to ampicillin, chloramphenicol inhibits growth instead of killing cells, and therefore the amplification of cells is simply tied to the amount of *amber* suppression in the given cells, without loss of cells containing weakly active synthetases. Chloramphenicol acts at the ribosome, and CAT is expressed in the cytosol, reducing the action *in trans* of this mechanism of resistance. Moreover, increasing the concentration of chloramphenicol increases the selective advantage for cells bearing wild-type synthetase versus cells containing an inactive synthetase, so a nearly arbitrary concentration of chloramphenicol can be used for selection from a library (T.J.Magliery and P.G. Schultz, unpublished results).

A variation on the general, double sieve selection has been introduced that uses fluorescence-activated cell sorting (FACS) and a variant of green fluorescent protein (GFP) as a reporter. In a first step, cells containing the gene for T7 RNA polymerase with multiple *amber* mutations, GFP under the control of the T7 promoter, orthogonal tRNA(CUA) and a library of variants of orthogonal aaRS are grown in the presence of unnatural amino acids. These cells are then examined for fluorescence, either by FACS, fluorimetry or visually on plates with long-wave UV irradiation. Fluorescent cells are diluted and grown in the absence of unnatural amino acids. Here, cells that fail to fluoresce must contain a synthetase that is able to reject natural amino acids but is known to be active toward the unnatural amino acid substrate from the first screen. This system has been shown to be useful for the *ScGln* and *MjTyr* pairs, and is being used as a method for both the screening of libraries and the characterization of selectants from antibiotic selections. FACS offers the additional advantage that one can select the appropriate level of fluorescence as a cut-off for sorting, thereby altering the stringency of either the negative or positive step (S.W. Santoro and P.G. Schultz, unpublished results). Modern FACS is capable of sorting a billion bacterial cells a day, which is comparable to the largest libraries that can be conveniently generated in *E. coli*.

Although direct selections for unnatural amino acid insertion are very difficult to conceive, direct screens are possible. Pastrnak and Schultz developed an antibody recognition-based approach to the screening of M13 phage displaying a surface epitope into which amino acids can be inserted via *amber* suppression [150]. Here, M13 phage harbor genes directing the expression of a variant aaRS (the engineered O-*ScAspRS* or *ScGlnRS*) and the corresponding orthogonal tRNA(CUA). Helper-phage VCSM13 was modified to display an *amber* mutant of the immunogenic C3 epitope derived from poliovirus; also, since the *amber* mutation corresponds to a residue near the N-terminus of the C3-pIII fusion, phage production requires *amber* suppression. Therefore, this is both a selection for *amber* suppression (*i.e.* active aaRS) and a means of displaying an amino acid on the surface of M13 phage for screening. Antibodies were elicited from synthetic peptides of the C3 epitope containing the amino acid of choice at the position corresponding to the *amber* mutation in the C3-pIII fusion. To use this system with the relatively weak O-*ScAspRS*, an RF-1 deficient version of the excellent cloning strain DH10B was constructed. In a model selection in which M13 phage with C3-pIII fusions with an Asp residue in the C3 epitope were screened from a large excess of C3-pIII-containing phage with Asn at the same residue, 900-fold enrichments were possible with a monoclonal antibody elicited against C3(Asp) peptide [150]. Since antibodies against C3 epitopes with natural or unnatural amino acids can be generated, both positive and negative screening can be carried out.

