Chapter 1

Introduction

Role of Aminoacyl-tRNA Synthetases (AARS) in Cellular Protein Synthesis

Cellular protein synthesis is directed by the genetic code and takes place in two major steps. The first step, transcription, is the synthesis of messenger RNA (mRNA) from DNA. This is followed by translation, the synthesis of protein from amino acids as directed by the mRNA. Transcription is in some ways similar to DNA replication: chemical structures and polymer properties of RNA and DNA are similar enough that the ability of complementary sequences to form duplexes through base pairing exists for both DNA and RNA molecules, as well as DNA-RNA hybrids. Hence, DNA is an intrinsic template for RNA polymerization, as it is for its own replication. However, such sequence complementarity does not exist between amino acids that have diverse chemical structures and their polymerization template, the mRNA in the translation step. Therefore a translator is required in order to match every specific code on the mRNA (codon) to its corresponding amino acid.

Transfer RNA (tRNA) molecules act as adapters between amino acids and the codons on the mRNA in translation. Two recognition events characterize this process: codon recognition by tRNA and amino acid–tRNA matching. The recognition of a codon on the mRNA by the complementary anticodon on the tRNA is straightforward due to base pairing. The process of matching the correct amino acid with the correct tRNA is more difficult and is overseen by a class of enzymes called aminoacyl-tRNA synthetases (AARS). These enzymes catalyze the reaction that attaches an amino acid to the 3'-end of its corresponding tRNA molecule. Aminoacyl-tRNA synthesis takes place in two steps:

$$\text{AARS} + \textit{aa} + \text{ATP} \Leftrightarrow \text{AARS}\bullet\textit{aa}\text{-AMP} + \text{PP}_i \,, \qquad (1.1)$$

$$\text{AARS}\bullet\textit{aa}\text{-AMP} + \text{tRNA}^{aa} \Leftrightarrow \textit{aa}\text{-tRNA}^{aa} + \text{AARS} + \text{AMP} \,, \qquad (1.2)$$

where *aa* is an amino acid, and AARS and tRNA^{aa} are the corresponding aminoacyl-tRNA synthetase and tRNA, respectively. In the first step, the acid group on the amino acid is activated by forming a mixed anhydride with the α-phosphate on ATP, resulting in a pyrophosphate leaving group. In the second step, AMP is displaced by a nucleophilic attack from the 2'- or 3'-hydroxyl of the tRNA to give aminoacyl-tRNA [1]. These steps are catalyzed by the

corresponding AARS enzyme. To ensure the fidelity of information transfer from the nucleus to functional cellular proteins, the AARS has to be able to recognize with high accuracy the cognate amino acid out of all the small molecules in the cell, as well as the correct tRNA molecules from the 20 classes present. Upon binding, AARSs make extensive contacts with the tRNA molecules over a large surface area (2,470 to 5,650 $Å^2$) [2, 3]. Such large areas of interaction enable AARSs to distinguish the sequence specific features on the tRNAs [4] and ensure correct recognition. Hence, the error rates in tRNA recognition are usually less than 1 in 10^6 [5]. Amino acids, however, are small molecules and have much fewer features that distinguish one from another. Particularly AARSs that aminoacylate tRNAs with Ile, Thr, and Ala need to discriminate against amino acids that have one fewer methyl group, Val, Ser, and Gly, respectively. Pauling [6] calculated that the binding-energy difference between Ile and Val for the isoleucyl-tRNA synthetase (IleRS) could not result in an error rate better than 1 in 5 for the misactivation of Val, based on the contribution of the extra methyl group on Ile. However, in reality, this rate is only 1 in 3000 for IleRS [7] and around 1 in 10^4 to 10^5 for AARSs in general [8]. This observed high specificity of the AARSs for their cognate amino acid is ensured by additional proofreading or editing activities that hydrolyse the product when noncognate amino acids are activated or charged onto tRNAs.

