

Unnatural Protein Engineering: Producing Proteins with Unnatural Amino Acids

Thomas J. Magliery*

Yale University, Department of Molecular Biophysics & Biochemistry, P.O. Box 208114, New Haven, CT 06520-8114, USA

Abstract: Less than a decade ago, the ability to generate proteins with unnatural modifications was a Herculean task available only to specialty labs. Recent advances make it possible to generate reasonable quantities of protein with unnatural amino acids both *in vitro* and *in vivo*. The combination of solid-phase peptide synthesis and enzymatic or chemoselective ligation now permits construction of entirely-synthetic proteins as large as 25 kD. Incorporation of recombinant fragments (expressed protein ligation) allows unnatural modifications near protein termini in proteins of virtually any size. Site-specific modification of small quantities of protein at any position can be achieved using chemically-acylated tRNA. Microelectroporation now extends this method to cells like mammalian neurons, and combination with RNA-display makes unnatural proteins compatible with combinatorial methods. Widespread or residue-specific methods of amino acid replacement are especially suitable for the production of biomaterials, and bacteria have been engineered to expand the repertoire of amino acids available for this technique. Excitingly, the wholesale addition of engineered tRNAs and synthetases to bacteria and yeast now makes site-specific incorporation of unnatural amino acids possible in living cells with no chemical intervention. Methods of expanding the genetic code at the nucleic acid level, including 4-base codons and unnatural base pairs, are becoming useful for the addition of multiple amino acids to the genetic code. These recent advances in unnatural protein engineering are reviewed with an eye toward future challenges, including methods of creating nonpeptidic molecules using templated synthesis.

Keywords: Unnatural amino acids; protein ligation; aminoacyl-tRNA synthetase; genetic code; amino acid replacement; site-specific incorporation.

INTRODUCTION

One of the great paradoxes of protein science is that the remarkable diversity of protein structure and function arises from just twenty amino acids, but that this small amino acid repertoire also limits our ability to study proteins. (Occasionally, selenocysteine [1] or pyrrolysine [2, 3] are also translationally incorporated.) Nature did not provide many of the kinds of amino acids that would be most useful to biochemists and biophysicists. Tryptophan fluorescence is poor, and there are no amino acids with unpaired spins. The amino acids tend to differ in stereoelectronic (such as pK_A) and steric properties so significantly that it is difficult to interpret the effects of exchange of one residue for another. Nature has solved the problem of specificity in posttranslational modification using enzymatic recognition of extended motifs rather than unique reactive handles. There is nothing useful for trapping interacting proteins. There are neither alkenes nor alkynes. There are no useful electrophiles—not a ketone or an aldehyde. The heaviest atom to be found is a sulfur, and of course isotopes useful for NMR are only at natural abundance.

Those taking a chemical approach to biology have known about this issue for some time. Without doubt, to this day,

the most important tool for engineering approaches to understanding protein structure and function is site-directed mutagenesis, which allows us to swap one natural amino acid for another at a specific site in a protein [4, 5]. While there is no access to unnatural side chains, there is the substantial importance of ease of creation of the mutant protein at essentially any scale necessary. Modern synthesis of DNA oligonucleotides can provide 100-mers or more, and PCR-based methods make it simple to stitch together genes of essentially any sequence. In appropriate vectors, these genes can be used to express or overexpress proteins in bacteria, yeast, insect cells, mammalian cells or even living animals.

Until recently, no method of incorporating unnatural amino acids into proteins could compete in specificity, simplicity or scale with the ability to trade natural amino acids with site-directed mutagenesis. Chemical synthesis of peptides, allowing every monomer to be altered at will, was only practical for the creation of small quantities of peptides at lengths barely corresponding to the smallest proteins or domains known. Co-opting the protein synthesis machinery allowed access to proteins of virtually any size, but there were significant tradeoffs to be accepted. Near-natural amino acid homologs (like selenomethionine) could sometimes be substituted for natural amino acids in living cells, but few suitable amino acids were known, and the substitution could occur at any position with the related natural residue, although it often occurred with less than 100% replacement. On the other hand, chemical acylation of *amber* stop codon-

*Address correspondence to this author at the Yale University, Department of Molecular Biophysics & Biochemistry, P.O. Box 208114, New Haven, CT 06520-8114, USA; Tel: 203-432-9841; Fax: 203-432-5175; E-mail thomas.magliery@yale.edu

suppressing tRNA with unnatural amino acids permitted site-specific incorporation of virtually any amino acid, but it requires a substantial synthetic undertaking and resulted in poor yields of protein from *in vitro* translation mixtures. In order to achieve this same type of incorporation in living cells, microinjection of RNA and chemically-acylated tRNA was required, essentially restricting the user to cells like the large *Xenopus* oocyte and vanishingly small amounts of protein.

This is not to say that these methods were not useful, but just that they were not practical for the typical biochemistry or biophysics lab to undertake, and of course these are the fields that stand to benefit most from unnatural amino acid incorporation. Fortunately, the last several years have seen tremendous strides in methods to incorporate unnatural amino acids into proteins. Proteins of useful size can now be generated synthetically or semi-synthetically, using strategies to ligate smaller polypeptides. The scale and utility of chemically-acylated tRNA methods has improved, including the ability to produce proteins in mammalian cells, such as neurons. The number of amino acids that can be inserted by widespread (residue-specific) replacement has expanded, partially through enzyme engineering approaches, and the methods are sufficiently reliable that nearly 100% incorporation can be achieved. But perhaps most significantly, the protein biosynthetic machineries of bacteria and yeast have been engineered to allow site-specific insertion of an expanding array of unnatural amino acids without chemical acylation of tRNA. This has raised the question of expanding the genetic code at the DNA level, and a number of useful insertion signals including 4-base codons and unnatural base pairs have been introduced. Finally, this sort of reprogramming of the templated synthesis of proteins has inspired the templated synthesis of other kinds of molecules, as well.

My colleagues and I wrote several reviews about the state of unnatural protein engineering in the last several years [6-8], and this article is intended to be an update to those reviews. A number of excellent reviews of this topic have appeared more or less recently, and they offer some substantial insights that are beyond the focus of this article [9-12].

CHEMICAL SYNTHESIS AND SEMI-SYNTHESIS OF PROTEINS

Modern synthetic methods provide access to essentially any peptide less than about 50 residues in length, but this is only the size of the smallest proteins known (about 6 kD) [13]. Solid-phase peptide synthesis (SPPS) has solved a number of problems related to the insolubility of protected peptides and the difficulty of driving step-wise reactions to near completion, and it has made peptides widely available through automated procedures. However, even if every coupling cycle could result in a 99% yield of desired product, the overall yield would only be about 60% after 50 cycles, or 37% after 100 cycles. Thus, the step-wise formation of extremely small amounts of side-products makes it essentially impossible to substantially increase the scope of peptide synthesis. A general solution to this problem is to ligate a number of peptides into a protein of desired length. Two different strategies (Fig. 1) have been introduced to this end: enzymic ligation and chemoselective ligation (especially

native chemical ligation, also called expressed protein ligation when one of the fragments is recombinant).

Enzymic ligation of synthetic peptides is achieved by treatment of peptides with proteolytic enzymes under conditions where hydrolysis is not favored. Typically, a C-terminal ester derivative of one protein is amidated by the N-terminal amine of a second peptide in a medium high in organic solvent [14]. The relatively dry conditions favor the reaction of the N-terminal amide with the acyl-enzyme intermediate in the active site of the protease. Probably the most effective enzymic ligation system is a modified subtilisin, called subtiligase, and its derivatives, which have significantly reduced rates of proteolysis and therefore allow reaction in aqueous media [15-19]. Enzymatic ligation has been used to incorporate moieties such as biotin, heavy atoms and sugars into synthetic or partially synthetic proteins. However, this method has seen less use in the last several years, in large part due to peptide solubility issues, the fact that proteases have sequence requirements that limit the generality of this approach, and the residual spurious proteolytic activity of even modified enzymes.

Chemoselective reactions have found considerable utility in assembling fully synthetic proteins, including those bearing unnatural amino acids or posttranslational modifications that would be impossible to obtain uniformly (or at all) from cellular expression. Here, the strategy is to ligate two unprotected peptide fragments using some chemistry that is orthogonal to the intrinsic reactivities of the amino acids. The original applications of the strategy resulted in an unnatural backbone linkage (such as a thioester) [20]. A significant improvement in this methodology, introduced by Kent and coworkers, involves the reaction of a peptide bearing an N-terminal cysteine with a second peptide bearing a C-terminal thioester. The cysteinyl side chain first transthioesterifies onto the second peptide, and a spontaneous S→N acyl shift ensues to yield the native amide linkage (hence, the name 'native chemical ligation'). The reaction occurs at neutral pH in aqueous medium, usually under denaturing conditions to prevent structure-related impediments to reaction [13, 21].

Three difficulties associated with this methodology are the creation of the C-terminal thioester, the requirement for the N-terminal cysteine (and hence a Xxx-Cys bond in the mature protein) and the need to sequentially ligate several peptides to yield proteins of useful length [22]. The creation of Fmoc-based SPPS methodology for the introduction of the -thioester has made it more compatible with typical commercial apparatus and acid-sensitive moieties like phosphate groups [23]. (The original methodology relied on Boc-SPPS, which uses strong acid in deprotection.) To remove the requirement for N-terminal cysteine, a number of traceless auxiliaries have been introduced that can be mildly excised after ligation [24, 25], and approaches for desulfurization or deseleniumization (if SeCys is used instead of Cys) have also been employed [26-28]. Both solution-phase [29] and solid-phase [30] methods have been demonstrated to allow sequential ligations of multiple peptides, and wholly-synthetic proteins as large as 25 kD are accessible as a result.

A number of other chemoselective chemistries have also been used to assemble proteins and protein derivatives, per-

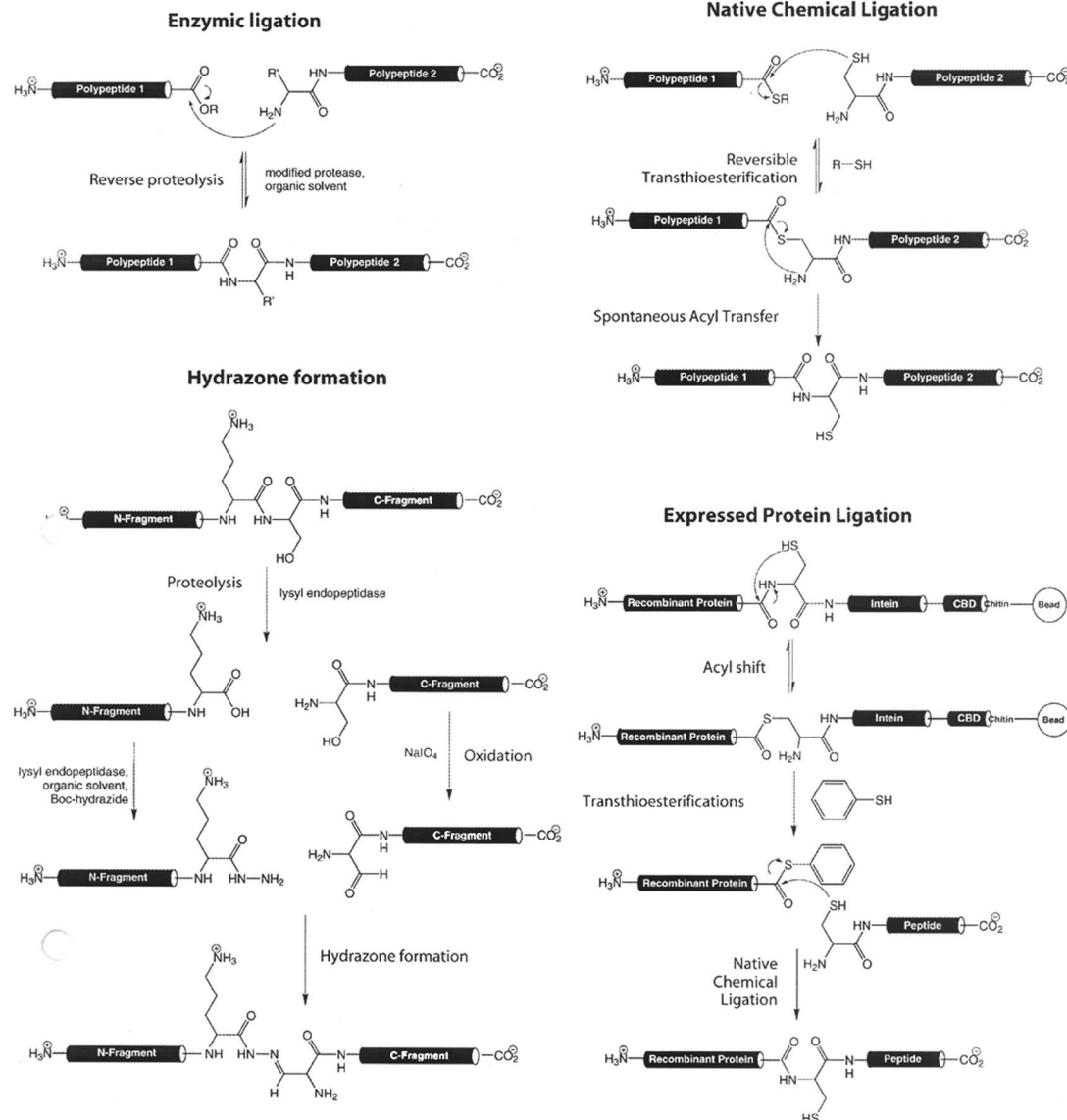


Fig. (1). Peptide ligation strategies. Schematic representation of enzymic ligation by reverse proteolysis, chemoselective ligation by hydrazone formation, native chemical ligation, and expressed protein ligation (EPL). Hydrazone formation could be carried out on wholly synthetic fragments, but the steps to ligate fragments from expressed proteins are shown.

haps most notably the hydrazone/oxime chemistry of Offord & Rose, wherein an N-terminal aldehyde or ketone (which can be created by selective oxidation of N-terminal Ser or Thr) is attacked with a C-terminal hydrazide or a hydroxylamine [31, 32]. The hydrazide can be introduced into a peptide in a way that is similar to the use of subtiligase, wherein a protease is used to ligate a hydrazine donor at high concentration in partially organic media.

Still, when one runs an *E. coli* lysate out on a gel, there is a conspicuous preponderance of proteins above the 25 kD marker, and this is even more true in eukaryotes. Muir and coworkers have developed an extremely clever and useful method of protein semisynthesis based on native chemical ligation, wherein the γ -thioester is produced recombinantly from a partially-incapacitated protein splicing system [12, 33]. Basically, the protein fragment of interest is expressed with an intein domain fused at its C-terminus (there are

commercially-available kits to do this). That intein domain has a mutation that prevents the intramolecular splicing reaction, and so the γ -thioester can be transthioesterified with some other thiol intermolecularly. A synthetic peptide with an N-terminal Cys can then be used to transthioesterify onto the recombinant fragment, and amide formation follows from acyl rearrangement. This allows one to essentially append a fragment altered at will by chemical synthesis onto the C-terminus of a recombinant protein of virtually any size.

Broadly considered, expressed protein ligation (EPL) can be viewed as any native chemical ligation in which one of the reactants (the γ -thioester or the N-terminal cysteinyl fragment) is synthetic and one is recombinant. As mentioned above, it is possible to introduce either of the appropriate functionalities synthetically, and it is possible to recombinantly produce γ -thioesters. It is also possible to produce recombinant proteins with cysteinyl N-termini, either

through removal of initiating methionine by endogenous aminopeptidases [34], or through protease treatment such that scission occurs at an Xxx-Cys bond [35]. Thus, it is possible to introduce unnatural amino acids within about 50 residues of the N- or C-termini of proteins using a single ligation. There are clearly some limitations, however. The protein has to tolerate the critical Cys at the junction site, or some strategy for alteration of the Cys has to be undertaken. If one wants to make modifications further in than 50 residues, then multiple ligations are required to build the protein, and this rapidly becomes difficult. Also, the ligation achieves its best yields in the absence of steric complications and at high concentrations of reactants, and therefore the identity of Xxx at the Xxx-Cys junction is important [29], and it is helpful if the reaction can occur under denaturing conditions followed by protein "refolding."