The critical question in this research is “can living organisms be engineered to allow site-directed mutagenesis with unnatural amino acids?” The answer is “yes”. Taking advantage of the fact that mutants of PheRS are known that reject *p*-F-Phe, Furter was partially successful in engineering a bacterium capable of inserting *p*-F-Phe in a site-selective manner [151]. The G37A mutant of yeast tRNA^{Phe}(CUA)

was found to be nearly orthogonal, although it was a poor substrate of LysRS in *E. coli*. When yeast PheRS was co-expressed with this tRNA in *E. coli*, 95 % of the amino acid inserted into an *amber* mutant of dihydrofolate reductase (DHFR) via O-tRNA^{Phe}(CUA) was phenylalanine. When this pair was expressed in a *p*-F-Phe-resistant, Phe-auxotrophic strain of *E. coli*, growth in the presence of high levels of *p*-F-Phe resulted in largely site-specific insertion at *amber* mutations. Under optimal conditions, about 75 % of the *amber*-encoded site in DHFR was occupied by *p*-F-Phe, while 20 % was Phe and 5 % was Lys. This indicates both that the O-tRNA^{Phe}(CUA) is being promiscuously acylated by *Ec*LysRS and that the *Sc*PheRS inserts Phe in addition to *p*-F-Phe. Also, when the same site in DHFR was replaced with a Phe codon, 93 % of this site was occupied by Phe, but 7 % was *p*-F-Phe, indicating that the endogenous *Ec*PheRS incorporates a small amount of *p*-F-Phe in addition to Phe. The yield of DHFR was high, however (about 10 mg L⁻¹ culture), and since the natural amino acids are silent in ¹⁹F NMR, this system may be useful for this application despite its lack of specificity. Ultimately, engineering this pair for better orthogonality and amino acid specificity would be required, as would engineering the *E. coli* synthetase to better reject the unnatural amino acid. Of course, the most significant limitation to this approach is that it requires synthetases already known to use an unnatural amino acid as a substrate.

Recently, Schultz and coworkers applied a general, directed library strategy combined with the improved, double-sieve selection based on chloramphenicol resistance to find a mutant of *Mj*TyrRS capable of acylating the selected O-*Mj*tRNA^{Tyr}(CUA)* specifically with the methyl ether of tyrosine (*O*-Me-Tyr). A small library was constructed by randomization of five of the residues of *Mj*TyrRS proximal to the tyrosine binding site. This library was subjected to positive, chloramphenicol-based selection in the presence of *para*-substituted analogs of tyrosine and negative, barnase-based selection in the absence of the unnatural amino acids. Selectants were pooled, subjected to DNA shuffling and reselected twice. However, the final selectant with activity toward *O*-Me-Tyr had mutations only at the originally-randomized positions from the first library, suggesting that the shuffling steps were unnecessary here. The relative importance of the four amino acid changes in the protein (Tyr32→Gln, Glu107→Thr, Asp158→Ala, Leu162→Pro) is not yet known. However, the specificity of this mutant *Mj*OMeTyrRS is remarkable. When the tRNA and synthetase are present with an *amber* mutant of the gene encoding dihydrofolate reductase, no DHFR can be detected unless *O*-Me-Tyr is added to the media. Moreover, using Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, the DHFR isolated after growth in the presence of the unnatural amino acid was confirmed to carry the *O*-Me-Tyr at a single position. Not only is this mutant synthetase capable of adenylating *O*-Me-Tyr about 8-fold faster than tyrosine at saturation (as measured by pyrophosphate exchange), the *K_M* for tyrosine is about 13-fold higher than for *O*-Me-Tyr [152]. Efforts are currently underway to “transplant” this active site into TyrRSs from other organisms, in order to transfect mammalian cells with a heterologous aaRS/tRNA(CUA) pair capable of specifically

inserting *O*-Me-Tyr (S.W. Santoro, J. Chin, T.J. Magliery and P.G. Schultz, unpublished).

The major challenges that lie ahead in this field are substantial; however, significant progress has been made to show that they are surmountable. Application of selections for novel unnatural amino acid specificity toward side chains significantly different from the natural set will be needed to make useful tools for cell biology. As these new aaRSs are generated, novel means of inserting these unnatural amino acids site-specifically will have to be demonstrated (e.g. using four-base codons and unnatural codons). Finally, transitioning these systems into eukaryotes like yeast and mammalian cells will make unnatural cell biology possible in a milieu in which human disease and drug targets can be studied with remarkable precision.

REFERENCES

- [1] Zoller, M.J. and Smith, M. (1983) *Methods in Enzymology*, **100**(11), p. 468-500.
- [2] Ulmer, K.M. (1983) *Science*, **219**, p. 666-671.
- [3] Oxender, D.L. and Fox, C.F. eds. (1987) *Protein Engineering*. Tutorials in molecular and cell biology, ed. Fox, C.F., Liss, A. R: New York.