Recognition of the cognate amino acid by the AARS is critical for the fidelity of the translation process. Theoretical models show that, rather than ribosomal errors, tRNA charging by the AARSs plays a key role in controlling the long-term stability of the cell. Error-prone ribosomes are more likely to produce inactive protein products [9], whereas mistakes in tRNA charging can lead to an increase in error rates by producing “mutant” AARSs in the next generation protein synthesis. A high AARS error rate was predicted to cause an “error catastrophe” [10]. It was recently shown that the viability of cells containing different error-prone valyl-tRNA synthetases (ValRS) is decreased in proportion to ValRS error rates when competitor (α-aminobutyrate) concentrations were increased [11].

Incorporation of Noncanonical Amino Acids into Proteins

The lethality caused by errors from AARS is in part due to the absence of a control mechanism that checks for correct amino acid-tRNA pairing at the ribosome level. The lack of error correction allows amino acids on misacylated tRNAs to be incorporated into proteins. At the same time, this also makes it possible to bypass the recognition and charging steps altogether by chemically charging the tRNAs with the amino acid of choice *in vitro* and providing them directly to the translational machinery. The preparation of chemically misacylated tRNAs was first described in 1984 [12]. In 1989, Schultz and Chamberlin groups independently combined this technique with cell-free translation methods, and successfully incorporated noncanonical amino acids into proteins at amber codon (UAG) sites [13, 14].

The 20 natural amino acids represent a very limited set of structures and functionalities present in organic chemistry. The introduction of noncanonical amino acids expands this set to compounds with diverse electronic and molecular structures. By specifying the exact chemical structure on the amino acid side chain, rather than mutating whole residues, perturbations can be applied to individual atoms in a protein, allowing us to probe protein structure and function more accurately. The ability to introduce noncanonical amino acids also allows unique chemical functionalities and reactions, such as light triggered cross-linking, to be applied to proteins. Due to the enormous potential of noncanonical amino acids, methods for their incorporation drew considerable attention. Since 1989 more than a hundred noncanonical amino acids have been incorporated into proteins using amber suppression and other similar methods [15]. Noncanonical amino acids incorporated this way include structural probes and fluorescent, isotopic and spin labels [16-18]. Successful incorporation was reported both *in vitro* [19] and *in vivo*, through microinjection of misacylated tRNA into *Xenopus* oocytes [20]. In oocytes, mutations to noncanonical amino acids were introduced into the nicotinic acetylcholine receptor (nAChR). Receptors that exhibit constitutive activity at acidic pH were discovered through the

incorporation of modified tyrosines [21]. Effects of backbone hydrogen bonding was probed by mutating the amide linkage on the protein backbone to an ester through hydroxyacid incorporation [22]. Recently, *cis-trans* isomerization at a proline residue in nAChR was modulated by the incorporation of proline analogs with varying preferences for the *cis*-conformer [23]. Despite many successes, the above methods present one major limitation: the yield of engineered protein is intrinsically limited by the amount of chemically aminoacylated tRNA, synthesis of which is laborious. Thus, the yield of expressed protein is usually low at around one milligram or less per reaction [24]. The low protein yield is especially a problem if the engineered protein is to be used in biomaterial applications, where material is required in bulk.

Residue-specific Incorporation

Unlike chemical polymerization, the synthesis of proteins is directed precisely by the genetic code. This control over sequence and size translates into a precise control over the properties of the polymer synthesized. Genetic engineering allows us to modify the genetic code, and therefore, engineer novel protein-based materials using cellular protein synthesis machinery [25, 26]. Tirrell and co-workers have utilized these ideas to prepare artificial proteins that have controllable crystallization properties [27], form liquid crystal phases and ordered films [28, 29], reversible hydrogels [30], and matrix proteins for potential use in vascular reconstruction [31-33]. The utilization of noncanonical amino acids with novel side chains has the potential to provide properties to the peptides not possible with the natural set, and expanding the utility of their applications. When changes in bulk properties are desired, site-specific incorporation is not necessary for this case. However, *in vitro* translation with good efficiency is required in order to get a high protein yield.