Nevertheless, this method has been very useful in inserting functionalities that cannot be introduced translationally into large, interesting proteins. Fluorescence probes, post-translational modifications, isotopes and various unnatural amino acids have been used fairly extensively now [12, 22]. It is worth noting that this method may always be superior for some kinds of modifications, such as the site-specific incorporation of isotopes (for NMR, for example). There will never be a modified aminoacyl-tRNA synthetase that is able to distinguish Lys from [¹⁵N]-Lys (although a protected [¹⁵N]-Lys is in principle not out of the question). Finally, all of these methods produce protein *in vitro*, which makes it considerably less useful for the investigation of cellular function (although some progress has been made in the introduction of protein into cells) [36, 37].

IN VITRO INCORPORATION OF UNNATURAL AMINO ACIDS

Crick's adapter hypothesis stated that tRNAs are merely adapter molecules that connect an amino acid with a codon based on the anticodon of the tRNA [38]. (That is, for the most part, the ribosome does not prevent misacylated tRNAs from delivering their cargo.) Essentially, this says that the fidelity of the genetic code depends on both the codon:anticodon interaction and the acylation of the tRNA with the proper amino acid, which is carried out by enzymes called aminoacyl-tRNA synthetases, or aaRSs. The corollary of this hypothesis is then that changing the genetic code requires two things: a tRNA for specifying the codon, and a method of acylating the tRNA with an amino acid of choice.

The most straightforward way of selecting a codon as an insertion signal is to choose from one of the 64 three-base codons in the standard genetic code. One could select one of the 61 sense codons as an insertion signal, and the usage of some of these in a given organism is sufficiently rare that the amino acid of choice would be inserted at relatively few positions in the proteome. However, a better approach has been to select one of the three stop codons [39], and in particular the *amber* (UAG) stop codon has been successful in the *Escherichia coli* translation system. *Amber* is both the least used stop codon in *E. coli*, and a number of efficient 'suppressor' tRNAs (i.e., those with a CUA anticodon) are known [40]. (Note that throughout this review all nucleic acid sequences are written 5'-to-3'.) There has been some

recent success in generating other robust insertion signals (see below).

The second issue is how to acylate the tRNA bearing the selected anticodon with an amino acid of choice. This is a field with a long history, dating back to the Raney nickel reduction of Cys-tRNA^{Cys} to Ala-tRNA^{Cys} that proved the adapter hypothesis [38, 41]. (Ala is inserted at a Cys codon in an *in vitro* translation mix using the misacylated tRNA.) The state-of-the-art technology for chemical misacylation of tRNA represents the work of the Hecht, Chamberlin and Schultz groups (Fig. 2) [42-49]. The tRNA scaffold missing the last two (3') invariant CA nucleotides is generated by *in vitro* transcription. Then, an unprotected pdCpA molecule is synthesized, and it is acylated with the cyanomethyl ester of the amino acid of choice (where the amine is typically protected with a photolabile NVOC group). The pdCpA-aa-NVOC is then ligated into the tRNA_{CA} using T4 RNA ligase (and the lack of the nucleophilic 2' hydroxyl on C does not disturb either this ligation or the ability of the tRNA to act in translation). Finally, the -amine is deprotected (with photolysis for NVOC) to yield the mature acylated tRNA. This chemistry is widely applicable (Fig. 3), although some amino acids are not compatible with the cyanomethyl activation, and photosensitive amino acids cannot be used with the

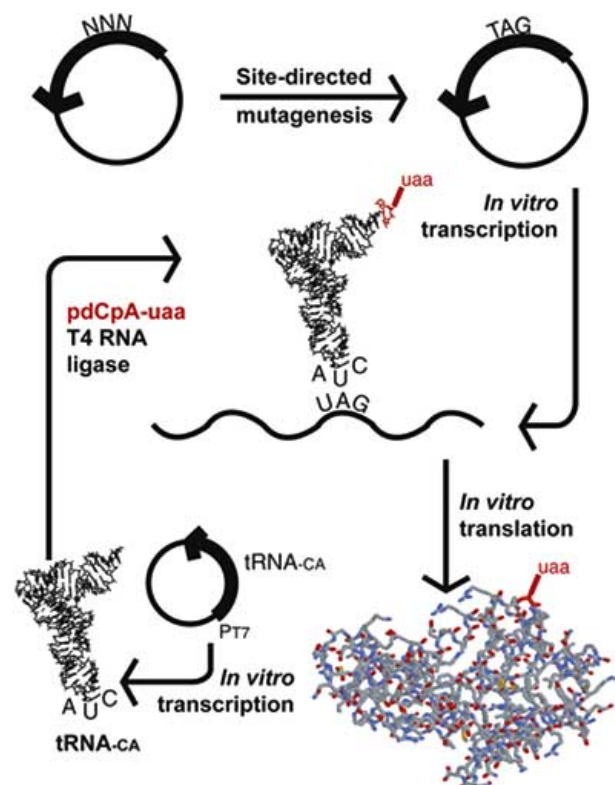


Fig. (2). Cell-free synthesis using chemically-acylated tRNA. The site of unnatural amino acid incorporation is specified at the genetic level by site-directed mutagenesis (here, to an *amber* TAG stop codon). Chemical acylation of tRNA is carried out by *in vitro* transcription of tRNA_{CA}, chemical acylation of pdCpA dinucleotide and ligation of the acylated pdCpA to the immature tRNA. This is then added to an *in vitro* transcription/translation reaction along with the DNA encoding the gene of interest.

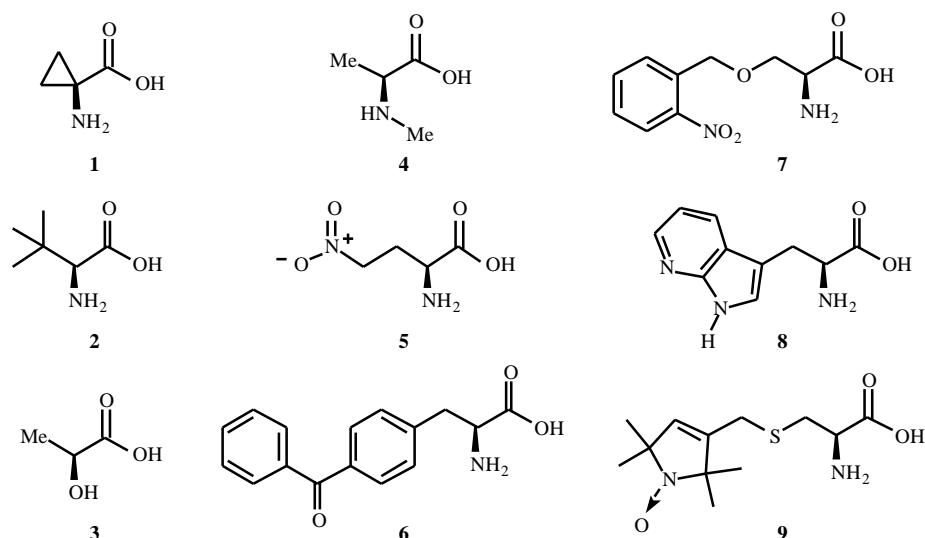


Fig. (3). Examples of amino acids incorporated by chemical acylation of tRNA. Over a hundred amino acids and analogs have been incorporated into proteins by chemical acylation of tRNA. Some types of analogs include sterically restricted amino acids (1-2), peptide backbone mimetics (like α -hydroxy acids, 3, and α -amino alkyl derivatives, 4), amino acids that differ in electronic properties (glutamine analog 5), reactive amino acids (for photoaffinity labeling, 6, or photodecaging, 7), fluorescence probes (8) and spin-labels (9).

NVOC protecting group. Also, not all amino acids are accepted by the translational machinery, likely in part due to the action of elongation factor Tu [50]. The chief liability of this method is that synthesis of both pdCpA and protected, activated amino acid is challenging, and most biochemistry labs are simply not equipped to do this. Ninomiya *et al.* recently showed that the transesterification reaction between the cyanomethyl-activated amino acid and pdCpA can be improved using cationic micelles [51], and pdCpA is now available commercially, but the method is still expensive and difficult.

Another approach to *in vitro* aminoacylation of tRNA with unnatural amino acid is to evolve a ribozyme for this purpose [52, 53]. The method is particularly attractive, since the tRNA specificity can be altered rationally. Although this approach is less general than synthetic acylation in the sense that one might have to evolve a separate ribozyme for a given amino acid, it may provide a tool that is simple to use and applicable to a wide array of amino acids. This approach still requires amino acid activation (as the cyanomethyl ester), but it avoids the pdCpA synthesis and ligation steps, and can be accomplished without α -amine protection. Recently, a resin-immobilized ribozyme was introduced which has broad tRNA aminoacylation specificity toward cyanomethyl esters of phenylalanyl analogs [54]. This system makes *in vitro* acylation with unnatural amino acids considerably easier and faster than purely synthetic routes.

The final issue is how to get a tRNA bearing an anticodon (CUA) and unnatural amino acid of choice to function in translation. Either one has to get the acylated tRNA into the cell (see below), or one has to generate the protein *in vitro* using an transcription/translation extract mixture. Of course, the nature of the tRNA scaffold depends upon the choice of protein biosynthetic apparatus, since one must employ a tRNA that is neither acylated nor deacylated by en-

dogenous aaRSs, but that is competent to act in translation (i.e., binds to elongation factors, is accepted by the ribosome, etc.). The most effective system to date is based on *E. coli* extract, using *amber*-suppressing tRNA derived from yeast tRNA^{Phe} or, for small, polar amino acids, *E. coli* tRNA^{Asn} [44, 55]. The efficiency of *amber* suppression, and *in vitro* translation generally, have been improved considerably recently by inactivation of the release factor that recognizes *amber* stop codons, and by optimization of the cell-free synthesis system, respectively [56-58]. The latter is accomplished by continuous dialysis of the transcription/translation reaction to replenish necessary small molecules and remove inhibitory byproducts, resulting in yields as high as 6 mg mL⁻¹ (for normal translation, not *amber* suppression).

A last point worth mentioning is that the template (mRNA) for protein synthesis also needs to be present, obviously, with the *amber* mutation in the appropriate position. This is straightforward for cell-free transcription/translation, since one can simply add a plasmid or even synthetic DNA coding for the protein to the mixture. Site-directed mutagenesis can be used to create the mutation in the plasmid, and use of the T7 promoter allows one to simply add T7 RNA polymerase to the cell-free synthetic system to achieve large amounts of mRNA. The importance of this issue will be clear with *in vivo* systems or alternate insertion signals, discussed below.

Chemical acylation of tRNA offers the considerable advantage over synthetic methods that it is applicable to (soluble) proteins of virtually any size with the desired amino acid inserted site-specifically at any position in the protein (regardless of distance from the termini). It does not require denaturing conditions to produce the protein, so refolding is less likely to be necessary. Unlike *in vivo* residue-selective or specific methods (below), it is totally site-specific (although it allows the insertion of only a single unnatural

amino acid using the *amber* stop codon). It is slightly more limited than synthetic methods in terms of generality with respect to amino acid, but it is much broader in scope than any *in vivo* method right now. However, the liabilities of this method are substantial: it is synthetically challenging, the yields are generally poor (and thus it is labor-intensive and expensive to generate more than microgram quantities of protein), and, like any *in vitro* method, it does not allow direct investigation of cellular function.

mRNA Display

One of the serious issues with *in vitro* production of unnatural proteins is the question of what one can actually do with the unnatural protein once it is in hand. It is difficult to produce large amounts of protein, so it is consequently difficult (or expensive, or both) to carry out material-hungry biophysical characterization like CD spectroscopy, NMR, or X-ray crystallography. (All of these experiments have been done, however.) One can carry out tests for binding or enzymatic activity, which require only a small amount of protein, but combinatorial methods are essentially not an option. The protein is not linked to its genetic material and thus would be impossible to identify in a mixture.

An elegant solution to the *in vitro* production of mRNA-linked proteins has substantially expanded the scope of *in vitro* unnatural protein engineering. Essentially, one can produce mRNA that ends with a puromycin (Pur) moiety, which acts as an aminoacyl-tRNA surrogate at the end of translation, covalently linking the protein to the encoding mRNA [59, 60]. This puromycin-linked RNA can be produced totally synthetically, or transcribed from DNA *in vitro* and then linked to a short RNA-Pur fusion using T4 RNA ligase (and a DNA splint). This mRNA-Pur fusion can then be added to a cell-free translation system. The resulting protein, which is labeled with its encoding RNA, can then be subjected to screening (for example, binding to an immobilized ligand) and identified by RT-PCR from as little as, in principle, a single molecule [61].

Of course, this protein has a large RNA appended to it, which is a complicating factor. One would like to know that one is selecting for a property of the protein, not the RNA, so it is necessary to reverse transcribe the RNA to produce an "inert" double-stranded nucleic acid. But this can hardly be thought of as an inconsequential spectator: a 100-residue protein (13 kD) would bear 300 residues of dsRNA/DNA hybrid (195 kD), and at minimum this is likely to have a profound effect on the solubility of the fusion (i.e., there is essentially no selection for soluble protein since the highly-charged fusion is virtually guaranteed to be soluble). Also, the cell-free system that has been the most efficient is based on rabbit reticulocytes rather than *E. coli* extract, and therefore one must find a tRNA scaffold and insertion signal that is applicable for delivery of an unnatural amino acid. Fortunately, it was known that THG73, a modified tRNA^{Gln} from *Tetrahymena thermophila*, is an efficient suppressor of *amber* in other eukaryotic systems, and other insertion signals have been explored (see below for both). Unnatural amino acids like biocytin (lysyl biotin) and *N*-methyl-Phe have been inserted this way, and selection of biocytin-containing fusions using streptavidin has been demonstrated, making

this a promising approach [62, 63]. A related technology, ribosome display (in which protein and mRNA remain on-board stalled ribosomes), is also potentially amenable to this sort of approach [64].

RECODING – BEYOND AMBER SUPPRESSION

Amber stop codon suppression has been the workhorse of site-specific biosynthetic unnatural protein engineering, but there has been significant development of other insertion signals. The motivation for this is quite simple: one would like to be able to insert more than one unnatural amino acid into a single protein site-specifically (as one can do with total synthesis). There are also more subtle motivations. *Amber* suppression is not generally as efficient as normal insertion of amino acids, presumably in large part due to competition with the release factor, and some sites are not amenable to any significant level of suppression. Moreover, there is the problem of read-through, which is caused by the misreading of the *amber* stop codon by another tRNA (usually with an anticodon similar to CUA). This results in competitive protein products in which the selected site is occupied by one of the natural amino acids. Also, although *amber* suppression is possible in eukaryotic systems, *amber* stop codons are used much more frequently in eukaryotes. Essentially, there are four alternatives to the *amber* stop codon: another stop codon, a sense codon, an extended (e.g., 4-base) codon, or an unnatural codon. All of these options have been explored in recent years to varying degrees of success.

The most efficient stop codon in *E. coli* is *ochre* (UAA), which is recognized by both release factors, and therefore is not a good candidate for suppression. There are, however, natural contexts in which *opal* (UGA) is suppressed in both prokaryotes and eukaryotes, such as with selenocysteine insertion [65]. In fact, the Schultz group has recently engineered an *opal* suppressor for use in *E. coli* [66]. Moreover, suppression of both *amber* and *ochre* stop codons has been demonstrated in mammalian cells [67]. Attempts have been made to use suppressors of rare codons in *E. coli* expression systems, and depletion of endogenous tRNAs from the cell-free extract has made this more tractable [68-72]. However, readthrough from tRNAs with near-anticodons and the need to heavily mutagenize genes to remove the selected sense codons makes this strategy problematic. Recently, Frankel & Roberts selected for tRNAs with NNC anticodons to suppress NNN codons in the rabbit reticulocyte expression system. They found that suppression of GUA was comparable to *amber* suppression [73]. A related approach may come from recent cell-free translation systems reconstituted entirely from recombinant proteins, since all of the tRNA and aaRS identities and concentrations can be altered at will to produce more "blanks" in the genetic code [74].