- [4] Leatherbarrow, R.J. and Fersht, A.R. (1986) *Protein Eng.*, **1**(1), p. 7-16.
- [5] Knowles, J.R. (1987) *Science*, **236**(4806), p. 1252-8.
- [6] Trower, M.K. ed. (1996) *In vitro mutagenesis protocols*. Methods in Molecular Biology. Vol. **57**. Humana Press: Totowa, NJ. 408.
- [7] Fersht, A.R., Leatherbarrow, R.J. and Wells, T.N.C. (1987) *Biochemistry*, **26**, p. 6030-6038.
- [8] Cunningham, B.C. and Wells, J.A. (1989) *Science*, **244**, p. 1081-1085.
- [9] Bartlett, P.A. (1988) *The interplay between enzyme mechanism, protein structure, and inhibitor and catalyst design*, in *Protein Structure and Protein Engineering*, Winnacker, E.-L. and Huber, R. Editors. Springer Verlag: Berlin. p. 86-95.
- [10] Bryson, J.W., Betz, S.F., Lu, H.S., Suich, D.J., Zhou, H.X., O'Neil, K.T. and DeGrado, W.F. (1995) *Science*, **270**, p. 935-941.
- [11] Kaiser, E.T., Lawrence, D.S. and Rokita, S.E. (1985) *Ann. Rev. Biochem.*, **54**, p. 565-595.
- [12] Distefano, M.D., Kuang, H., Qi, D. and Mazhary, A. (1998) *Curr. Opin. Struct. Biol.*, **8**, p. 459-465.
- [13] Hermanson, G.T. (1996) *Bioconjugate Techniques*, New York: Academic Press. **785**.
- [14] Lundblad, R.L. (1994) *Techniques in Protein Modification*, Boca Raton: CRC Press.
- [15] Tsutsumi, Y., Onda, M., Nagata, S., Lee, B., Kreitman, R.J. and Pastan, I. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, p. 8548-53.

- [16] Pullen, R.A., Lindsay, D.G., Wood, S.P., Tickle, I.J., Blundell, T.L., Wollmer, A., Krail, G., Brandenburg, D., Zahn, H., Gliemann, J. and Gammeltoft, S. (1976) *Nature*, **259**, p. 369-373.
- [17] Polgar, L. and Bender, M.L. (1966) *J. Am. Chem. Soc.*, **88**(13), p. 3153-3154.
- [18] Polgar, L. and Bender, M.L. (1967) *Biochemistry*, **6**(2), p. 610-620.
- [19] Neet, K.E. and Koshland, D.E. Jr. (1966) *Proc. Natl. Acad. Sci. USA*, **56**, p. 1606-1611.
- [20] Neet, K.E., Nanci, A. and Koshland, D.E. Jr. (1968) *J. Biol. Chem.*, **243**(24), p. 6392-6401.
- [21] Wu, Z.-P. and Hilvert, D. (1989) *J. Am. Chem. Soc.*, **111**, p. 4513-4514.
- [22] Pollack, S.J., Nakayama, G. R. and Schultz, P.G. (1988) *Science*, **242**, p. 1038-1040.
- [23] Gloss, L.M. and Kirsch, J.F. (1995) *Biochemistry*, **34**(12), p. 3990-8.
- [24] Offord, R.E., Gaertner, H.F., Wells, T.N. and Proudfoot, A.E. (1997) *Methods in Enzymology*, **287**(5310), p. 348-69.
- [25] Merrifield, R.B. (1963) *J. Am. Chem. Soc.*, **85**, p. 2149-2154.
- [26] Merrifield, B. (1986) *Science*, **232**, p. 343-347.
- [27] Kent, S.B.H. (1988) *Ann. Rev. Biochem.*, **57**, p. 957-989.
- [28] Fitzgerald, M.C. and Kent, S.B.H. (1998) *Total Chemical Synthesis of Proteins*, in *Bioorganic Chemistry: Peptides and Proteins*, Hecht, S.M. Editor. Oxford University Press: New York, p. 65-99.