Multisite (residue-specific) incorporation of amino-acid analogs in auxotrophic cell strains was observed as early as the 1950s [34, 35] and to a large extent suits the needs described above [36]. This method relies on the construction of a cell strain that is unable to synthesize a particular amino acid,

and the ability of the corresponding AARS to activate its structural analogs. First, auxotrophs to a certain amino acid are grown in media containing 20 natural amino acids. At the onset of the stationary growth phase, this amino acid is depleted and is replaced by one of its structural analogs through a media shift. Upon induction of protein synthesis, the analog taken up by the cell is charged by the cellular AARS onto the corresponding tRNAs, replacing the depleted amino acid. This way, the analog is incorporated into all proteins synthesized by the cell in place of the natural amino acid. Through this strategy, noncanonical amino acids that carry alkene [37], alkyne [38], azido [39], fluoro [40-42] and ketone [43] functionalities on their side chains have so far been incorporated into recombinant proteins in *E. coli* for protein engineering applications.

In the last decade, amino acids that carry noncanonical side chains have become useful protein engineering tools. Introduction of selenomethionine and other amino acids carrying heavy atoms into proteins greatly simplifies determination of x-ray crystal structures through multiwavelength anomalous diffraction [44, 45]. Fluorinated amino acids have been used as ^{19}F-NMR probes in proteins [46]. The hyperhydrophobicity of fluorocarbons have led to the construction of proteins with fluorous cores [40, 47-54]. These proteins exhibit large gains in thermostability compared with their non-fluorinated counterparts. The tryptophan analog 4-amino-tryptophan was used to modify the chromophore of the enhanced cyan fluorescent protein (ECFP) [55]. The resulting protein exhibited a redshift not achieved by laboratory evolution or screening.

The value and versatility of noncanonical amino acids have significantly increased through their use as bioorthogonal reactive handles for conjugation [56, 57], detection [58, 59] and selective isolation [60] of proteins. Upon incorporation into proteins, aryl halides allow palladium-catalyzed Heck and Sonogashira couplings [61, 62], and the acetyl group can be reacted selectively with hydrazides [43]. Among the reactive groups introduced into proteins, azides are perhaps the most significant [63]. Azides can be ligated under physiological conditions to ester-functionalized triaryl phosphines through the Staudinger ligation or to terminal alkynes through Cu(I)-catalyzed [3+2] azide-alkyne

cycloaddition, both in bioorthogonal fashion. Recently a strain-promoted version of the azide-alkyne cycloaddition that does not require the Cu(I) catalyst was reported [64], which allows this chemistry to be applied to live cells. The utility of the introduction of bioorthogonally reactive amino acids into proteins was demonstrated through their application to the study of proteomic changes in cells in response to stimuli [58-60]. Proteins synthesized following a stimulus were metabolically labeled with azide- or alkyne-bearing methionine analogs. Through the use of affinity tags for separation and identification, or fluorescent tags for visualization, the newly synthesized proteins were studied. Photoreactive aryl azides, such as *p*-azidophenylalanine were incorporated into proteins, and used in the construction of ligand-bearing [56] or lithographically patterned surfaces [57].

Modification of AARS Activities

The most prominent limitation for the *in vivo* incorporation of noncanonical amino acids is its reliance on the host's native AARSs to charge the amino acid of interest on to tRNAs. This requirement limits the set of noncanonical amino acids that could be incorporated to only those that are similar in size, shape and electrostatic structure to the natural amino acids they replace. Therefore, the diversity in structural and chemical functionality of analogs obtained using chemical misacylation-based techniques is limited. In cases where the associated AARS shows low activity for the desired analog, it has been possible to increase the overall activity by engineering *E. coli* hosts that overexpress the AARS of choice to achieve incorporation [65]. It is also possible that the analog of choice is activated or charged, but then is hydrolyzed through editing/proofreading. Disrupting the editing activity of the AARSs resulted in successful activation of amino acid analogs in valyl- [66] and leucyl-tRNA synthetases [67]. However, many noncanonical amino acids differ significantly from their natural counterparts in structure and are not substrates for natural AARSs. For these molecules, it is necessary to engineer AARS activity specifically. Point mutations at the active sites of certain AARSs, such as