Naturally occurring frameshift suppressors in yeast and *Salmonella* are known, typically functioning by decoding a 4-base codon using a tRNA with an extended anticodon loop (8 nt instead of 7 nt) [65]. Suppression of 4-base and 5-base codons has been demonstrated in *E. coli* and in *E. coli* cell-free translation systems. Magliery *et al.* selected for tRNA^{Ser} with randomized, extended anticodon loops (8 or 9 nt instead of 7 nt) that would suppress randomized four-base codons in the gene for β -lactamase, and identified very efficient anti-

codon loop/4-base codon pairs for the codons AGGA, UAGA, CCCU and CUAG [75]. A similar approach afforded fairly efficient 5-base codon suppressors of CUAGU, CUACU, AGGAU [76]. These studies established a number of factors leading to efficient suppression of codons: Watson-Crick pairing at all positions; $N+4$ nt in the anticodon loop of a tRNA suppressing a codon with N nt, where 2 nt are on either side of the anticodon; and the presence of a U 5' to the anticodon and an A 3' to it. In general, suppressible codons are based on rare 3-base codons. Four-base codons have been used both for site-directed mutagenesis and to insert unnatural amino acids into proteins [77-79]. Sisido's group has incorporated multiple unnatural amino acids into proteins site-specifically using AGGU or GGGU and CGGG codons [80, 81]. Five-base codons of the type CGGNN were also investigated, and CGGAC was found to be the most efficiently suppressed [82]. Extended codons have a distinct advantage of the *amber* stop codon in that "readthrough" by natural tRNAs results in an out-of-frame product that typically terminates rapidly. Thus, material purified with a C-terminal tag is virtually guaranteed to contain unnatural amino acid at the selected site (except for the possibility of spontaneous frameshifts).

Unnatural base pairs (Fig. 4) are potentially the most general route to developing unique insertion signals. (See the

references cited here for reviews [83, 84].) However, one must incorporate the unnatural bases into both the tRNA and mRNA, and eventually one would like to do this in living cells. This means that the unnatural base pair must be a substrate for both DNA and RNA polymerases, EF-Tu and the ribosome. Each unnatural base must uniquely elicit its cognate (and not pair with any of the four natural bases), it must be reasonably stable, and it cannot result in chain termination after incorporation. This has turned out to be a tall order. The approach was first demonstrated by Benner and colleagues, who developed an iso-C/iso-G pair, which worked fairly well, except that d-iso-G elicited dT in addition to d-iso-C [85]. Benner, Chamberlin and coworkers demonstrated that synthetic RNA containing an (iso-C)AG codon could be efficiently suppressed by a tRNA with a CU(iso-G) anticodon [86]. The resulting peptide was only 17 residues long, however, due to the need to synthesize the RNAs chemically. Other bases with alternative hydrogen-bonding patterns were developed, but none were strictly orthogonal to the natural bases, especially in transcription [87-89].

Yokoyama and colleagues have synthesized a series of hydrogen-bonding shape mimics of natural nucleobases, and some, particularly 2-amino-6-(2-thienyl)purine (ds) and pyridine-2-one (dy), have good stability and elicit each other with fair fidelity in DNA replication [90-93]. Interestingly,

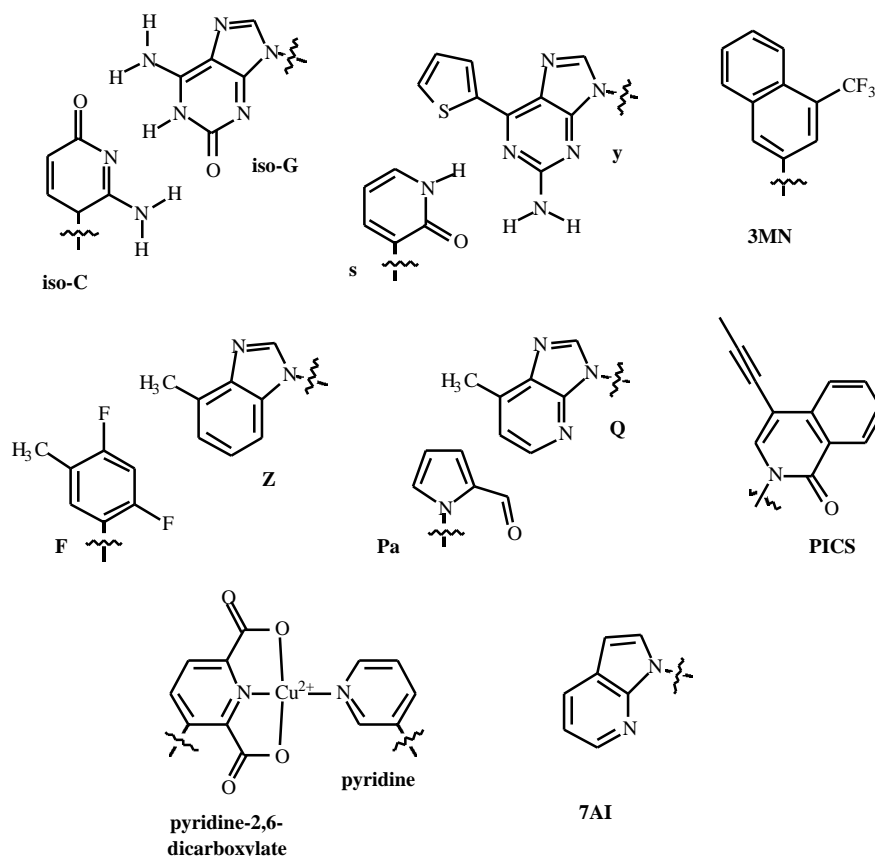


Fig. (4). Unnatural bases used for genetic code expansion. The bases 3MN, PICS and 7AI are drawn without a partner because they self-pair.

this pair was used for *in vitro* site-specific unnatural amino acid incorporation [94]. DNA was synthesized to contain a CTs sequence that could be transcribed into a yUG codon by T7 RNA polymerase. A yeast tyrosyl tRNA with a CUs anticodon was prepared with a combination of synthesis and ligation, and this was aminoacylated with 3-chlorotyrosine using yeast tyrosyl synthetase (which accepts C₁Tyr as a substrate). The C₁Tyr-tRNA_{CUs} was added to an *E. coli* cell-free transcription/translation system along with the DNA template and T7 RNA polymerase, and full-length protein was produced in good yield (40% compared to a natural codon) with high fidelity (>90% of the amino acid at the site was C₁Tyr).

Kool and coworkers suggested the idea of circumventing problems with using hydrogen-bonding schemes for base recognition by instead employing hydrophobic base pairs. For example, difluorotoluene (dF) was found to base-pair with adenine [95]. The base could also elicit dA in DNA replication, or 4-methylbenzimidazole (dZ) to form a stable, all-hydrophobic base pair [96]. However, both dF and dZ also elicit natural bases, and so are not suitably orthogonal. Improvements by the Kool and Yokoyama groups have led to the development of pyrrole-2-carbaldehyde (dPa) and 9-methylimidazo[4,5-b]pyridine (dQ), which forms a stable base pair in DNA that is effectively replicated and extended, although dPa is still compromised by spurious insertion of dA [97]. These bases are essentially shape-complementary to the natural purine:pyrimidine pairs, and they all possessed minor-groove hydrogen-bonding functionalities.

Romesberg, Schultz and coworkers have explored a series of hydrophobic bases that are, in general, neither shape-complementary nor in possession of any hydrogen-bonding functionality. It has been possible to develop stable base pairs of this type, sometimes incorporated into DNA with high fidelity, but chain termination after insertion has been a significant problem (e.g. d3MN, dPICS and d7AI) [98, 99]. Interestingly, the most efficient pairs have been "self-pairs," which is to say a base that elicits itself in DNA synthesis, which is perfectly reasonable for genetic code expansion. An interesting approach to solve the chain-termination problem is to explore the use of other natural polymerases, or to engineer new polymerases [100, 101]. For example, the d7AI self-pair is efficiently incorporated by Klenow fragment of DNA polymerase I, and it is efficiently extended by mammalian polymerase . Finally, a number of nucleobases that pair due to metal ion ligation have also been explored [102-105].

Despite the obvious advantages of the unnatural base approach, it is still fairly far from being broadly useful, even for *in vitro* applications. At minimum, an unnatural base synthetically incorporated into a DNA template needs to specify the efficient *in vitro* transcription of unnatural base-containing mRNA and tRNA. (This is similar to what Hirao *et al.* have accomplished, although the tRNA preparation was somewhat complicated [94].) It would be considerably more convenient if the unnatural base could be inserted into DNA with PCR and/or plasmid replication, which requires applicable thermophilic and cellular polymerases. In the case of cellular replication, it also requires that the unnatural nucleosides or nucleotides get into cells and be (or remain)

appropriately phosphorylated. Currently, *amber* suppression and four-base codon suppression are the best approaches in *E. coli*, and other stop codons or sense codons may be equally useful in eukaryotic systems.

IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS

Microinjection/Microelectroporation

So far, I have discussed the use of chemically-acylated tRNA in cell-free transcription/translation reactions. This is a straightforward way of generating large, soluble proteins with a site-specific unnatural amino acid mutation, and proteins can be produced in reasonable quantities with effort. However, to harness the power of unnatural protein engineering to investigate cellular function, the protein must be produced in living cells. Moreover, membrane-bound proteins are not amenable to the cell-free approach, and the Dougherty group has elegantly investigated ion channel function by adapting the chemical acylation approach for use *in vivo*. Most of the work so far has been carried out on the nicotinic acetylcholine receptor (nAChR) using *Xenopus* oocytes, which are sufficiently large that one can microinject reagents into them [106, 107]. Here, one must inject both mRNA and chemically-acylated tRNA, and that tRNA must be orthogonal to *Xenopus* aaRSs but competent to act in the translational apparatus. A modified version of *Tetrahymena thermophila* tRNA^{Gln}(CUA) called THG73, 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 [108].

An exciting development in this field has allowed extension of this scheme to mammalian cells [109]. Both CHO-K1 cells and rat hippocampal neurons were subject to delivery of DNA, mRNA and tRNA using microelectroporation. By administering chemically acylated *amber* suppressor (the THG73 tRNA) and nAChR mRNA to neurons in this fashion, functional receptors site-specifically modified with unnatural amino acids were produced and probed electrophysiologically. This method has the substantial advantage over other *in vivo* methods that virtually any amino acid can be inserted, but only very small quantities of protein can be produced, requiring very sensitive assays (like voltage clamping). RajBhandary has also shown that acylated tRNA can be introduced into mammalian cells (COS-1 cells) using transfection reagents [67].

Widespread/Residue-Specific Incorporation

In the context of protein science, the term *amino acids* typically means *proteinogenic amino acids*—the 20 common monomers that are incorporated into proteins. Recently, other (rarely) translationally incorporated amino acids like selenocysteine and pyrrolysine have been discovered [1-3]. However, there are many other amino acids that act in the biological milieu, including natural non-proteinogenic amino acids, amino acids created by posttranslational modification, and amino acids (natural and unnatural) that act on cells in one way or another. Citrulline and ornithine are intermediates in the biosynthesis of arginine, and are not incorporated into proteins, despite their relative similarity to arginine and

lysine, respectively. Posttranslational modification results in phosphorylated and sulfated amino acids, lipidated and glycosylated amino acids, methylated and acetylated amino acids, hydroxylated amino acids, and biotinylated amino acids (biocytin). On the other hand, L-canavanine, an oxy-analog of arginine, is produced by legumes and is toxic to bacteria and animals due to misincorporation into proteins (among other effects).

This raises an interesting question: how does the cell maintain the fidelity of the genetic code with so many amino acids around? The simple answer is that the aminoacyl-tRNA synthetases are exquisitely good at their job, and manage to correctly distinguish their amino acid substrate (there is generally one aaRS per amino acid per organism) from all the other amino acids in the cytosol (or organelle). The error rate in protein synthesis is only about 1 in every 10,000 residues, and part of this can be blamed on incorrect tRNA selection at the ribosome and errors in transcription. Many aaRSs employ specific mechanisms to “edit” (hydrolyze) improperly-acylated tRNAs, in addition to their exquisite recognition of amino acid and tRNA in the first instance.

But the more complicated answer is that the fidelity of all aaRSs is limited, especially when confronted with near-analogs of natural amino acids which have never been encountered in a biological context. The fact is that aaRSs only need to have evolved ways of excluding other cellular amino acids (proteinogenic or not), and this opens the door to a whole array of residue-selective (or, better, residue-specific) protein modifications achieved by adding unnatural amino acids to the growth medium. This fact has revolutionized protein biophysics: high-level incorporation of selenomethionine affords a general solution to the phase problem in X-ray crystallography through multiwavelength anomalous diffraction (MAD) [110, 111]. But it is also a route to both specific unnatural modifications of proteins and the generation of unnatural biomaterials.

For widespread replacement of an amino acid to be useful, it has to occur as specifically as possible, which is to say it should replace only one amino acid at any level, and that level should be 100%. Needless to say, it is not generally possible to grow cells under normal conditions and achieve 100% replacement of one of the natural amino acids; some critical, lethal replacements are bound to occur. In general, one gets the best incorporation of an amino acid analog with (1) an auxotrophic strain for the related natural amino acid, with sufficient growth to remove all the natural amino acid from the growth medium followed by (2) high-level induction of protein expression with concomitant addition of the unnatural amino acid [112]. Nearly quantitative replacement is possible under these conditions. The second issue with this method is that any given protein is likely to contain at least one of each of the natural amino acids, and so one must mutate the protein appropriately to achieve labeling at the desired position(s). This may not always be practical. Finally, the array of appropriate amino acid analogs essentially has to be determined empirically, although some engineering has been helpful, and thus the selection of analogs does not compare to what is possible through synthetic methods.

Nevertheless, some interesting and useful amino acids have been incorporated this way (Fig. 5). The 4-, 5-, and 6-

fluoro derivatives of tryptophan and *o*-, *m*- and *p*-fluoro derivatives of phenylalanine permit site-specific labeling for ¹⁹F NMR [113-115]. Besides SeMet, other Met analogs have been incorporated, such as 2-aminohexanoic acid (nor-leucine), ethionine, telluromethionine and *S*-nitrosohomocysteine [112, 116]. Thiaproline is inserted by ProRS [117]. These sorts of amino acids permit the characterization of proteins with what Budisa and coworkers call “atomic mutations,” changes of as little as a single atom in a whole macromolecule. Tirrell, Fournier and colleagues became interested in the use of unnatural amino acids to produce homogeneous, biopolymeric materials (periodic proteins) [118]. For example, 3-thienylalanine, azidohomoalanine, trifluoroisoleucine, hexafluoroisoleucine, homoallylglycine, and homopropargylglycine have been inserted into such proteins [119-123].

Some unnatural amino acids are only incorporated efficiently if MetRS is overexpressed, such as the olefinic and acetylenic amino acids *cis*- and *trans*-2-amino-4-hexenoic acid, 2-butynylglycine and allylglycine [124, 125]. Engineering of a PheRS (by Ibbá & Hennecke [126]) resulted in a larger binding pocket, which has allowed the insertion of *para*-substituted fluoro-, chloro-, bromo-, iodo-, azido-, cyano- and ethynyl-phenylalanine [127]. A related approach is mutation of an aaRS to abrogate hydrolytic editing of misacylated products, which has found purchase with ValRS and LeuRS since these enzymes require editing to remove other hydrophobic amino acids that are erroneously esterified. For example, low-level incorporation of aminobutyrate at Val codons was achieved by ablation of ValRS editing activity [128]. Tang & Tirrell have used LeuRS with reduced editing activity to introduce six new hydrophobic amino acids, both saturated and unsaturated [129].

Widespread methods of amino acid replacement have several positive points worth highlighting. Once a suitable amino acid is found, the method is technically simple, and large amounts of protein can be produced cheaply. So far, the engineering approaches applied to expand the scope of widespread amino acid replacement have been relatively simple and have yielded excellent results. However, the method is considerably less general with respect to amino acid than synthetic approaches. While it is an *in vivo* method, it requires growth conditions that make it much less useful for examining cellular processes, and it results in (often lethal) incorporation of unnatural amino acids all over the proteome under normal growth conditions. Indeed, the most serious problem is that it is not actually a method of expanding the genetic code, since it requires the sacrifice of one of the natural amino acids. At best, this is likely to be inconvenient in any specific protein. However, when this is not a concern, as with the homogeneous preparation of unnatural biomaterials, this method is particularly suitable.