- [29] Kaiser, E.T., Mihara, H., Laforet, G.A., Kelly, J.W., Walters, L., Findeis, M.A. and Sasaki, T. (1989) *Science*, **243**, p. 187-192.
- [30] Sieber, P., Kamber, B., Hartmann, A., Jöhl, A., Riniker, B. and Rittel, W. (1974) *Hel. Chim. Acta*, **57**, p. 2617-2621.
- [31] Sieber, P., Eisler, K., Kamber, B., Riniker, B., Rittel, W., Märki, F. and de Gasparo, M. (1978) *Physiol. Chem.*, **359**, p. 113-123.
- [32] Akaji, K., Fujii, N., Yajima, H., Hayashi, K., Mizuta, K., Aono, M. and Moriga, M. (1985) *Chem. Pharm. Bull.*, **33**(1), p. 184-201.
- [33] Heath, W.F. and Merrifield, R.B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, p. 6367-6371.
- [34] Blake, J., Yamashiro, D. Ramasharma, K. and Li, C.H. (1986) *Int. J. Peptide Protein Res.*, **28**, p. 468-476.
- [35] Clark-Lewis, I., Aebersold, R., Ziltener, H., Schrader, J.W., Hood, L.E. and Kent, S.B.H. (1986) *Science*, **231**, p. 134-139.
- [36] Schneider, J. and Kent, S.B.H. (1988) *Cell*, **54**, p. 363-368.
- [37] Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J. and Kent, S.B.H. (1989) *Science*, **245**, p. 616-621.
- [38] Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.-T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L. and Springer, J.P. (1989) *Nature*, **337**, p. 615-620.
- [39] Humphrey, J.M. and Chamberlin, A.R. (1997) *Chemical Reviews*, **97**(6), p. 2243-2266.
- [40] Halab, L., Gosselin, F. and Lubell, W.D. (2000) *Biopolymers (Peptide Science)*, **55**, p. 101-122.
- [41] Gisin, B.F., Davis, D.G., Borowska, Z.K., Hall, J.E. and Kobayashi, S. (1981) *J. Am. Chem. Soc.*, **103**, p. 6373-6377.
- [42] Veber, D.F., Holly, F.W., Nutt, R.F., Bergstrand, S.J., Brady, S.F., Hirschmann, R., Glitzer, M.S. and Saperstein, R. (1979) *Nature*, **280**, p. 512-514.
- [43] Freidinger, R.M., Veber, D.F., Perlow, D.S., Brooks, J.R. and Saperstein, R. (1980) *Science*, **210**, p. 656-658.
- [44] Rajashekhar, B. and Kaiser, E.T. (1986) *J. Biol. Chem.*, **261**(29), p. 13617-13623.
- [45] Blanc, J.P. and Kaiser, E.T. (1984) *J. Biol. Chem.*, **259**(15), p. 9549-9556.
- [46] Kaiser, E.T. and Kézdy, F.J. (1984) *Science*, **223**, p. 249-255.
- [47] Veliçelebi, G., Patthi, S. and Kaiser, E.T. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, p. 5397-5399.
- [48] Schnolzer, M. and Kent, S.B.H. (1992) *Science*, **256**, p. 221-224.
- [49] Braisted, A.C., Judice, J.K. and Wells, J.A. (1997) *Methods in Enzymology*, **289**, p. 298-313.
- [50] Cotton, G.J. and Muir, T.W. (1999) *Chemistry & Biology*, **6**(9), p. R247-R256.
- [51] Dawson, P.E. and Kent, S.B.H. (2000) *Annual Review of Biochemistry*, **69**, p. 923-960.
- [52] Offord, R.E. (1987) *Protein Eng.*, **1**(3), p. 151-157.
- [53] Wallace, C.J.A. ed. (1999) *Protein Engineering by Semi-synthesis*, CRC Press: Boca Raton.
- [54] Wang, B., Lodder, M., Zhou, J., Baird, T.T., Brown, K.C., Craik, C.S. and Hecht, S.M. (2000) *J. Amer. Chem. Soc.*, **122**, p. 7402-3.