phenylalanyl- [68] and tyrosyl-tRNA synthetase [69], are known to relax the binding specificity of AARS and allow the incorporation of amino acid analogs that are normally not incorporated. Enlarging the amino acid binding pocket of the *E. coli* phenylalanyl-tRNA synthetase (PheRS) by the A294G mutation made the introduction of various *para*-substituted phenylalanine analogs into proteins [61, 62, 70], as well as the photoreactive amino acid benzofuranyl-alanine [71] possible. An additional mutation (T251G) identified in the PheRS binding pocket by a computational protein design algorithm allowed the introduction of *p*-acetyl-phenylalanine into proteins [43]. Carrying this mutation over to the yeast homolog of this enzyme (T415G) enabled the site-specific incorporation of a series of tryptophan analogs in *E. coli* through the use of this enzyme with a mutant yeast phenylalanine amber suppressor tRNA [72]. A single mutation (C443G) in the proline binding pocket of *E. coli* prolyl-tRNA synthetase (ProRS) also permitted the incorporation of a series of proline surrogates, including *2S*-pipecolic acid, which is a proline analog with a six-membered ring [73].

Schultz and co-workers have employed directed evolution in order to identify mutations on AARSs that lead to specific activity toward a noncanonical amino acid of interest. Their studies focused on the modification of tyrosyl-tRNA synthetase from the archaebacterium *Methanococcus jannaschii* (*Mj*-TyrRS), for its use with its associated $tRNA^{Tyr}$ as a 21^{st} AARS-tRNA pair in *E. coli* in their site-specific *in vivo* incorporation studies [74]. In order to set a clear background for their screen, five residues thought to be in the active site were mutated to alanines. Using oligonucleotide-encoded DNA shuffling, a library of *Mj*-TyrRS mutants was created on these five positions, which was then screened in two steps. Positive selection for active *Mj*-TyrRS mutants was carried out in the first step, where *Mj*-TyrRS activity is required for cells to survive in antibiotic supplemented media. In the second step, negative selection is applied against the *Mj*-TyrRS mutants that retain their activity toward tyrosine. Here, a lethal gene is introduced into the cells, which are then grown in media that contains tyrosine but not its unnatural analog. Cells containing any *Mj*-TyrRS that is active toward tyrosine express the lethal gene and die. After several cycles of

mutagenesis and a two-step selection, the cells surviving the final positive selection are analyzed. This method has produced *Mj*-TyrRS mutants that exhibit high specificities to tyrosine analogs, and allowed the *in vivo* site-specific incorporation of a wide variety of noncanonical amino acids [75-78], including a glucose-modified serine [79]. Application of this methodology to a *E. coli* $tRNA^{Leu}$-leucyl-tRNA synthetase (LeuRS) pair enabled identification of LeuRS variants with diverse activities [80-83] and achieved site-specific incorporation of fluorescent and photocaged amino acids in yeast, among others.

In order to enable the incorporation of translationally silent methionine (Met) analogs into proteins, several screening strategies were developed for methionyl-tRNA synthetase (MetRS) libraries in the Tirrell group. Following their demonstration of the display of unnatural amino acids in *E. coli* cell surface [84], Link and co-workers showed that the display of unnatural amino acids on the cell surface could be used as a reporter for the cellular AARS activity toward an unnatural substrate [85]. Four residues in the *E. coli* MetRS were randomized and the MetRS library was screened for the incorporation of an azide-bearing amino acid (azidonorleucine, Anl) into the surface-accessible Met positions on *E. coli* outer-membrane protein C (OmpC). Only the clones carrying MetRS variants active toward Anl can display azide groups on the cell surface and exhibit fluorescence upon labeling with a fluorescent probe. These cells can be enriched on a fluorescence-activated cell sorter (FACS). Analysis of the enriched clones revealed that a single mutation to glycine (L13G) was sufficient to relax the specificity of MetRS to allow azidonorleucine incorporation. The above methodology requires bioorthogonal derivatization of reactive side chains presented on the cell surface. Yoo and Tirrell reported a more generally applicable strategy that takes advantage of a green fluorescent protein (GFP) variant, which was engineered to permit the incorporation of methionine analogs without any loss of fluorescence [86]. A five-position saturation-mutagenesis library was expressed in cells overproducing this GFP variant in the presence of 6,6,6-trifluoronorleucine (Tfn). Cells exhibiting the highest yields of GFP were isolated by FACS. The enriched population was also subjected to negative

selection, where GFP production was induced in the presence of 19 amino acids (-Met) and cells with low fluorescence were retained. After multiple rounds of selection, the L13S-Y260L-H301L mutant of MetRS was enriched, and quantitative replacement of Met with Tfn was demonstrated through the use of this mutant.