Site-Specific Incorporation

The drawbacks and advantages of the *in vivo* chemical acylation and widespread replacement methods provide a blueprint for what one ultimately wants in a method of unnatural protein engineering. Ideally, the method will be suitable for use in living cells, so that it is technically simple and inexpensive to produce large amounts of protein or examine cellular processes (i.e. “unnatural cell biology”). It should

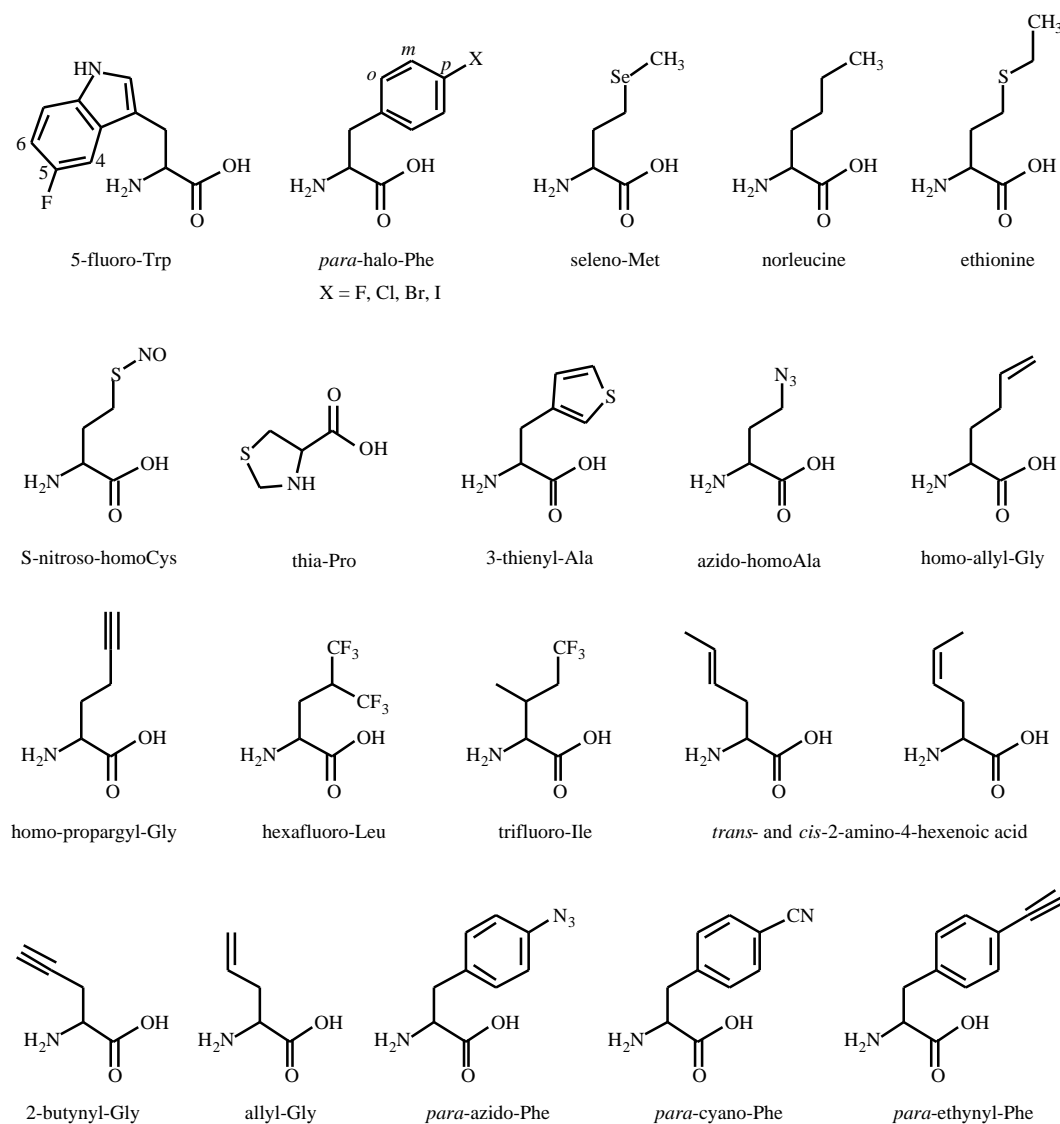


Fig. (5). Examples of amino acids incorporated by residue-specific replacement.

require little or no organic synthesis (except perhaps of the free amino acid), and should be as simple as transforming plasmids and adding amino acid to the growth medium. On the other hand, it should be broadly applicable to as many amino acids as possible. It should expand, not just recode, the genetic code, and it should do so in a site-specific manner, ideally so specifically that only the protein of interest will contain the unnatural amino acid, and only at the selected site.

How this can be accomplished returns us to the corollary to Crick's adapter hypothesis: essentially, we can invade the genetic code with a unique tRNA (and anticodon) and aminoacyl-tRNA synthetase. Specifically, we need the following:

1. A host organism. Due to the extensive engineering required, well-understood systems like *E. coli* and yeast have proven good starting points.
2. An insertion signal (codon). In *E. coli*, the *amber* stop codon is a good first choice. Expansion into four-base and unnatural codons is necessary to think about adding multiple amino acids to the genetic code of a single organism.
3. An "orthogonal" tRNA bearing the insertion signal, that is neither acylated nor deacylated ("edited") by any endogenous aaRS.
4. An orthogonal aaRS, that acylates only the orthogonal tRNA, only with an unnatural amino acid. For simplicity, an aaRS without hydrolytic editing is a good choice.
5. An unnatural amino acid that is not toxic nor incorporated into proteins by endogenous aaRSs, but that is readily uptaken by the cell (or produced in it).

Generating an Orthogonal tRNA/Synthetase Pair

The initial work by the Schultz group to this end focused on evolving such an “unnatural organism” in two stages: first, generate an orthogonal tRNA/aaRS pair, and, second, change the specificity of the aaRS with respect to amino acid. All initial work was carried out in *E. coli* using *amber*-suppressing tRNA. We first employed an engineering approach to generate an orthogonal tRNA/synthetase pair based on the exceedingly well-characterized *E. coli* glutamyl-tRNA synthetase (GlnRS)/tRNA^{Gln}(CUA) pair. (GlnRS naturally acylates an efficient *amber* suppressor in *E. coli*.) Based on the co-crystal structure [130], three mutations in tRNA^{Gln} were identified that together abrogated activation by GlnRS [131]. This tRNA was not appreciably acylated by any aaRS in an *E. coli* transcription/translation reaction, but it was competent for amino acid delivery when chemically acylated. A mutant GlnRS was then engineered using a DNA shuffling [132, 133] strategy coupled with a selection for *amber* suppression (survival in galactose in a strain with an *amber* mutation in the gene for β -galactosidase) [134]. Despite a remarkable change in specificity toward the orthogonal tRNA, the mutant GlnRS was still capable of efficiently acylating the wild-type tRNA^{Gln}, as well. If the amino acid specificity of this enzyme had been subsequently altered, the unnatural amino acid would have been inserted at Gln codons throughout the proteome in addition to *amber* codons.

Work by the Schimmel group suggested that another approach might be easier to obtain an orthogonal aaRS/tRNA pair: importation of a heterologous pair from another organism [135, 136]. Based on the fact that *Saccharomyces cerevisiae* tRNA^{Gln} (SctRNA^{Gln}) was not acylated by *E. coli* GlnRS (*EcGlnRS*), Liu & Schultz demonstrated that a modified SctRNA^{Gln}(CUA) and *ScGlnRS* constitute a functional, orthogonal pair in *E. coli* [137]. (Hereafter, the notation *OotRNA*^{Aaa}(NNN) and *OoAaaRS* shall be used analogously to the examples in the previous sentence, where *Oo* denotes the organism, Aaa denotes the amino acid, and NNN denotes the anticodon.) However, despite nearly 10 man-years of work engineering this pair, we were never able to alter the amino acid specificity of *ScGlnRS*, even though virtually all of the selection and library-construction technology was worked out on this system (D.R. Liu, T.J. Magliery, M. Pastrnak, S.W. Santoro & P.G. Schultz, unpublished, in addition to the cited references) [137-139].

Based on a similar observation from the Schimmel group [140, 141], Wang *et al.* showed that the *Methanococcus jannaschii* tRNA^{Tyr}(CUA) and *MjTyrRS* constituted an orthogonal pair in *E. coli*, as well [142]. However, while the *MjTyrRS* is considerably more active than *ScGlnRS*, the *MjTyrRS* was not “as orthogonal” as SctRNA^{Gln}(CUA), as measured by the survival of *E. coli* expressing these tRNAs in a strain bearing an *amber* mutant of β -lactamase on varying concentration of ampicillin. (That is: *E. coli* with *MjTyrRS*(CUA) survive at higher ampicillin concentrations due to greater *amber* suppression from spurious acylation by *E. coli* endogenous synthetases.) Wang & Schultz used a semi-rational design strategy to increase the orthogonality of this tRNA [143], employing a variation on the double-sieve selection introduced by Liu & Schultz (see

below) [137]. First, a library of *MjTyrRS*(CUA) variants was co-produced in a strain with a multi-*amber* mutant of barnase, a toxic RNase in *E. coli*. Surviving cells bore tRNAs less capable of being acylated by endogenous synthetases. The selected tRNAs were then co-produced in a strain with an *amber* mutant of β -lactamase and expressing *MjTyrRS*. Survival at this step ensures that mutant tRNAs are still competent for protein biosynthesis. The resulting orthogonal tRNA and has been by far the most useful so far, due to the high activity of the synthetase and good orthogonality of the tRNA, in addition to the relatively large, hydrophobic Tyr binding pocket in *TyrRS*. Recently, solution of the *MjTyrRS* crystal structure permitted design of a mutant synthetase with better recognition of the *MjTyrRS*(CUA), which may be useful in further work altering the amino acid specificity of the synthetase [144].

A number of other pairs have also been introduced. Pastrnak *et al.* engineered an orthogonal pair based on the *amber*-suppressing SctRNA^{Asp}(CUA) and a Asp \rightarrow Lys mutant of *ScAspRS* known to acylate that tRNA [138]. (Actually, the homologous mutation was known in the *E. coli* enzyme and tRNA. The anticodon is a key recognition element of AspRS, making the *amber* suppressor instantly orthogonal.) However, the amount of *amber* suppression was weak, although more reasonable in an RF1-deficient *E. coli* strain, since RF1 competes with *amber* suppressors for recognition of UAG codons. The RajBhandary group also selected for a mutant *ScTyrRS* that was orthogonal in *E. coli* but capable of acylating the orthogonal *amber*-suppressing mutant of *E. coli* initiator tRNA^{Met} [145]. Anderson & Schultz have shown that *Methanobacterium thermoautotrophicum* LeuRS and derivatives of *Halobacterium* NRC-1 tRNA^{Leu} constitute orthogonal pairs capable of suppression of *amber*, *opal* (UGA) and 4-base (AGGA) codons in *E. coli* [66]. Santoro *et al.* also showed that *Pyrococcus horikoshii* GluRS and a consensus-derived archaeal tRNA^{Glu}(CUA) are an orthogonal pair in *E. coli* [146]. However, there has been no success altering the amino acid specificity of any of the synthetases described in this paragraph. Recently, the Schultz group has selected a mutant *Pyrococcus horikoshii* LysRS derivative that is capable inserting homoglutamine in *E. coli* in response to the four-base codon AGGA using a consensus-derived archaeal tRNA^{Lys}(UCCU) [147]. This is only the second orthogonal pair shown to be useful in *E. coli*, and the fact that it allows four-base suppression means that it is now possible to simultaneously, site-specifically introduce two unnatural amino acids into bacterial proteins.

The first orthogonal pair for use in eukaryotes was characterized by the RajBhandary group, consisting of *EcGlnRS* and an *amber*-suppressing derivative of *EctRNA*^{Met} used in yeast [145]. RajBhandary and coworkers also demonstrated that coexpression of *EcGlnRS* and *EctRNA*^{Gln}(CUA) is necessary and sufficient for *amber* suppression in mammalian COS-1 and CV-1 cells, although the orthogonality of the synthetase was not assessed (beyond the fact that it is not lethal) [148]. However, the amino acid specificity of *EcGlnRS* has not subsequently been altered. The Yokoyama group found that a mutant *E. coli* tyrosyl-tRNA synthetase that inserts 3-iodo-Tyr and *EctRNA*^{Tyr}(CUA) constitutes an orthogonal pair in wheat germ extract [149]. When expression of this pair was attempted in mammalian cells (using a

D-arm mutant of the tRNA to create an internal promoter), no *amber* suppression was detected [150]. It was found that the same synthetase was capable of acylating a *Bacillus stearothermophilus* tRNA^{Tyr}(CUA), allowing *amber* suppression in mammalian CHO cells. Again inspired by work from the Schimmel group [151-153], Chin *et al.* demonstrated that *E. coli* tyrosyl-tRNA synthetase and *EctRNA*^{Tyr}(CUA) constitute an orthogonal pair in yeast, and they have successfully engineered mutants of *EcTyrRS* that acylate with unnatural amino [154, 155]. Recently, Zhang *et al.* have introduced a mutant *Bacillus subtilis* TrpRS that inserts 5-hydroxy-Trp, and an opal-suppressing mutant of *BstRNA*^{Trp}(UCA), for selective incorporation in mammalian cells [156].

Altering Amino Acid Specificity

Alteration of the amino acid specificity of the aaRSs has been the most challenging aspect of the development of this methodology. AaRSs are exceedingly good at what they do, and it has proven no simple matter to coax them to do otherwise. The problem is complicated considerably by at least five facts. (1) The small amino acid substrates are contacted by a large number of residues concentrated in three-dimensional space but often spread over the primary structure of the synthetase. This makes it technically difficult to construct reasonable mutagenic libraries. (2) In spite of this, mutations far from the active site of aaRSs are known to affect aminoacylation kinetics. (3) Mutations to the synthetase can also affect recognition of tRNA, thereby altering the orthogonality of the enzyme. (4) It is extremely difficult to alter the specificity of an aaRS without reducing its aminoacylation activity, but a certain level of activity is required to support high-level protein synthesis. Thus, engineering of synthetases with weak aminoacylation activity toward native substrates has proven exceedingly difficult. (5) Selecting for unnatural amino acid specificity requires that the survival of the organism be tied to the insertion of an unnatural amino acid, and no direct way to do this has been developed.

A small number of attempts have been made to alter the specificity of aaRSs by inspection or semi-empirically, with varying results. From genetic screens, it was known that the *EcPheRS* Ala294→Ser mutant resisted incorporation of *p*-F-Phe. Ibba & Hennecke showed that the Ala294→Gly mutant of this enzyme, with an expanded binding pocket, is capable of inserting *p*-Cl-Phe and *p*-Br-Phe, and the Tirrell group has shown that *para*-iodo-, azido-, cyano- and ethynyl-Phe are also accepted [126, 127]. Furter used the associated *p*-F-Phe resistant *E. coli* strain to engineer the first bacterium able to site-selectively insert an unnatural amino acid [157]. Yeast PheRS (which accepts *p*-F-Phe) and tRNA^{Phe}(CUA) were expressed in this strain. When *p*-F-Phe was added to the medium, about 75% of the *amber*-encoded sites in the target protein (DHFR) were found to contain *p*-F-Phe (the rest was Phe and Lys), and about 7% of Phe sites contained *p*-F-Phe. This indicates both that *EcPheRS*(A294S) accepts *p*-F-Phe weakly, and that the *SctRNA*^{Phe}(CUA) is also acylated with Phe and Lys (presumably by *EcLysRS*, in the latter case). While not ideal, this system is probably suitable to obtain protein site-specifically fluorinated for NMR. Kiga *et al.* converted *EcTyrRS* into an enzyme capable of accepting 3-iodotyrosine as a substrate by selecting three sites for

mutagenesis based on structure and examining the aminoacylation properties of 50 mutants *in vitro* [149]. Interestingly, while >95% of *amber*-encoded sites were occupied by 3-I-Tyr in a wheat-germ extract translation reaction, Tyr-containing protein was produced if 3-I-Tyr was left out of the reaction mixture. This is potentially a problem when moving this technology into cells, since the concentration of unnatural amino acid cannot be controlled arbitrarily. A number of attempts have been made to change GlnRS into GluRS, but the resulting enzymes still prefer Gln (although by a considerably smaller amount than wild-type GlnRS) [158, 159]. Additionally, some attempts to computationally predict mutations to alter amino acid specificity have been reported [160, 161].