- [55] Tanuichi, H., Parr, G.R. and Juillerat, M.A. (1986) *Meth. Enzymol.*, **131**, p. 185-217.
- [56] Tschesche, H., Beckmann, J., Mehlich, A., Schnabel, E., Truscheit, E. and Wenzel, H.R. (1987) *Biochim. Biophys. Acta*, **913**, p. 97-101.
- [57] Fruton, J.S. (1982) *Adv. Enzymol.*, **53**, p. 239-306.
- [58] Jakubke, H.-D., Kuhl, P. and Könnecke, A. (1985) *Angew. Chem. Int. Ed. Eng.*, **24**(2), p. 85-93.
- [59] Jakubke, H.-D. (1987) *Enzymatic peptide synthesis*, In *The Peptides: Analysis, Synthesis, Biology*, Meienhofer J. and Udenfriend, S. Editors. Academic Press: Orlando. p. 103-165.
- [60] Homandberg, G.A., Mattis, J.A. and Laskowski, M. (1978) *Biochemistry*, **17**(24), p. 5220-5227.

- [61] Barbas, C.F., Matos, J.R., West, J.B. and Wong, C.-H. (1988) *J. Am. Chem. Soc.*, **110**, p. 5162-5166.
- [62] Vogel, K., Cook, J. and Chmielewski, J. (1996) *Chemistry and Biology*, **3**, p. 295-9.
- [63] Abrahmsen, L., Tom, J., Burnier, J., Butcher, K.A., Kossiakoff, A. and Wells, J.A. (1991) *Biochemistry*, **30**, p. 4151-4159.
- [64] Chang, T.K., Jackson, D.Y., Burnier, J.P. and Wells, J.A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, p. 12544-12548.
- [65] Jackson, D.Y., Burnier, J., Quan, C., Stanley, M., Tom, J. and Wells, J.A. (1994) *Science*, **266**(5183), p. 243-247.
- [66] Atwell, S. and Wells, J.A. (1999) *Proc. Natl. Acad. Sci. USA*, **96**(17), p. 9497-502.
- [67] Dawson, P.E., Muir, T.W., Clark-Lewis, I. and Kent, S.B.H. (1994) *Science*, **266**(5186), p. 776-779.
- [68] Canne, L.E., Bark, S.J. and Kent, S.B.H. (1996) *J. Am. Chem. Soc.*, **118**(25), p. 5891-5896.
- [69] Canne, L.E., Botti, P., Simon, R.J., Chen, Y.J., Dennis, E.A. and Kent, S.B.H. (1999) *J. Am. Chem. Soc.*, **121**(38), p. 8720-8727.
- [70] Xu, M.Q. and Perler, F.B. (1996) *EMBO J.*, **15**(19), p. 5146-5153.
- [71] Chong, S.R., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A., Pelletier, J.J., Paulus, H. and Xu, M.Q. (1997) *Gene*, **192**(2), p. 271-281.
- [72] Severinov, K. and Muir, T.W. (1998) *J. Biol. Chem.*, **273**(26), p. 16205-16209.
- [73] Muir, T.W., Sondhi, D. and Cole, P.A. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, p. 6705-6710.
- [74] Cotton, G.J., Ayers, B., Xu, R. and Muir, T.W. (1999) *J. Am. Chem. Soc.*, **121**(5), p. 1100-1101.
- [75] Cotton, G.J. and Muir, T.W. (2000) *Chemistry & Biology*, **7**(4), p. 253-261.
- [76] Blaschke, U.K., Cotton, G.J. and Muir, T.W. (2000) *Tetrahedron*, **56**(48), p. 9461-9470.
- [77] Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B. and Zinoni, F. (1991) *Mol. Microbiol.*, **5**(3), p. 515-20.
- [78] Stadtman, T.C. (1996) *Annu. Rev. Biochem.*, **65**, p. 83-100.
- [79] Rodi, D.J. and Makowski, L. (1999) *Curr. Opin. Biotechnol.*, **10**(1), p. 87-93.
- [80] Patten, P.A., Howard, R.J. and Stemmer, W.P.C. (1997) *Pharmaceutical Biotechnology*, **8**, p. 724-733.