Computational Protein Design and Its Application to AARSs

The success of redirecting AARS activity to noncanonical amino acids through mutations in the binding pocket suggests that changing the binding specificity alone may be sufficient for new AARS activities. Not surprisingly, computational methods utilized for the identification of new AARS activities have focused on optimizing the binding interactions of the new substrate with the binding pocket. Fidelity of phenylalanyl- [87], methionyl- [88], and seryl-tRNA synthetases [89] to binding their natural substrates has been reliably demonstrated through computational methods. The relationship between enzymatic activities of AARSs toward unnatural ligands, and the corresponding computational binding energies was investigated in several studies. Wang et al. reported a good correlation between the translational activities of a set of phenylalanine analogues in *E. coli* and the binding energies of these analogs calculated based on a phenylalanyl-tRNA synthetase crystal structure from *T. thermophilus* (*Tt*-PheRS) [90]. In a similar study, Datta et al. showed that binding energies computed between *E. coli* MetRS and a series of methionine analogs agree with experimental free energies calculated based on kinetic parameters of activation [88]. In each case, the authors presented well-behaved scoring functions that can predict AARS binding, and therefore, the likelihood of incorporation, which suggests the possibility of computationally designing AARS binding sites for each analog of interest.

One such example has been presented by Datta et al. [43], who used a previously established design algorithm (ORBIT) [91, 92], which utilizes an empirical force field, a pairwise-decomposable description of energies and optimization methods based on the dead-end elimination theorem. Eleven

binding site residues were scanned to identify two mutations on the *Tt*-PheRS structure that should assist *p*-acetyl-phenylalanine binding. Corresponding mutations on the *E. coli* PheRS were then shown to allow the incorporation of this analog into proteins in *E. coli*. A different strategy was employed by Zhang et al. [93, 94], who focused on using a full molecular mechanics force field and a scoring function based on binding energies. Unlike ORBIT, this method relies on keeping the combinatorial complexity of the problem low by focusing on the few residues around the ligand. The success of Schultz and co-workers in selecting high fidelity enzymes from libraries of 5-fold mutants verifies this approach. In this method, named "clash-opportunity progressive design" (COP), first the mutations that relieve clashes and then those that improve binding were identified. These mutations were carried out combinatorially, and from this set, the mutants that prefer binding to the analog over any natural amino acids were identified. COP was successful in identifying two key mutations that allow the binding of *O*-methyl-tyrosine to *Mj*-TyrRS, previously discovered by Wang et al. [75]. Even though binding specificity of analogs to AARS is a necessary but not sufficient condition for their activation, the results above show that computational methods may be an alternative to library screening in altering AARS activity.

Discovery of MetRS Mutants Active Toward Azidonorleucine

Although attempts were made to design AARS binding sites, mutational data to verify the computational procedures have been a limiting factor in these studies. Thus, the experimental screens developed for the *E. coli* MetRS system [85, 86] and the encouraging results from computational studies of unnatural AARS activities [88, 93] together present a unique opportunity to evaluate the effectiveness of computational procedures. In this thesis, we take advantage of the methods developed in previous work to a explore a MetRS library both experimentally and computationally.

The work presented here has two main goals: identification of a variety of MetRS mutants that allow the efficient incorporation of azidonorleucine (Anl) into proteins, and evaluation of computational methods using the experimentally

acquired data. In Chapter 2, we build on the work by Link et al. [85] and identify a wide variety of MetRS mutants that enable Anl incorporation through screens of a three-position MetRS saturation mutagenesis library. In Chapter 3, we show a good agreement between computational binding energies and results from *in vitro* activation kinetics experiments. We also report the results from an *in silico* screen of a MetRS saturation mutagenesis library and identify where the computational model succeeds and fails by comparing the *in vivo* and *in silico* results. In Chapter 4, we discuss the implications of our results to future efforts in engineering new MetRS activities.