Liu & Schultz introduced the concept of a double-sieve selection to isolate aaRS mutants capable of uniquely inserting unnatural amino acids [137]. First, selection of a library of aaRS variants is carried out in the presence of the tRNA(CUA) and an *amber* mutant of β -lactamase, with ampicillin and unnatural amino acid supplementation of the medium. (Minimal medium aids the uptake of unnatural amino acid.) Use of a permissive site in β -lactamase ensures that survivors of the selection contain aaRSs that are capable of acylating the tRNA(CUA)—thus producing β -lactamase and conferring resistance to ampicillin—but which amino acid was esterified onto the tRNA is unknown. The synthetases from the survivors of this positive selection are then expressed in bacteria bearing the tRNA(CUA) and a multi-*amber* gene for barnase, which is toxic in *E. coli*. No unnatural amino acids are added to the medium, so survival at this stage ensures that no cellular amino acids (proteinogenic or otherwise) are a substrate for the mutant synthetases. Survivors of both selections are then known to be active toward *some* amino acid, but not *endogenous* amino acids; thus, they must be active toward the unnatural amino acid. The initial implementation of this strategy, with the *ScGlnRS/SctRNA*^{Gln}(CUA) orthogonal pair, a large library of mostly commercially-available amino acids, and synthetase libraries created by random mutation (DNA shuffling), resulted in no useful synthetases.

Two critical technical improvements to the methodology were (1) the replacement of the β -lactamase positive selection with chloramphenicol acetyltransferase-based selection, and (2) use of a directed, semi-rational library construction method [8, 75, 138, 162]. Due to the fact that ampicillin is bacteriocidal and periplasmically active, it was impossible to know the level of ampicillin that was appropriate for selection, and rescue of cells without β -lactamase activity was possible *in trans*. Chloramphenicol is bacteriostatic and acts cytosolically, allowing a broad range of chloramphenicol concentration to be suitable for selection. Random mutagenesis of wild-type GlnRS had two negative consequences. First, most of the library members were very active toward Gln, requiring exceptional performance from the negative selection. Second, the probability of mutation in the proximity of the amino acid substrate was low. Instead, guided by the ternary X-ray crystal structure of *EcGlnRS*, tRNA^{Gln} and a Gln-AMP analog [130], 5-10 residues proximal to the substrate were first mutated to Ala and then randomized to all 20 amino acids. The resulting libraries contained few synthetases with activity toward Gln, allowing one to forego the

negative selection in early rounds, and were certain to have an altered amino acid binding pocket.

Unfortunately, libraries designed for use with carboxamide N-alkylated Gln analogs, analogs elaborated from the α -carbon, and the β -hydroxy acid analog of Gln were not found to contain any synthetases capable of activating the unnatural amino acids at useful levels (T.J. Magliery, M. Pastrnak & P.G. Schultz, unpublished, in addition to [75]). However, using the very same library construction and selection methodology, the Schultz group has been able to alter the specificity of *Mj*TyrRS, allowing the delivery of several useful tyrosyl analogs, such as *O*-methyltyrosine, *p*-benzoylphenylalanine, 3-(2-naphthyl)alanine, *m*-acetylphenylalanine, and *O*-allyltyrosine (Fig. 6) [162-166]. Importantly, the site-specificity associated with these systems is excellent. Typically, no protein can be detected in the absence of the synthetase, tRNA or unnatural amino acid, and mass spec-

trometry indicates modification at only the single amber-encoded position. It is very likely that the high activity of the *Mj*TyrRS and the excellent orthogonality of the synthetase and tRNA account for why this enzyme has been so much more amenable to substrate specificity changes than *Sc*GlnRS, *Sc*AspRS, *Mt*LeuRS or *Ph*GluRS. It is also probably no coincidence that the synthetase with the most hydrophobic binding pocket was the easiest to alter. Incidentally, all of the libraries of *Mj*TyrRS were designed using the *Bacillus stearothermophilus* structure, but the structure of the *Methanococcus jannaschii* protein has recently been solved [144]. There are a number of differences in the tyrosine binding pocket that will improve further library design with the enzyme, and may even suggest ways to improve some of the existing mutant synthetases, such as that for *O*-methyltyrosine (Fig. 7).

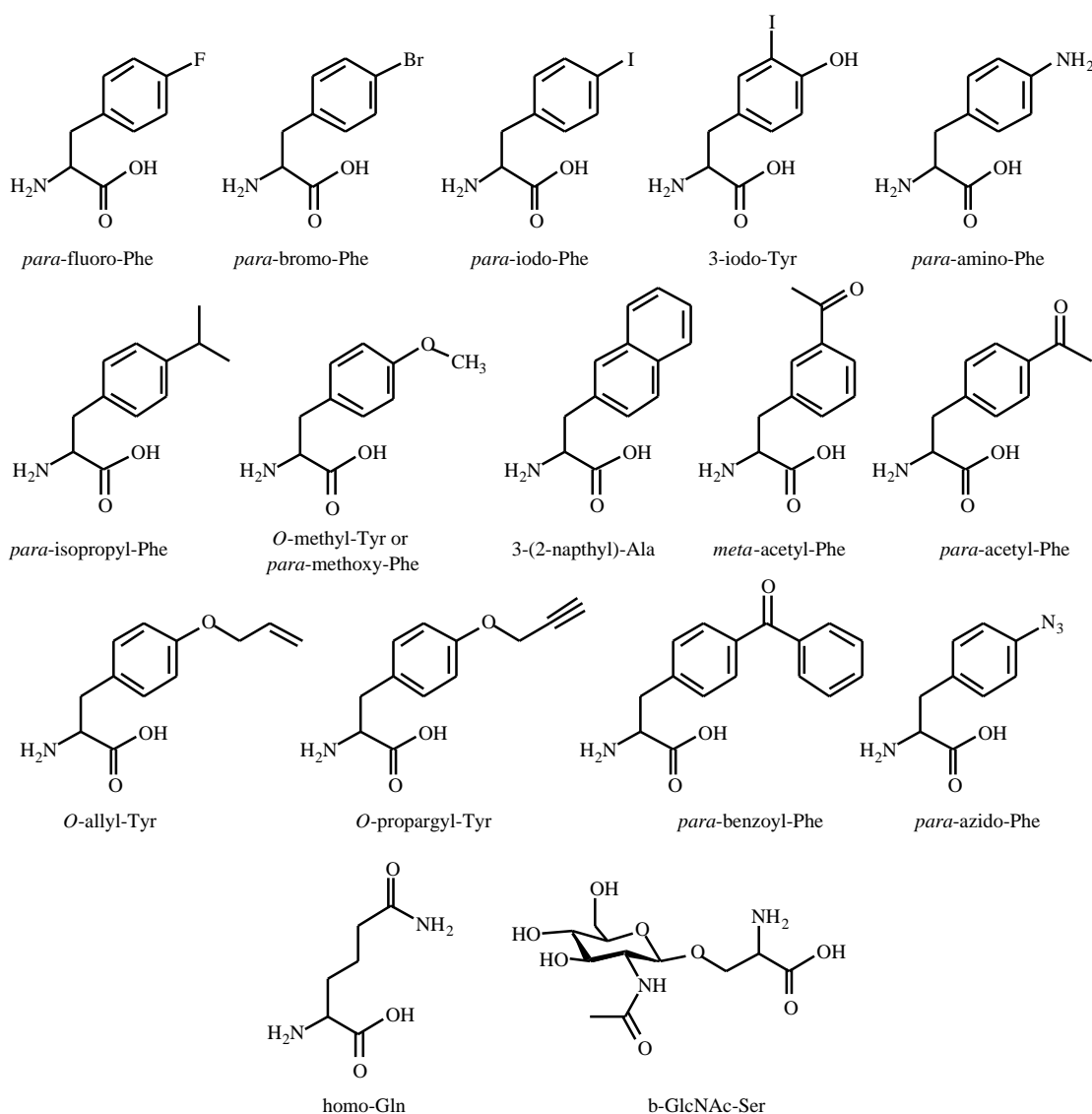


Fig. (6). Unnatural amino acids incorporated site-specifically in living cells with modified aminoacyl-tRNA synthetases. Except for homoGln, all of these are inserted by mutants of PheRS or TyrRS, which explains their structural similarity.

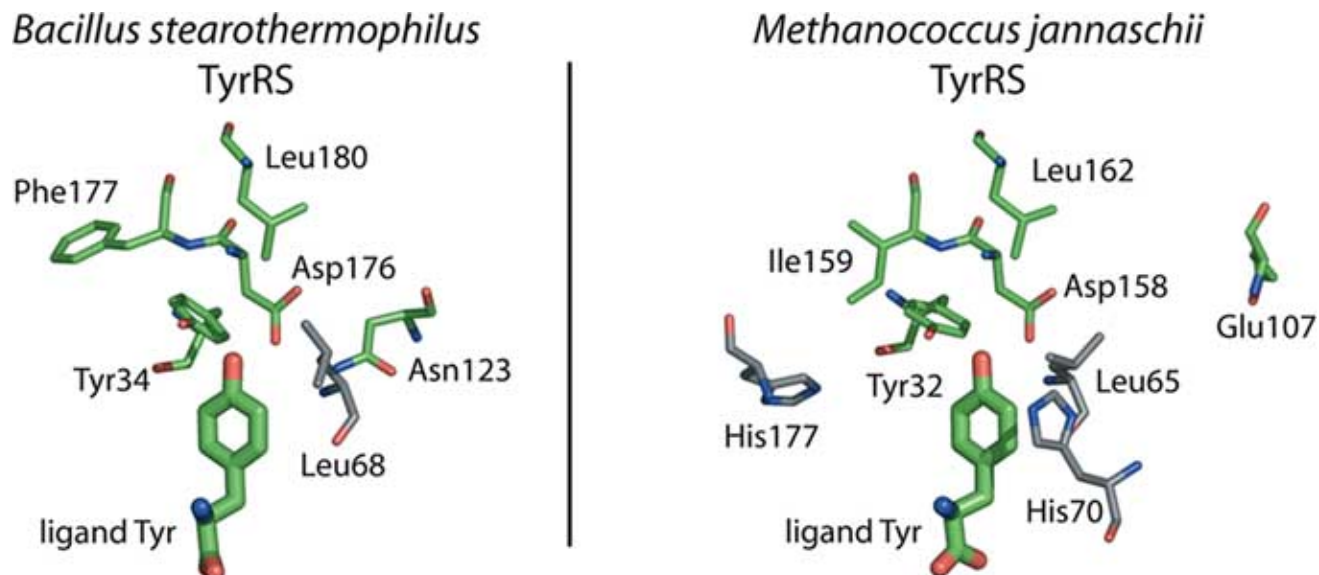


Fig. (7). The active site of TyrRS. Amino acids near the substrate tyrosine of *Bacillus stearothermophilus* (left) and *Methanococcus jannaschii* (right) TyrRS. Nearly all of the *Mj*TyrRS derivatives used for insertion of unnatural amino acids were generated by randomization of the Tyr32 (Tyr34), Glu107 (Asn123), Asp158 (Asp176), Ile159 (Phe177) and Leu162 (Leu180), where the corresponding amino acids in *Bs*TyrRS are noted parenthetically. These amino acids were selected from the *Bs*TyrRS structure, which was available at the time, but recent solution of the *Mj*TyrRS structure suggests some possible improvements. Notably, Glu107 is further from the ligand than expected, and nearby His70 and His177 have no analogs in *Bs*TyrRS. Another nearby amino acid, Leu65 (Leu68), has not been altered. Rendered using PyMOL (Warren Delano, 2002, <http://www.pymol.org>) from PDB entries 4TS1 and 1J1U.

Santoro *et al.* extended the selection technology by developing a fluorescence-based screen for *amber* suppression [167]. An *amber* mutant of T7 RNA polymerase was used to drive the transcription of green fluorescent protein (GFP). In combination with fluorescence-activated cell sorting (FACS), this system can be used as both a positive screen for *amber* suppression (where one sorts for fluorescent cells in the presence of unnatural amino acids) and a negative screen against *amber* suppression (sorting for dim cells in the absence of unnatural amino acids). A multivalent system, where chloramphenicol-based selection is possible in addition to screening, has been especially effective. Amino acids such as *p*-aminophenylalanine, *p*-isopropylphenylalanine, *O*-allyltyrosine, and *p*-azidophenylalanine have been added to the repertoire through this approach, again using the *Mj*TyrRS orthogonal pair [167, 168]. An interesting extension of this work was the engineering of a bacterium capable of biosynthesizing *p*-aminophenylalanine using the PapABC enzymes of *Streptomyces venezuelae* to convert chorismate [169]. With the mutant synthetase and altered *Mj*tRNA^{Tyr}(CUA), this bacterium has a *bona fide* 21-amino acid genetic code, as it is not dependent upon the addition of the unnatural amino acid to the medium. Such an organism does not require minimal medium for culturing, and so may be suitable for more ambitious organismal engineering projects (with the caveat that containment of the bacterium is exceedingly important).

Recently, Chin *et al.* reported an important type of double-sieve selection useful in yeast employing the *Ec*-TyrRS/*Ec*tRNA^{Tyr}(CUA) orthogonal pair [154]. An *amber* stop codon was inserted in the gene for the transcription factor GAL4, such that only the DNA-binding domain would

be produced without *amber* suppression, but full-length protein including the activation domain would be produced with *amber* suppression. For positive selection in the presence of unnatural amino acids, complementation of auxotrophy with the *HIS3* allele (under the control of the GAL4 promoter) was used, since the activity of the dehydratase that it encodes can be modulated dose-dependently with 3-aminotriazole. Negative selection (in the absence of unnatural amino acids) was accomplished by modifying the “reverse two-hybrid” system, wherein the *URA3* gene product converts 5-fluoroorotic acid to a toxic product. Amino acids including *para*-acetyl-, benzoyl-, azido-, and iodo-phenylalanine have been added to the yeast repertoire this way, as has *O*-methyl- and *O*-propargyl-tyrosine [155, 170].

Site-specific *in vivo* methods of inserting unnatural amino acids into proteins are an exciting development for the protein biochemistry and biophysics community. In either bacteria or yeast, specific unnatural modifications can be made to proteins by simply creating an *amber* mutant using oligonucleotide-directed mutagenesis of the gene of interest, co-transforming a plasmid bearing that gene with a plasmid for orthogonal tRNA and mutant synthetase production, and growing the cells in media supplemented with the unnatural amino acid. The modifications are highly specific, and the target protein can be expressed at low levels or overexpressed to obtain sufficient quantities for *in vitro* methods like NMR or X-ray crystallography. Moreover, there is already a useful array of unnatural functionalities available for both bacterial or yeast work, including affinity labels, unique reactive handles and heavy atoms.

On the other hand, the Achilles’ heel of this method is that for every amino acid that one is interested in, one must

engineer a new synthetase. The technology for this is robust, and the range of functionalities does not appear to be limited, but so far only large, hydrophobic analogs of tyrosine have been readily accessible. This situation can likely be improved by finding further synthetase/tRNA pairs (especially using non-*amber* insertion signals) that are highly active and orthogonal. However, it is likely to be inherently difficult to modify active sites that are designed to bind to polar amino acids, since hydrogen-bonding is much more difficult to achieve geometrically than hydrophobic packing. There will also be certain interesting amino acids that are not amenable to this approach, such as near-analogs of natural amino acids that are misactivated by endogenous synthetases, or amino acids that are not imported into the cell. Finally, the ability to do “unnatural cell biology” on mammalian cells is sure to be one of the most exciting and useful possibilities, but it will be increasingly difficult to develop selectable systems in higher eukaryotic cells. Fortunately, it may be possible to “transplant” tRNA/ synthetase pairs engineered in yeast directly into mammalian cells.

UTILITY OF UNNATURAL PROTEIN ENGINEERING

Throughout this review, I have highlighted many of the kinds of amino acids that have been successfully incorporated into proteins using both *in vitro* and *in vivo* methodologies. Excellent reviews have been written on the uses of unnatural amino acids, and I refer the reader to those for a more comprehensive discussion of this topic [7, 8, 10-12, 22, 171, 172]. Here, I will only provide a brief overview and mention some of the more recent applications of unnatural protein engineering.

Biophysical Probes

Isotopically labeled amino acids [173], spin-label amino acids and fluorophores [33, 42, 174, 175] have been incorporated into proteins. Aryl iodides and bromides may be useful as heavy-atoms for X-ray crystallography [127, 155]. Expressed protein ligation (EPL) has been especially useful for the segmental isotopic labeling of proteins for NMR, allowing solution study of very large proteins without the spectral overlap problems inherent in such work [176]. Incorporation of two fluorophores with overlapping emission and absorption spectra has allowed fluorescence resonance energy transfer (FRET) studies of protein dynamics. For example, dual-dye incorporation into a version of c-Crk-II permitted real-time monitoring of phosphorylation by c-Abl kinase [177]. Incorporation of biotin-containing amino acids, which are strongly bound by streptavidin, has been used both to map transmembrane topology [178] and to demonstrate the power of mRNA-protein fusion selections [73]. Incorporation of trimethylammoniumalkyl groups (a ‘tethered agonist’) has aided the understanding of acetylcholine recognition by the nicotinic receptor [179].