- [81] Zhao, H., Giver, L., Shao, Z., Affholter, J.A. and Arnold, F.H. (1998) *Nature Biotech.*, **16**, p. 258-61.
- [82] Schmidt-Dannert, C. and Arnold, F.H. (1999) *TIB Tech.*, **17**, p. 135-6.
- [83] Sandman, K.E., Benner, J.S. and Noren, C.J. (2000) *J. Am. Chem. Soc.*, **122**(5), p. 960-1.
- [84] Cunningham, B.C. and Wells, J.A. (1997) *Curr. Opin. Structl. Biol.*, **7**, p. 457-462.
- [85] Struthers, M.D., Cheng, R.P. and Imperiali, B. (1996) *Science*, **271**, p. 342-345.
- [86] Dahiyat, B.I. and Mayo, S.L. (1997) *Science*, **278**, p. 82-87.
- [87] Li, B., Tom, J.Y.K., Oare, D., Yen, R., Fairbrother, W.J., Wells, J.A. and Cunningham, B.C. (1995) *Science*, **270**, p. 1657-60.
- [88] Chin, J.W. and Schepartz, A. (2001) *J. Am. Chem. Soc.*, **123**, p. 2929-30.
- [89] Nguyen, J.T., Turck, C.W., Cohen, F.E., Zuckermann, R.N. and Lim, W.A. (1998) *Science*, **282**, p. 2088-92.
- [90] Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C. and Schultz, P.G. (1989) *Science*, **244**, p. 182-188.
- [91] Bain, J.D., Glabe, C.G., Dix, T.A., Chamberlin, A.R. and Diala, E.S. (1989) *J. Am. Chem. Soc.*, **111**, p. 8013-8014.
- [92] Bossi, L. and Roth, J.R. (1980) *Nature*, **286**, p. 123-127.
- [93] Kim, D.M., Kigawa, T., Choi, C.Y. and Yokoyama, S. (1996) *European Journal of Biochemistry*, **239**(3), p. 881-886.
- [94] Cload, S.T., Liu, D.R., Froland, W.A. and Schultz, P.G. (1996) *Chemistry & Biology*, **3**(12), p. 1033-1038.
- [95] 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**(1), p. 6-11.
- [96] Thorson, J.S., Cornish, V.W., Barrett, J.E., Cload, S.T., Yano, T. and Schultz, P.G., Martin, R. Editor. (1998) *Protein Synthesis: Methods and Protocols*. Humana Press: Totowa, NJ. p. 43-73.
- [97] Short, G.F., Golovine, S.Y. and Hecht, S.M. (1999) *Biochemistry*, **38**(27), p. 8808-8819.
- [98] Magliery, T.J., Pastrnak, M., Anderson, J.C., Santoro, S.W., Meggers, E., Herberich, B., Wang, L. and Schultz, P.G. (2002) *Incorporation of unnatural amino acids into proteins in Translation Mechanisms*, Jacques Lapointe (ed), Landes, Bioscience., in press.
- [99] Kool, E.T. (2000) *Curr. Opin. Chem. Biol.*, **4**(6), p. 602-608.
- [100] Magliery, T.J., Anderson, J.C. and Schultz, P.G. (2001) *J. Mol. Biol.*, **307**, p. 755-769.
- [101] Cornish, V.W., Mendel, D. and Schultz, P.G. (1995) *Ang. Chem. Intl. Ed. Eng.*, **34**(6), p. 621-633.
- [102] Dougherty, D.A. (2000) *Curr. Opin. Chem. Biol.*, **4**(6), p. 645-652.
- [103] Ellman, J.A., Mendel, D. and Schultz, P.G. (1992) *Science*, **255**(5041), p. 197-200.
- [104] Mendel, D., Ellman, J. and Schultz, P.G. (1993) *J. Am. Chem. Soc.*, **115**(10), p. 4359-4360.
- [105] Mendel, D., Ellman, J.A., Chang, Z.Y., Veenstra, D.L., Kollman, P.A. and Schultz, P.G. (1992) *Science*, **256**(5065), p. 1798-1802.