References

[1] Ibba M and Soll D, "Aminoacyl-tRNA synthesis," *Annu Rev Biochem*. **2000**, *69*:617.

[2] Nadassy K, Wodak SJ and Janin J, "Structural features of protein-nucleic acid recognition sites," *Biochemistry*. **1999**, *38*:1999.

[3] Rich A and Schimmel PR, "Structural organization of complexes of transfer RNAs with aminoacyl transfer RNA synthetases," *Nucleic Acids Res*. **1977**, *4*:1649.

[4] Cusack S, "Aminoacyl-tRNA synthetases," *Curr Opin Struct Biol*. **1997**, *7*:881.

[5] Sankaranarayanan R and Moras D, "The fidelity of the translation of the genetic code," *Acta Biochimica Polonica*. **2001**, *48*:323.

[6] Pauling L, "Festschrift fur Prof. Dr. Arthur Stoll," Birkhauser Verlag: Basel, **1958**; 597

[7] Loftfield RB, "Frequency of errors in protein biosynthesis," *Biochem J*. **1963**, *89*:82.

[8] Loftfield RB and Vanderjagt D, "Frequency of errors in protein-biosynthesis," *Biochem J*. **1972**, *128*:1353.

[9] Kowald A and Kirkwood TBL, "Accuracy of transfer-RNA charging and codon–anticodon recognition — relative importance for cellular-stability," *J Theor Biol*. **1993**, *160*:493.

[10] Freist W, Sternbach H, Pardowitz I and Cramer F, "Accuracy of protein biosynthesis: Quasi-species nature of proteins and possibility of error catastrophes," *J Theor Biol*. **1998**, *193*:19.

[11] Nangle LA, Lagardt VD, Dodring V and Schimmel P, "Genetic code ambiguity — cell viability related to severity of editing defects in mutant tRNA synthetases," *J Biol Chem*. **2002**, *277*:45729.

[12] Heckler TG, Chang LH, Zama Y, Naka T, Chorghade MS and Hecht SM, "T4 RNA ligase mediated preparation of novel chemically misacylated transfer-RNA^{Phe}s," *Biochemistry*. **1984**, *23*:1468.

[13] Bain JD, Glabe CG, Dix TA, Chamberlin AR and Diala ES, "Biosynthetic site-specific incorporation of a non-natural amino-acid into a polypeptide," *J Am Chem Soc*. **1989**, *111*:8013.

[14] Noren CJ, Anthonycahill SJ, Griffith MC and Schultz PG, "A general-method for site-specific incorporation of unnatural amino-acids into proteins," *Science*. **1989**, *244*:182.

[15] Wang P, "Expanding the biosynthetic capacity of the aminoacyl-tRNA synthetases," *PhD Thesis,* California Institute of Technology (Pasadena, CA). **2003**.

[16] Hohsaka T and Sisido M, "Incorporation of non-natural amino acids into proteins," *Curr Opin Chem Biol*. **2002**, *6*:809.

[17] Dougherty DA, "Unnatural amino acids as probes of protein structure and function," *Curr Opin Chem Biol*. **2000**, *4*:645.

[18] Cornish VW, Mendel D and Schultz PG, "Probing protein-structure and function with an expanded genetic-code," *Angew Chem Int Ed Engl*. **1995**, *34*:621.

[19] Gilmore MA, Steward LE and Chamberlin AR, "Incorporation of noncoded amino acids by *in vitro* protein biosynthesis" in "*Implementation and redesign of catalytic function in biopolymers*," Ed., **1999**; 77

[20] Nowak MW, Kearney PC, Sampson JR, Saks ME, Labarca CG, Silverman SK, Zhong W, Thorson J, Abelson JN, Davidson N, Schultz PG, Dougherty DA and Lester HA, "Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells," *Science*. **1995**, *268*:439.

[21] Petersson EJ, Choi A, Dahan DS, Lester HA and Dougherty DA, "A perturbed pK(a) at the binding site of the nicotinic acetylcholine receptor: Implications for nicotine binding," *J Am Chem Soc*. **2002**, *124*:12662.