Fluorinated Amino Acids

Fluorinated amino acids are interesting both because they are a biophysical probe (^{19}F is NMR active with $I=1/2$ and at 100% abundance) and have profound effects on the solubility, and consequently stability, of proteins. Tirrell and co-

workers have shown that coiled-coils can achieve higher chemical and thermal stability upon replacement of hydrophobic amino acids with fluorinated amino acids, and some of these proteins retain activity (fluorous-GCN4 binds to DNA, for example) [122, 180, 181]. This is also interesting since Kumar and colleagues have shown that fluorinated interfaces can direct protein-protein interactions, particularly in membrane proteins [182-184]. Fluorophenylalanines have also been shown to affect the UV absorption properties of proteins, which may be a useful new biophysical probe [115]. Different stereochemical substitution of fluorine in proline affects the barrier to *cis-trans* isomerization as well as the position of the equilibrium [185]. Frieden has used fluorinated amino acids to investigate the kinetics of side-chain stabilization during protein folding by NMR [186].

Unnatural Fluorescent Proteins

Tryptophan analogs have been introduced to alter the intrinsic fluorescence of proteins. 4-Aminotryptophan, for example, results in pH-dependent fluorescence [187]. Replacement by that amino acid of the two Trp positions of the cyan variant of GFP resulted in an extreme red-shifted variant of GFP dubbed “gold” fluorescent protein [188]. Both to expand the spectral properties of GFP variants and to better understand structure-activity relationships, the tyrosine in the GFP chromophore was replaced with various unnatural tyrosyl analogs *in vivo* [189]. Two of these in particular, with *p*-amino-Phe and *p*-methoxy-Phe, may be useful as replacements for EGFP (which allows FRET with fluorescein) and BFP (which can be used in FRET with EGFP, but is limited by poor quantum yield and rapid photobleaching), respectively.

Unique Reactive Handles

Expanding the reactivity of proteins, particularly in a site-specific way, is potentially the most useful alteration that can be made to proteins. Chemical modification of natural proteins is basically limited to use of the nucleophilic groups: cysteine’s thiol, amines on lysine and at the N-terminus, histidine’s imidazole, and the alcohols, depending upon pH. However, it is often difficult to prevent cross-reaction with other nucleophilic groups in the protein, and extensive mutation (for example, to yield a single-Cys version of a protein) is often required. It is possible to introduce an electrophile into a protein by mild oxidation of N-terminal Ser or Thr, but more general approaches to unique reactive handles have many potential applications [190].

The introduction of ketones and aldehydes allows chemoselective formation of hydrazones and oximes using hydrazides and hydroxylamines, respectively (Fig. 8). This has been demonstrated with *in vitro* [191] and *in vivo* [166, 192] protein production, and it has been used extensively by the Bertozzi group for synthesizing glycopeptide mimetics and engineering the cell surface [193, 194]. It has been extended with EPL and *in vivo* methods to large glycoprotein mimetics using a keto functionality [195-198]. Recently, the Schultz group demonstrated that hydrazone formation is possible in living cells to label ketone-containing protein with fluorophore hydrazides [192]. Both cytosolic and outer-membrane proteins were labeled in this fashion. Ketones

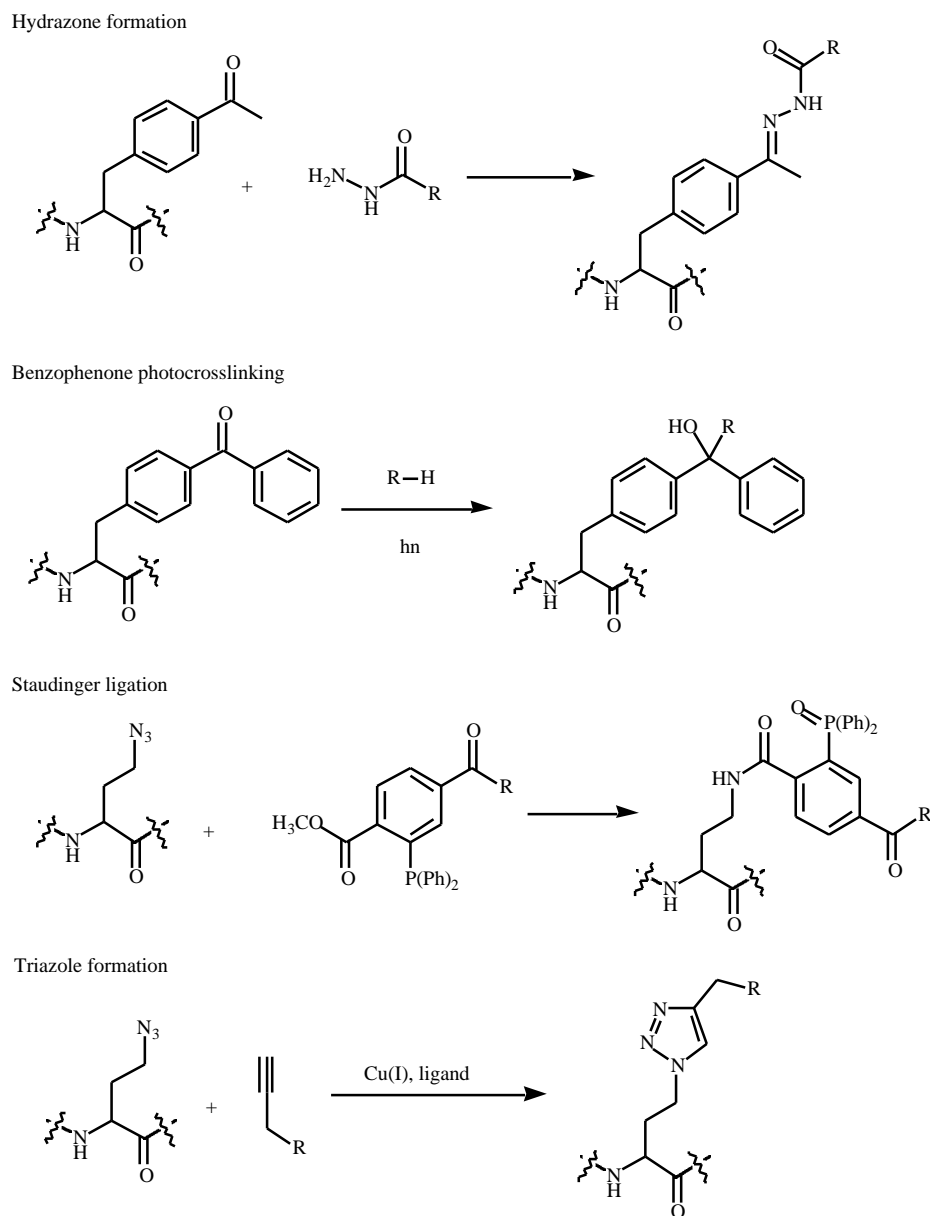


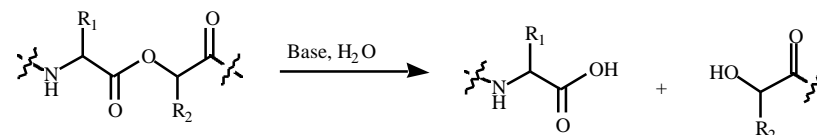
Fig. (8). Expanding the chemical reactivity of proteins. Four chemoselective reactions are shown that are possible with the incorporation of unnatural amino acids. These reactions are sufficiently orthogonal to cellular molecules that the first two have been demonstrated in living bacteria, and the second two have been used to modify the surfaces of living cells.

also allow for the possibility of protein stabilization by formation of an imine with lysine. Chin *et al.* have shown that incorporation of *p*-benzoylphenylalanine into proteins in bacteria allows photoaffinity labeling to interacting proteins upon irradiation [163, 199]. The covalently-linked proteins can then be isolated or visualized electrophoretically after *in vitro* or *in vivo* crosslinking. The Bertozzi group has made use of the Staudinger ligation to selectively modify sugar moieties on cell surfaces, and Kiick and colleagues have incorporated azidohomoalanine into proteins using *in vivo* amino acid replacement to allow selective modification of proteins with triarylphosphines [121, 200]. This same amino acid can be used for Cu(I)-catalyzed triazole formation with an alkyne, and cell-surface labeling was accomplished in this

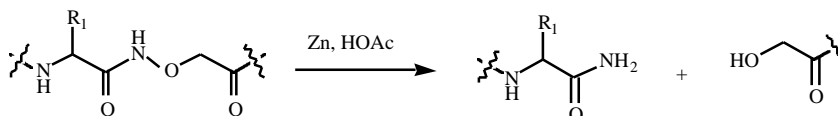
way [201]. Presumably, this reaction can also be used for alkynyl amino acids, much as olefin metathesis could be used for olefinic amino acids [202].

Several unnatural amino acids have been useful for engineering chemical scission sites into proteins (Fig. 9). Hydroxyacids inserted through chemical acylation of tRNA result in an ester linkage that can be cleaved in base [203]. Similarly, specific cleavage at aminooxyacetic acid can be achieved with zinc and acetic acid, and cleavage of the protein backbone at allylglycine is possible upon treatment with iodine [204, 205]. Photochemical proteolysis occurs upon irradiation at 2-nitrophenylglycines [206]. Photocaged groups, such as nitrobenzyl ethers and esters, shield reactive

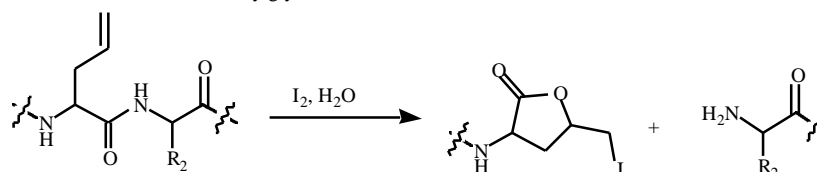
Ester hydrolysis at α -hydroxy acid



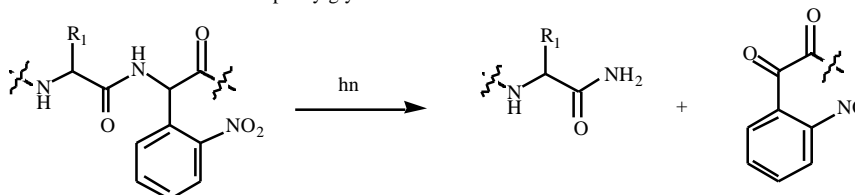
Acid hydrolysis at aminooxyacetic acid



Iodine-mediated scission at allylglycine



Photochemical scission at 2-nitrophenylglycine



Photodeprotection of "caged" nitrobenzyl ethers, esters

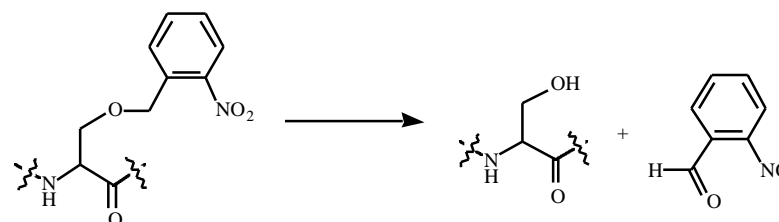


Fig. (9). Chemical scission and deprotection of proteins. Several methods have been introduced to specifically cleave the backbone of proteins at the site of unnatural amino acid (or analog) incorporation. Also, photodeprotection of "caged" Ser and Tyr has been demonstrated, including in *Xenopus* oocytes.

moieties until irradiative deprotection, and can be used to control enzymatic activity or protein-protein interaction in a time-resolved manner [207-209]. These types of functionalities are only available through synthetic methods at the moment but will prove especially useful for cell biological studies when they are available for translational incorporation.

Posttranslational Modifications

Understanding the roles of proteins in cellular function necessarily means understanding the extent and meaning of posttranslational modification, particularly in eukaryotes. However, it is difficult to obtain specifically, homogeneously modified protein, not least because many posttransla-

tional modifications do not occur in the bacteria in which proteins are often most conveniently expressed. EPL has been useful for the homogeneous preparation of phosphorylated proteins, and has been used to understand the role of phosphorylation in TGF β signaling [210-213]. The Cole group has incorporated phosphoserine, a non-hydrolyzable analog of phosphoserine, using EPL, which creates an "always on" protein state [214-216]. As mentioned above, the Bertozzi group has used a chemoselective mimetic approach to generate neoglycoproteins, employing hydrazone, oxime and Staudinger chemistry [197]. Zhang *et al.* have translationally incorporated β -GlcNAc-serine in bacteria, where proteins are not normally glycosylated, permitting homogeneous preparation of large amounts of glycoprotein [217]. Lipidation, such as introduction of a geranylgeranyl group

into the GTPase Rab7, has also been accomplished with EPL [218, 219].

TEMPLATED SYNTHESIS OF NON-PEPTIDIC MOLECULES

While extensive discussion of this topic is beyond the scope of this review, it is worth noting that, in the limit, chemists would like to be able to program and control the synthesis of nonpeptidic molecules as well. Just as physical association of nucleic acid permits facile production, selection and identification of proteins, it could ultimately do the same for peptidomimetics, like polyesters, or even less-related molecules. Much of the pioneering work in this field has been carried out by the Orgel group, largely studying the nonenzymatic synthesis of RNA from DNA and other templates [220, 221]. Recently, the Liu group has reported the DNA-templated polymerization of peptide nucleic acid aldehydes [222]. Liu and colleagues have also devised systems for the DNA-programmed multistep syntheses of small molecules [223], and have demonstrated that *in vitro* selections on DNA-linked small molecules are possible as well [224, 225]. Harbury and colleagues recently have introduced a technique called "DNA display" for the programmed synthesis of peptides and other polymers using DNA-directed spatial separation [226-228].

CONCLUSION

Less than a decade ago, unnatural protein engineering was mostly an idea, accessible to only a few labs of specialists, and decidedly in the methodology-development stage of its existence. Today, unnatural protein engineering is a reality accessible to any biochemistry or biophysics lab at reasonable cost and without heroic synthetic efforts. In particular, expressed protein ligation, widespread amino acid replacement, and site-specific translational incorporation in living cells allow the protein scientist to endow a target protein with unique reactive handles, biophysical probes, and posttranslational modifications in a controlled, homogenous manner. These possibilities herald the beginning of a new phase in the study and engineering of proteins, a phase that will vastly aid our understanding of cellular function, and a phase in which molecules with entirely novel properties can be created. There is, of course, still need for significant improvements to the existing technologies, including improving the scope of peptide synthesis and working around current limitations in ligation chemistry; dramatically improving the ease of chemical synthesis of acylated tRNA, perhaps using ribozymes; identifying or engineering many more orthogonal tRNA/synthetase pairs and radically expanding the repertoire of unnatural amino acids that can be inserted in living cells; perfecting alternative coding schemes, including unnatural codons; developing systems for facile use in mammalian cells or whole organisms; and working towards the templated creation of molecules that are far different from proteins.

ACKNOWLEDGEMENTS

T.J.M. is a fellow of the National Institutes of Health (GM065750). The author thanks Lynne Regan (Yale University, New Haven, CT) for additional support during the

preparation of this manuscript. Thanks to P.G. Schultz (The Scripps Research Institute, La Jolla, CA) for allowing me to cite unpublished results, and to S.W. Santoro (Harvard University, Cambridge, MA) for critical reading of this manuscript.