- [106] Thorson, J.S., Chapman, E. and Schultz, P.G. (1995) *J. Am. Chem. Soc.*, **117**(36), p. 9361-9362.
- [107] Thorson, J.S., Chapman, E., Murphy, E.C., Schultz, P.G. and Judice, J.K. (1995) *J. Am. Chem. Soc.*, **117**(3), p. 1157-1158.
- [108] Ting, A.Y., Shin, I., Lucero, C. and Schultz, P.G. (1998) *J. Am. Chem. Soc.*, **120**(28), p. 7135-7136.
- [109] Zhong, W.G., Gallivan, J.P., Zhang, Y.N., Li, L.T., Lester, H.A. and Dougherty, D.A. (1998) *Proc. Natl. Acad. Sci. USA*, **95**(21), p. 12088-12093.
- [110] Ellman, J.A., Volkman, B.F., Mendel, D., Schultz, P.G. and Wemmer, D.E. (1992) *J. Am. Chem. Soc.*, **114**(20), p. 7959-7961.
- [111] 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**(8), p. 2910-2914.
- [112] Cornish, V.W., Hahn, K.M. and Schultz, P.G. (1996) *J. Am. Chem. Soc.*, **118**(34), p. 8150-8151.
- [113] Gallivan, J.P., Lester, H.A. and Dougherty, D.A. (1997) *Chemistry & Biology*, **4**(10), p. 739-749.
- [114] Murakami, H., Hohsaka, T., Ashizuka, Y. and Sisido, M. (1998) *J. Am. Chem. Soc.*, **120**(30), p. 7520-7529.
- [115] Mendel, D., Ellman, J.A. and Schultz, P.G. (1991) *J. Am. Chem. Soc.*, **113**(7), p. 2758-2760.
- [116] Miller, J.C., Silverman, S.K., England, P.M., Dougherty, D.A. and Lester, H.A. (1998) *Neuron*, **20**(4), p. 619-624.
- [117] England, P.M., Lester, H.A., Davidson, N. and Dougherty, D.A. (1997) *Proc. Natl. Acad. Sci. USA*, **94**(20), p. 11025-11030.
- [118] Short, G.F., Lodder, M., Laikhter, A.L., Arslan, T. and Hecht, S.M. (1999) *J. Am. Chem. Soc.*, **121**(2), p. 478-479.
- [119] Pollitt, S.K. and Schultz, P.G. (1998) *Ang. Chem. Intl. Ed. Eng.*, **37**(15), p. 2104-2107.
- [120] Nowak, M.W., Kearney, P.C., Sampson, J.R., Saks, M.E., Labarca, C.G., Silverman, S.K., Zhong, W.G., Thorson, J., Abelson, J.N., Davidson, N., Schultz, P.G., Dougherty, D.A. and Lester, H.A. (1995) *Science*, **268**(5209), p. 439-442.
- [121] 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**(38), p. 23169-23175.
- [122] Budisa, N., Minks, C., Alefelder, S., Wenger, W., Dong, F.M., Moroder, L. and Huber, R. (1999) *FASEB J.*, **13**(1), p. 41-51.
- [123] Hendrickson, W.A., Horton, J.R. and Lemaster, D.M. (1990) *EMBO J.*, **9**(5), p. 1665-1672.
- [124] Geisow, M. (1991) *Trends Biotech.*, **9**(1), p. 4-5.
- [125] Budisa, N., Steipe B., Demange, P., Eckerskorn, C., Kellermann, J. and Huber, R. (1995) *Eur. J. Biochem.*, **230**(2), p. 788-796.
- [126] Jakubowski, H. (2000) *J. Biol. Chem.*, **275**(29), p. 21813-21816.
- [127] Budisa, N., Huber, R., Golbik, R., Minks, C., Weyher, E. and Moroder, L. (1998) *Eur. J. Biochem.*, **253**(1), p. 1-9.
- [128] Browne, D.R., Kenyon, G.L. and Hegeman, G.D. (1970) *Biochem. Biophys. Res. Comm.*, **39**(1), p. 13-9.