REFERENCES

- [1] Bock, A.; Forchhammer, K.; Heider, J.; Baron, C. *Trends Biochem. Sci.* **1991**, *16*, 463.
- [2] Hao, B.; Gong, W.; Ferguson, T.K.; James, C.M.; Krzycki, J.A.; Chan, M.K. *Science* **2002**, *296*, 1462.
- [3] Srinivasan, G.; James, C.M.; Krzycki, J.A. *Science* **2002**, *296*, 1459.
- [4] Braman, J. *In Vitro Mutagenesis Protocols*, 2nd ed.; Humana Press: Totowa, NJ, **2002**.
- [5] Zoller, M.J.; Smith, M. *Methods Enzymol.* **1983**, *100*, 468.
- [6] Anthony-Cahill, S.J.; Magliery, T.J. *Curr. Pharm. Biotechnol.* **2002**, *3*, 299.
- [7] Magliery, T.J.; Pastrnak, M.; Anderson, J.C.; Santoro, S.W.; Meggers, E.; Herberich, B.; Wang, L.; Schultz, P.G. In *Translation Mechanisms*, Lapointe, J.; Brakier-Gingras, L., Ed. Landes Bioscience: Georgetown, TX, **2003**; pp. 95-114.
- [8] Magliery, T.J.; Liu, D.R. In *The Genetic Code and the Origin of Life*, Ribas de Pouplana, L., Ed. Landes Bioscience: Georgetown, TX, **2004**.
- [9] Wang, L.; Schultz, P.G. *Chem. Commun.* **2002**, *1*.
- [10] Link, A.J.; Mock, M.L.; Tirrell, D.A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603.
- [11] Hohsaka, T.; Sisido, M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 809.
- [12] Muir, T.W. *Annu. Rev. Biochem.* **2003**, *72*, 249.
- [13] Dawson, P.E.; Kent, S.B.H. *Annu. Rev. Biochem.* **2000**, *69*, 923.
- [14] Vogel, K.; Chmielewski, J. *J. Am. Chem. Soc.* **1994**, *116*, 11163.
- [15] Abrahmsen, L.; Tom, J.; Burnier, J.; Butcher, K.A.; Kossiakoff, A.; Wells, J.A. *Biochemistry* **1991**, *30*, 4151.
- [16] Atwell, S.; Wells, J.A. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9497.
- [17] Braisted, A.C.; Judice, J.K.; Wells, J.A. *Methods Enzymol.* **1997**, *289*, 298.
- [18] Chang, T.K.; Jackson, D.Y.; Burnier, J.P.; Wells, J.A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 12544.
- [19] Jackson, D.Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J.A. *Science* **1994**, *266*, 243.
- [20] Schnolzer, M.; Kent, S.B.H. *Science* **1992**, *256*, 221.
- [21] Dawson, P.E.; Muir, T.W.; Clarklewis, I.; Kent, S.B.H. *Science* **1994**, *266*, 776.
- [22] Hofmann, R.M.; Muir, T.W. *Curr. Opin. Biotechnol.* **2002**, *13*, 297.
- [23] Sewing, A.; Hilvert, D. *Angew. Chem., Int. Ed.* **2001**, *40*, 3395.
- [24] Low, D.W.; Hill, M.G.; Carrasco, M.R.; Kent, S.B.H.; Botti, P. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6554.
- [25] Canne, L.E.; Bark, S.J.; Kent, S.B.H. *J. Am. Chem. Soc.* **1996**, *118*, 5891.
- [26] Yan, L.Z.; Dawson, P.E. *J. Am. Chem. Soc.* **2001**, *123*, 526.
- [27] Gieselmann, M.D.; Zhu, Y.T.; Zhou, H.; Galonic, D.; van der Donk, W.A. *ChemBiochem.* **2002**, *3*, 709.
- [28] Zhu, Y.; van der Donk, W.A. *Org. Lett.* **2001**, *3*, 1189.
- [29] Hackeng, T.M.; Griffin, J.H.; Dawson, P.E. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10068.
- [30] Canne, L.E.; Botti, P.; Simon, R.J.; Chen, Y.J.; Dennis, E.A.; Kent, S.B.H. *J. Am. Chem. Soc.* **1999**, *121*, 8720.
- [31] Gaertner, H.F.; Rose, K.; Cotton, R.; Timms, D.; Camble, R.; Oford, R.E. *Bioconj. Chem.* **1991**, *3*, 262.
- [32] Rose, K. *J. Am. Chem. Soc.* **1994**, *116*, 30.
- [33] Muir, T.W.; Sondhi, D.; Cole, P.A. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705.
- [34] Villain, M.; Vizzavona, J.; Rose, K. *Chem. Biol.* **2001**, *8*, 673.
- [35] Erlanson, D.A.; Chytil, M.; Verdine, G.L. *Chem. Biol.* **1996**, *3*, 981.
- [36] Morris, M.C.; Depollier, J.; Mery, J.; Heitz, F.; Divita, G. *Nat. Biotechnol.* **2001**, *19*, 1173.
- [37] Morris, M.C.; Chaloin, L.; Heitz, F.; Divita, G. *Curr. Opin. Biotechnol.* **2000**, *11*, 461.
- [38] Crick, F.H.C. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138.

- [39] Crick, F.H.C.; Barret, L.; Brenner, S.; Watts-Tobin, R. *Nature* **1961**, *192*, 1227.
- [40] Bossi, L.; Roth, J.R. *Nature* **1980**, *286*, 123.
- [41] Chapeville, F.; Lipmann, F.; von Ehrenstein, G.; Weisblum, B.; Ray, W.J., Jr.; Benzer, S. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 1086.
- [42] Steward, L.E.; Collins, C.S.; Gilmore, M.A.; Carlson, J.E.; Ross, J.B.A.; Chamberlin, A.R. *J. Am. Chem. Soc.* **1997**, *119*, 6.
- [43] Thorson, J.S.; Cornish, V.W.; Barrett, J.E.; Cload, S.T.; Yano, T.; Schultz, P.G. In *Protein Synthesis: Methods and Protocols*, Martin, R., Ed. Humana Press: Totowa, NJ, **1998**; Vol. 77, pp. 43-73.
- [44] Noren, C.J.; Anthony-Cahill, S.J.; Griffith, M.C.; Schultz, P.G. *Science* **1989**, *244*, 182.
- [45] Bain, J.D.; Glabe, C.G.; Dix, T.A.; Chamberlin, A.R.; Diala, E.S. *J. Am. Chem. Soc.* **1989**, *111*, 8013.
- [46] Payne, R.C.; Nichols, B.P.; Hecht, S.M. *Biochemistry* **1987**, *26*, 3197.
- [47] Baldini, G.; Martoglio, B.; Schachenmann, A.; Zugliani, C.; Bruner, J. *Biochemistry* **1988**, *27*, 7951.
- [48] Robertson, S.A.; Noren, C.J.; Anthonycahill, S.J.; Griffith, M.C.; Schultz, P.G. *Nucl. Acids Res.* **1989**, *17*, 9649.
- [49] Robertson, S.A.; Ellman, J.A.; Schultz, P.G. *J. Am. Chem. Soc.* **1991**, *113*, 2722.
- [50] LaRiviere, F.J.; Wolfson, A.D.; Uhlenbeck, O.C. *Science* **2001**, *294*, 165.
- [51] Ninomiya, K.; Kurita, T.; Hohsaka, T.; Sisido, M. *Chem. Commun.* **2003**, 2242.
- [52] Bessho, Y.; Hodgson, D.R.; Suga, H. *Nat. Biotechnol.* **2002**, *20*, 723.
- [53] Murakami, H.; Saito, H.; Suga, H. *Chem. Biol.* **2003**, *10*, 655.
- [54] Murakami, H.; Kourouklis, D.; Suga, H. *Chem. Biol.* **2003**, *10*, 1077.
- [55] Cload, S.T.; Liu, D.R.; Froland, W.A.; Schultz, P.G. *Chem. Biol.* **1996**, *3*, 1033.
- [56] Short, G.F.; Golovine, S.Y.; Hecht, S.M. *Biochemistry* **1999**, *38*, 8808.
- [57] Yokoyama, S. *Curr. Opin. Chem. Biol.* **2003**, *7*, 39.
- [58] Kigawa, T.; Yabuki, T.; Yoshida, Y.; Tsutsui, M.; Ito, Y.; Shibata, T.; Yokoyama, S. *FEBS Lett.* **1999**, *442*, 15.
- [59] Takahashi, T.T.; Austin, R.J.; Roberts, R.W. *Trends Biochem. Sci.* **2003**, *28*, 159.
- [60] Roberts, R.W.; Szostak, J.W. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297.
- [61] Liu, R.; Barrick, J.E.; Szostak, J.W.; Roberts, R.W. *Methods Enzymol.* **2000**, *318*, 268.
- [62] Frankel, A.; Li, S.W.; Starch, S.R.; Roberts, R.W. *Curr. Opin. Struct. Biol.* **2003**, *13*, 506.
- [63] Frankel, A.; Millward, S.W.; Roberts, R.W. *Chem. Biol.* **2003**, *10*, 1043.
- [64] Hanes, J.; Pluckthun, A. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4937.
- [65] Atkins, J.F.; Weiss, R.B.; Thompson, S.; Gesteland, R.F. *Annu. Rev. Genet.* **1991**, *25*, 201.
- [66] Anderson, J.C.; Schultz, P.G. *Biochemistry* **2003**, *42*, 9598.
- [67] Kohrer, C.; Xie, L.; Kellerer, S.; Varshney, U.; RajBhandary, U.L. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14310.
- [68] Kanda, T.; Takai, K.; Hohsaka, T.; Sisido, M.; Takaku, H. *Biochem. Biophys. Res. Commun.* **2000**, *270*, 1136.
- [69] Kanda, T.; Takai, K.; Yokoyama, S.; Takaku, H. *FEBS Lett.* **1998**, *440*, 273.
- [70] Kanda, T.; Takai, K.; Yokoyama, S.; Takaku, H. *Bioorg. Med. Chem.* **2000**, *8*, 675.
- [71] Kanda, T.; Takai, K.; Yokoyama, S.; Takaku, H. *J. Biochem. (Tokyo, Jpn.)* **2000**, *127*, 37.
- [72] Jackson, R.J.; Naphthine, S.; Brierley, I. *RNA* **2001**, *7*, 765.
- [73] Frankel, A.; Roberts, R.W. *RNA* **2003**, *9*, 780.
- [74] Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001**, *19*, 751.
- [75] Magliery, T.J.; Anderson, J.C.; Schultz, P.G. *J. Mol. Biol.* **2001**, *307*, 755.
- [76] Anderson, J.C.; Magliery, T.J.; Schultz, P.G. *Chem. Biol.* **2002**, *9*, 237.
- [77] Hohsaka, T.; Ashizuka, Y.; Murakami, H.; Sisido, M. *J. Am. Chem. Soc.* **1996**, *118*, 9778.
- [78] Kramer, G.; Kudlicki, W.; Hardesty, B. In *Protein Synthesis: Methods and Protocols*, Martin, R., Ed. Humana Press: Totowa, NJ, **1998**; Vol. 77, pp. 105-116.
- [79] Ma, C.H.; Kudlicki, W.; Odum, O.W.; Kramer, G.; Hardesty, B. *Biochemistry* **1993**, *32*, 7939.
- [80] Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. *Biochemistry* **2001**, *40*, 11060.
- [81] Hohsaka, T.; Ashizuka, Y.; Sasaki, H.; Murakami, H.; Sisido, M. *J. Am. Chem. Soc.* **1999**, *121*, 12194.
- [82] Hohsaka, T.; Ashizuka, Y.; Murakami, H.; Sisido, M. *Nucl. Acids Res.* **2001**, *29*, 3646.
- [83] Kool, E.T. *Acc. Chem. Res.* **2002**, *35*, 936.
- [84] Henry, A.A.; Romesberg, F.E. *Curr. Opin. Chem. Biol.* **2003**, *7*, 727.
- [85] Switzer, C.; Moroney, S.E.; Benner, S.A. *J. Am. Chem. Soc.* **1989**, *111*, 8322.
- [86] Bain, J.D.; Switzer, C.; Chamberlin, A.R.; Benner, S.A. *Nature* **1992**, *356*, 537.
- [87] Piccirilli, J.A.; Krauch, T.; Moroney, S.E.; Benner, S.A. *Nature* **1990**, *343*, 33.
- [88] Rappaport, H.P. *Biochemistry* **1993**, *32*, 3047.
- [89] Rappaport, H.P. *Nucl. Acids Res.* **1988**, *16*, 7253.
- [90] Ohtsuki, T.; Kimoto, M.; Ishikawa, M.; Mitsui, T.; Hirao, I.; Yokoyama, S. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4922.
- [91] Fujiwara, T.; Kimoto, M.; Sugiyama, H.; Hirao, I.; Yokoyama, S. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2221.
- [92] Hirao, I.; Mitsui, T.; Fujiwara, T.; Kimoto, M.; To, T.; Okuni, T.; Sato, A.; Harada, Y.; Yokoyama, S. *Nucl. Acids Res. Suppl.* **2001**, *17*.
- [93] Ishikawa, M.; Hirao, I.; Yokoyama, S. *Nucl. Acids Symp. Ser.* **1999**, 125.
- [94] Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T.; Kodama, K.; Nishikawa, K.; Yokoyama, S. *Nat. Biotechnol.* **2002**, *20*, 177.
- [95] Moran, S.; Ren, R.X.F.; Rumney, S.; Kool, E.T. *J. Am. Chem. Soc.* **1997**, *119*, 2056.
- [96] Morales, J.C.; Kool, E.T. *Nat. Struct. Biol.* **1998**, *5*, 950.
- [97] Mitsui, T.; Kitamura, A.; Kimoto, M.; To, T.; Sato, A.; Hirao, I.; Yokoyama, S. *J. Am. Chem. Soc.* **2003**, *125*, 5298.
- [98] Ogawa, A.K.; Wu, Y.Q.; McMinn, D.L.; Liu, J.Q.; Schultz, P.G.; Romesberg, F.E. *J. Am. Chem. Soc.* **2000**, *122*, 3274.
- [99] Henry, A.A.; Yu, C.Z.; Romesberg, F.E. *J. Am. Chem. Soc.* **2003**, *125*, 9638.
- [100] Tae, E.J.L.; Wu, Y.Q.; Xia, G.; Schultz, P.G.; Romesberg, F.E. *J. Am. Chem. Soc.* **2001**, *123*, 7439.
- [101] Xia, G.; Chen, L.J.; Sera, T.; Fa, M.; Schultz, P.G.; Romesberg, F.E. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6597.
- [102] Zimmermann, N.; Meggers, E.; Schultz, P.G. *Bioorg. Chem.* **2004**, *32*, 13.
- [103] Zimmermann, N.; Meggers, E.; Schultz, P.G. *J. Am. Chem. Soc.* **2002**, *124*, 13684.
- [104] Meggers, E.; Holland, P.L.; Tolman, W.B.; Romesberg, F.E.; Schultz, P.G. *J. Am. Chem. Soc.* **2000**, *122*, 10714.
- [105] Atwell, S.; Meggers, E.; Spraggon, G.; Schultz, P.G. *J. Am. Chem. Soc.* **2001**, *123*, 12364.
- [106] 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.; Lester, H.A. *Science* **1995**, *268*, 439.
- [107] Beene, D.L.; Dougherty, D.A.; Lester, H.A. *Curr. Opin. Neurobiol.* **2003**, *13*, 264.
- [108] Saks, M.E.; Sampson, J.R.; Nowak, M.W.; Kearney, P.C.; Du, F.Y.; Abelson, J.N.; Lester, H.A.; Dougherty, D.A. *J. Biol. Chem.* **1996**, *271*, 23169.
- [109] Monahan, S.L.; Lester, H.A.; Dougherty, D.A. *Chem. Biol.* **2003**, *10*, 573.
- [110] Hendrickson, W.A.; Horton, J.R.; Lemaster, D.M. *EMBO J.* **1990**, *9*, 1665.
- [111] Yang, W.; Hendrickson, W.A.; Crouch, R.J.; Satow, Y. *Science* **1990**, *249*, 1398.
- [112] Budisa, N.; Steipe, B.; Demange, P.; Eckerskorn, C.; Kellermann, J.; Huber, R. *Eur. J. Biochem.* **1995**, *230*, 788.
- [113] Luck, L.A.; Vance, J.E.; Oconnell, T.M.; London, R.E. *J. Biomol. NMR* **1996**, *7*, 261.

- [114] Minks, C.; Huber, R.; Moroder, L.; Budisa, N. *Biochemistry* **1999**, *38*, 10649.
- [115] Minks, C.; Huber, R.; Moroder, L.; Budisa, N. *Anal. Biochem.* **2000**, *284*, 29.
- [116] Jakubowski, H. *J. Biol. Chem.* **2000**, *275*, 21813.
- [117] Budisa, N.; Minks, C.; Medrano, F.J.; Lutz, J.; Huber, R.; Moroder, L. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 455.
- [118] McGrath, K.P.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *J. Am. Chem. Soc.* **1992**, *114*, 727.
- [119] Kothakota, S.; Mason, T.L.; Tirrell, D.A.; Fournier, M.J. *J. Am. Chem. Soc.* **1995**, *117*, 536.
- [120] Wang, P.; Tang, Y.; Tirrell, D.A. *J. Am. Chem. Soc.* **2003**, *125*, 6900.
- [121] Kiick, K.L.; Saxon, E.; Tirrell, D.A.; Bertozzi, C.R. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19.
- [122] Tang, Y.; Tirrell, D.A. *J. Am. Chem. Soc.* **2001**, *123*, 11089.
- [123] van Hest, J.C.M.; Kiick, K.L.; Tirrell, D.A. *J. Am. Chem. Soc.* **2000**, *122*, 1282.
- [124] Kiick, K.L.; van Hest, J.C.M.; Tirrell, D.A. *Angew. Chem. Int. Ed.* **2000**, *39*, 2148.
- [125] Kiick, K.L.; Weberskirch, R.; Tirrell, D.A. *FEBS Lett.* **2001**, *502*, 25.
- [126] Ibba, M.; Hennecke, H. *FEBS Lett.* **1995**, *364*, 272.
- [127] Kirshenbaum, K.; Carrico, I.S.; Tirrell, D.A. *ChemBiochem.* **2002**, *3*, 235.
- [128] Doring, V.; Mootz, H.D.; Nangle, L.A.; Hendrickson, T.L.; de Crecy-Lagard, V.; Schimmel, P.; Marliere, P. *Science* **2001**, *292*, 501.
- [129] Tang, Y.; Tirrell, D.A. *Biochemistry* **2002**, *41*, 10635.
- [130] Rath, V.L.; Silvian, L.F.; Beijer, B.; Sproat, B.S.; Steitz, T.A. *Structure* **1998**, *6*, 439.
- [131] Liu, D.R.; Magliery, T.J.; Schultz, P.G. *Chem. Biol.* **1997**, *4*, 685.
- [132] Patten, P.A.; Howard, R.J.; Stemmer, W.P.C. *Curr. Opin. Biotechnol.* **1997**, *8*, 724.
- [133] Stemmer, W.P.C. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10747.
- [134] Liu, D.R.; Magliery, T.J.; Pasternak, M.; Schultz, P.G. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10092.
- [135] Whelihan, E.F.; Schimmel, P. *EMBO J.* **1997**, *16*, 2968.
- [136] Wang, C.C.; Schimmel, P. *J. Biol. Chem.* **1999**, *274*, 16508.
- [137] Liu, D.R.; Schultz, P.G. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4780.
- [138] Pasternak, M.; Magliery, T.J.; Schultz, P.G. *Helv. Chim. Acta* **2000**, *83*, 2277.
- [139] Magliery, T.J. Ph.D. Thesis, University of California, Berkeley, CA, **2001**.
- [140] Steer, B.A.; Schimmel, P. *J. Biol. Chem.* **1999**, *274*, 35601.
- [141] Steer, B.A.; Schimmel, P. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13644.
- [142] Wang, L.; Magliery, T.J.; Liu, D.R.; Schultz, P.G. *J. Am. Chem. Soc.* **2000**, *122*, 5010.
- [143] Wang, L.; Schultz, P.G. *Chem. Biol.* **2001**, *8*, 883.
- [144] Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. *Nat. Struct. Biol.* **2003**, *10*, 425.
- [145] Kowal, A.K.; Kohrer, C.; Rajbhandary, U.L. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2268.
- [146] Santoro, S.W.; Anderson, J.C.; Lakshman, V.; Schultz, P.G. *Nucl. Acids Res.* **2003**, *31*, 6700.
- [147] Anderson, J.C.; Wu, N.; Santoro, S.W.; Lakshman, V.; King, D.S.; Schultz, P.G. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7566.
- [148] Drabkin, H.J.; Park, H.J.; Rajbhandary, U.L. *Mol. Cell Biol.* **1996**, *16*, 907.
- [149] Kiga, D.; Sakamoto, K.; Kodama, K.; Kigawa, T.; Matsuda, T.; Yabuki, T.; Shirouzu, M.; Harada, Y.; Nakayama, H.; Takio, K.; Hasegawa, Y.; Endo, Y.; Hirao, I.; Yokoyama, S. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9715.
- [150] Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S. *Nucl. Acids Res.* **2002**, *30*, 4692.
- [151] Wakasugi, K.; Quinn, C.L.; Tao, N.; Schimmel, P. *EMBO J.* **1998**, *17*, 297.
- [152] Edwards, H.; Schimmel, P. *Mol. Cell Biol.* **1990**, *10*, 1633.
- [153] Edwards, H.; Trezeguet, V.; Schimmel, P. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 1153.
- [154] Chin, J.W.; Cropp, T.A.; Chu, S.; Meggers, E.; Schultz, P.G. *Chem. Biol.* **2003**, *10*, 511.
- [155] Chin, J.W.; Cropp, T.A.; Anderson, J.C.; Mukherji, M.; Zhang, Z.W.; Schultz, P.G. *Science* **2003**, *301*, 964.
- [156] Zhang, Z.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D.S.; Schultz, P.G. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8882.
- [157] Furter, R. *Protein Sci.* **1998**, *7*, 419.
- [158] Agou, F.; Quevillon, S.; Kerjan, P.; Mirande, M. *Biochemistry* **1998**, *37*, 11309.
- [159] Hong, K.W.; Ibba, M.; Soll, D. *FEBS Lett.* **1998**, *434*, 149.
- [160] Datta, D.; Wang, P.; Carrico, I.S.; Mayo, S.L.; Tirrell, D.A. *J. Am. Chem. Soc.* **2002**, *124*, 5652.
- [161] Zhang, D.; Vaidehi, N.; Goddard, W.A., 3rd; Danzer, J.F.; Debe, D. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6579.
- [162] Wang, L.; Brock, A.; Herberich, B.; Schultz, P.G. *Science* **2001**, *292*, 498.
- [163] Chin, J.W.; Martin, A.B.; King, D.S.; Wang, L.; Schultz, P.G. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020.
- [164] Wang, L.; Brock, A.; Schultz, P.G. *J. Am. Chem. Soc.* **2002**, *124*, 1836.
- [165] Zhang, Z.W.; Wang, L.; Brock, A.; Schultz, P.G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2840.
- [166] Wang, L.; Zhang, Z.W.; Brock, A.; Schultz, P.G. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 56.
- [167] Santoro, S.W.; Wang, L.; Herberich, B.; King, D.S.; Schultz, P.G. *Nat. Biotechnol.* **2002**, *20*, 1044.
- [168] Chin, J.W.; Santoro, S.W.; Martin, A.B.; King, D.S.; Wang, L.; Schultz, P.G. *J. Am. Chem. Soc.* **2002**, *124*, 9026.
- [169] Mehl, R.A.; Anderson, J.C.; Santoro, S.W.; Wang, L.; Martin, A.B.; King, D.S.; Horn, D.M.; Schultz, P.G. *J. Am. Chem. Soc.* **2003**, *125*, 935.
- [170] Deiters, A.; Cropp, T.A.; Mukherji, M.; Chin, J.W.; Anderson, J.C.; Schultz, P.G. *J. Am. Chem. Soc.* **2003**, *125*, 11782.
- [171] Cornish, V.W.; Mendel, D.; Schultz, P.G. *Angew. Chem., Int. Ed.* **1995**, *34*, 621.
- [172] Dougherty, D.A. *Curr. Opin. Chem. Biol.* **2000**, *4*, 645.
- [173] Ellman, J.A.; Volkman, B.F.; Mendel, D.; Schultz, P.G.; Wemmer, D.E. *J. Am. Chem. Soc.* **1992**, *114*, 7959.
- [174] Cornish, V.W.; Benson, D.R.; Altenbach, C.A.; Hideg, K.; Hubbell, W.L.; Schultz, P.G. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2910.
- [175] Taki, M.; Hohsaka, T.; Murakami, H.; Taira, K.; Sisido, M. *Phosphorous, Sulfur Silicon Relat. Elem.* **2002**, *177*, 1929.
- [176] Cowburn, D.; Muir, T.W. *Methods Enzymol.* **2001**, *339*, 41.
- [177] Cotton, G.J.; Muir, T.W. *Chem. Biol.* **2000**, *7*, 253.
- [178] Gallivan, J.P.; Lester, H.A.; Dougherty, D.A. *Chem. Biol.* **1997**, *4*, 739.
- [179] Li, L.T.; Zhong, W.G.; Zacharias, N.; Gibbs, C.; Lester, H.A.; Dougherty, D.A. *Chem. Biol.* **2001**, *8*, 47.
- [180] Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D.T.; Goddard, W.A.; DeGrado, W.F.; Tirrell, D.A. *Biochemistry* **2001**, *40*, 2790.
- [181] Tang, Y.; Ghirlanda, G.; Petka, W.A.; Nakajima, T.; DeGrado, W.F.; Tirrell, D.A. *Angew. Chem., Int. Ed.* **2001**, *40*, 1494.
- [182] Bilgicer, B.; Fichera, A.; Kumar, K. *J. Am. Chem. Soc.* **2001**, *123*, 4393.
- [183] Bilgicer, B.; Xing, X.; Kumar, K. *J. Am. Chem. Soc.* **2001**, *123*, 11815.
- [184] Bilgicer, B.; Kumar, K. *Tetrahedron* **2002**, *58*, 4105.
- [185] Renner, C.; Alefelder, S.; Bae, J.H.; Budisa, N.; Huber, R.; Moroder, L. *Angew. Chem., Int. Ed.* **2001**, *40*, 923.
- [186] Frieden, C. *Biochemistry* **2003**, *42*, 12439.
- [187] Budisa, N.; Rubini, M.; Bae, J.H.; Weyher, E.; Wenger, W.; Golbik, R.; Huber, R.; Moroder, L. *Angew. Chem., Int. Ed.* **2002**, *41*, 4066.
- [188] Bae, J.H.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M.H.J.; Azim, M.K.; Kim, J.S.; Zumbusch, A.; Holak, T.A.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* **2003**, *328*, 1071.
- [189] Wang, L.; Xie, J.M.; Deniz, A.A.; Schultz, P.G. *J. Org. Chem.* **2003**, *68*, 174.
- [190] Offord, R.E.; Gaertner, H.F.; Wells, T.N.; Proudfoot, A.E. *Methods Enzymol.* **1997**, *287*, 348.
- [191] Cornish, V.W.; Hahn, K.M.; Schultz, P.G. *J. Am. Chem. Soc.* **1996**, *118*, 8150.
- [192] Zhang, Z.W.; Smith, B.A.C.; Wang, L.; Brock, A.; Cho, C.; Schultz, P.G. *Biochemistry* **2003**, *42*, 6735.
- [193] Mahal, L.K.; Yarema, K.J.; Bertozzi, C.R. *Science* **1997**, *276*, 1125.

- [194] Hang, H.C.; Bertozzi, C.R. *Acc. Chem. Res.* **2001**, *34*, 727.
- [195] Macmillan, D.; Bertozzi, C.R. *Tetrahedron* **2000**, *56*, 9515.
- [196] Tolbert, T.J.; Wong, C.H. *J. Am. Chem. Soc.* **2000**, *122*, 5421.
- [197] Grogan, M.J.; Pratt, M.R.; Marcaurelle, L.A.; Bertozzi, C.R. *Annu. Rev. Biochem.* **2002**, *71*, 593.
- [198] Liu, H.T.; Wang, L.; Brock, A.; Wong, C.H.; Schultz, P.G. *J. Am. Chem. Soc.* **2003**, *125*, 1702.
- [199] Chin, J.W.; Schultz, P.G. *Chembiochem.* **2002**, *3*, 1135.
- [200] Saxon, E.; Bertozzi, C.R. *Science* **2000**, *287*, 2007.
- [201] Link, A.J.; Tirrell, D.A. *J. Am. Chem. Soc.* **2003**, *125*, 11164.
- [202] Blackwell, H.E.; Sadowsky, J.D.; Howard, R.J.; Sampson, J.N.; Chao, J.A.; Steinmetz, W.E.; O'Leary, D.J.; Grubbs, R.H. *J. Org. Chem.* **2001**, *66*, 5291.
- [203] Ellman, J.A.; Mendel, D.; Schultz, P.G. *Science* **1992**, *255*, 197.
- [204] Eisenhauer, B.M.; Hecht, S.M. *Biochemistry* **2002**, *41*, 11472.
- [205] Wang, B.; Brown, K.C.; Lodder, M.; Craik, C.S.; Hecht, S.M. *Biochemistry* **2002**, *41*, 2805.
- [206] England, P.M.; Lester, H.A.; Davidson, N.; Dougherty, D.A. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11025.
- [207] Pollitt, S.K.; Schultz, P.G. *Angew. Chem., Int. Ed.* **1998**, *37*, 2104.
- [208] Miller, J.C.; Silverman, S.K.; England, P.M.; Dougherty, D.A.; Lester, H.A. *Neuron* **1998**, *20*, 619.
- [209] Mendel, D.; Ellman, J.A.; Schultz, P.G. *J. Am. Chem. Soc.* **1991**, *113*, 2758.
- [210] Flavell, R.R.; Huse, M.; Goger, M.; Trester-Zedlitz, M.; Kuriyan, J.; Muir, T.W. *Org. Lett.* **2002**, *4*, 165.
- [211] Wu, J.W.; Hu, M.; Chai, J.J.; Seoane, J.; Huse, M.; Li, C.; Rigotti, D.J.; Kyin, S.; Muir, T.W.; Fairman, R.; Massague, J.; Shi, Y.G. *Mol. Cell* **2001**, *8*, 1277.
- [212] Huse, M.; Muir, T.W.; Xu, L.; Chen, Y.G.; Kuriyan, J.; Massague, J. *Mol. Cell* **2001**, *8*, 671.
- [213] Huse, M.; Holford, M.N.; Kuriyan, J.; Muir, T.W. *J. Am. Chem. Soc.* **2000**, *122*, 8337.
- [214] Cole, P.A.; Courtney, A.D.; Shen, K.; Zhang, Z.S.; Qiao, Y.F.; Lu, W.; Williams, D.M. *Acc. Chem. Res.* **2003**, *36*, 444.
- [215] Lu, W.; Shen, K.; Cole, P.A. *Biochemistry* **2003**, *42*, 5461.
- [216] Lu, W.; Gong, D.Q.; Bar-Sagi, D.; Cole, P.A. *Mol. Cell* **2001**, *8*, 759.
- [217] Zhang, Z.W.; Gildersleeve, J.; Yang, Y.Y.; Xu, R.; Loo, J.A.; Uryu, S.; Wong, C.H.; Schultz, P.G. *Science* **2004**, *303*, 371.
- [218] Rak, A.; Pylypenko, O.; Durek, T.; Watzke, A.; Kushnir, S.; Brunsveld, L.; Waldmann, H.; Goody, R.S.; Alexandrov, K. *Science* **2003**, *302*, 646.
- [219] Alexandrov, K.; Heinemann, I.; Durek, T.; Sidorovitch, V.; Goody, R.S.; Waldmann, H. *J. Am. Chem. Soc.* **2002**, *124*, 5648.
- [220] Kozlov, I.A.; De Bouvere, B.; Van Aerschot, A.; Herdewijn, P.; Orgel, L.E. *J. Am. Chem. Soc.* **1999**, *121*, 5856.
- [221] Orgel, L.E. *Nature* **1992**, *358*, 203.
- [222] Rosenbaum, D.M.; Liu, D.R. *J. Am. Chem. Soc.* **2003**, *125*, 13924.
- [223] Gartner, Z.J.; Kanan, M.W.; Liu, D.R. *J. Am. Chem. Soc.* **2002**, *124*, 10304.
- [224] Gartner, Z.J.; Liu, D.R. *J. Am. Chem. Soc.* **2001**, *123*, 6961.
- [225] Doyon, J.B.; Snyder, T.M.; Liu, D.R. *J. Am. Chem. Soc.* **2003**, *125*, 12372.
- [226] Halpin, D.R.; Harbury, P.B. *PLoS Biol.* **2004**, *2*, E173.
- [227] Halpin, D.R.; Harbury, P.B. *PLoS Biol.* **2004**, *2*, E174.
- [228] Halpin, D.R.; Lee, J.A.; Wrenn, S.J.; Harbury, P.B. *PLoS Biol.* **2004**, *2*, E175.

Received: February 02, 2004

Accepted: October 10, 2004

This article is an update of the original article published in *Current Pharmaceutical Biotechnology*, Vol. 3, No. 4, December 2002, 299-315.