- [129] Kim, H.W., Perez, J.A., Ferguson, S.J. and Campbell, I.D. (1990) *FEBS Lett.*, **272**(1-2), p. 34-36.
- [130] Parsons, J.F., Xiao, G.Y., Gilliland, G.L. and Armstrong, R.N. (1998) *Biochemistry*, **37**(18), p. 6286-6294.
- [131] Luck, L.A. and Falke, J.J. (1991) *Biochemistry*, **30**(26), p. 6484-6490.
- [132] Luck, L.A., Vance, J.E., Oconnell, T.M. and London, R.E. (1996) *J. Biomolec. NMR.*, **7**(4), p. 261-272.
- [133] Minks, C., Huber, R., Moroder, L. and Budisa, N. (1999) *Biochemistry*, **38**(33), p. 10649-10659.
- [134] Ibba, M. and Soll, D. (2000) *Ann. Rev. Biochem.*, **69**, p. 617-650.
- [135] Liu, D.R., Magliery, T.J., Pasternak, M. and Schultz, P.G. (1997) *Proc. Natl. Acad. Sci. USA*, **94**(19), p. 10092-10097.
- [136] Liu, D.R., Magliery, T.J. and Schultz, P.G. (1997) *Chemistry & Biology*, **4**(9), p. 685-691.
- [137] Schimmel, P. and Soll, D. (1997) *Proc. Natl. Acad. Sci. USA*, **94**(19), p. 10007-10009.
- [138] Whelihan, E.F. and Schimmel, P. (1997) *EMBO J.*, **16**(10), p. 2968-2974.
- [139] Wang, C.C. and Schimmel, P. (1999) *J. Biol. Chem.*, **274**(23), p. 16508-16512.
- [140] Liu, D.R. and Schultz, P.G. (1999) *Proc. Natl. Acad. Sci. USA*, **96**(9), p. 4780-4785.
- [141] Huang, W.Z., Petrosino, J., Hirsch, M., Shenkin, P.S. and Palzkill, T. (1996) *J. Mol. Biol.*, **258**(4), p. 688-703.
- [142] Ibba, M., Kast, P. and Hennecke, H. (1994) *Biochemistry*, **33**(23), p. 7107-7112.
- [143] Ibba, M. and Hennecke, H. (1995) *FEBS Lett.*, **364**(3), p. 272-275.
- [144] 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**(48), p. 9443-9449.
- [145] Agou, F., Quevillon, S., Kerjan, P. and Mirande, M. (1998) *Biochemistry*, **37**(32), p. 11309-11314.
- [146] Hong, K.W., Ibba, M. and Soll, D. (1998) *FEBS Lett.*, **434**(1-2), p. 149-154.
- [147] Ibba, M., Hong, K.-W. and Soll, D. (1996) *Genes to Cells*, **1**, p. 421-427.
- [148] Rath, V.L., Silvian, L.F., Beijer, B., Sproat, B.S. and Steitz, T.A. (1998) *Structure*, **6**(4), p. 439-449.
- [149] Capone, J.P., Sedivy, J.M., Sharp, P.A. and RajBhandary, U.L. (1986) *Mol. Cell. Biol.*, **6**(9), p. 3059-67.
- [150] Pastrnak, M. and Schultz, P.G. (2001) *Bioorg. Med. Chem.*, **9**(9), 2373-9.
- [151] Furter, R. (1998) *Prot. Sci.*, **7**(2), p. 419-426.

- [152] Wang, L., Brock, A., Herberich, B. and Schultz, P.G. (2001) *Science*, **292**, p. 498-500.
- [153] Wang, L., Magliery, T.J., Liu, D.R. and Schultz, P.G. (2000) *J. Am. Chem. Soc.*, **122**(20), p. 5010-5011.
- [154] Kowal, A.K., Kohrer, C. and RajBhandary, U.L. (2001) *Proc. Natl. Acad. Sci. USA*, **98**(5), p. 2268-2273.
- [155] Pastrnak, M., Magliery, T.J. and Schultz, P.G. (2000) *Helv. Chim. Acta*, **83**(9), p. 2277-2286.