[22] England PM, Zhang Y, Dougherty DA and Lester HA, "Backbone mutations in transmembrane domains of a ligand-gated ion channel: Implications for the mechanism of gating," *Cell*. **1999**, *96*:89.

[23] Lummis SC, Beene DL, Lee LW, Lester HA, Broadhurst RW and Dougherty DA, "Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel," *Nature*. **2005**, *438*:248.

[24] Gilmore MA, Steward LE and Chamberlin AR, "Incorporation of noncoded amino acids by *in vitro* protein biosynthesis" in "*Implementation and redesign of catalytic function in biopolymers*," Schmidtchen FP, Ed., Springer: Berlin, **1999**; 77

[25] McGrath KP, Tirrell DA, Kawai M, Mason TL and Fournier MJ, "Chemical and biosynthetic approaches to the production of novel polypeptide materials," *Biotechnol. Prog*. **1990**, *6*:188.

[26] Fournier MJ, Creel HS, Krejchi MT, Mason TL, Tirrell DA, McGrath KP and Atkins EDT, "Genetic synthesis of periodic protein materials," *J Bioact Compat Pol*. **1991**, *6*:326.

[27] Krejchi MT, Atkins ED, Waddon AJ, Fournier MJ, Mason TL and Tirrell DA, "Chemical sequence control of beta-sheet assembly in macromolecular crystals of periodic polypeptides," *Science*. **1994**, *265*:1427.

[28] Yu SJM, Conticello VP, Zhang GH, Kayser C, Fournier MJ, Mason TL and Tirrell DA, "Smectic ordering in solutions and films of a rod-like polymer owing to monodispersity of chain length," *Nature*. **1997**, *389*:167.

[29] Yu SM and Tirrell DA, "Thermal and structural properties of biologically derived monodisperse hairy-rod polymers," *Biomacromolecules*. **2000**, *1*:310.

[30] Petka WA, Harden JL, McGrath KP, Wirtz D and Tirrell DA, "Reversible hydrogels from self-assembling artificial proteins," *Science*. **1998**, *281*:389.

[31] Di Zio K and Tirrell DA, "Mechanical properties of artificial protein matrices engineered for control of cell and tissue behavior," *Macromolecules*. **2003**, *36*:1553.

[32] Welsh ER and Tirrell DA, "Engineering the extracellular matrix: A novel approach to polymeric biomaterials. I. Control of the physical properties of

artificial protein matrices designed to support adhesion of vascular endothelial cells," *Biomacromolecules*. **2000**, *1*:23.

[33] Panitch A, Yamaoka T, Fournier MJ, Mason TL and Tirrell DA, "Design and biosynthesis of elastin-like artificial extracellular matrix proteins containing periodically spaced fibronectin cs5 domains," *Macromolecules*. **1999**, *32*:1701.

[34] Dittmer K, Goering HL, Goodman I and Cristol SJ, "The inhibition of microbiological growth by allylglycine, methallylglycine and crotylglycine," *J Am Chem Soc*. **1948**, *70*:2499.

[35] Rennert OM and Anker HS, "On incorporation of 5',5',5'-trifluoroleucine into proteins of *E. coli*," *Biochemistry*. **1963**, *2*:471.

[36] Budisa N, Minks C, Alefelder S, Wenger W, Dong FM, Moroder L and Huber R, "Toward the experimental codon reassignment *in vivo*: Protein building with an expanded amino acid repertoire," *Faseb Journal*. **1999**, *13*:41.

[37] van Hest JCM and Tirrell DA, "Efficient introduction of alkene functionality into proteins *in vivo*," *FEBS Lett*. **1998**, *428*:68.

[38] van Hest JCM, Kiick KL and Tirrell DA, "Efficient incorporation of unsaturated methionine analogues into proteins *in vivo*," *J Am Chem Soc*. **2000**, *122*:1282.

[39] Kiick KL, Saxon E, Tirrell DA and Bertozzi CR, "Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation," *Proc Natl Acad Sci U S A*. **2002**, *99*:19.

[40] Tang Y, Ghirlanda G, Petka WA, Nakajima T, DeGrado WF and Tirrell DA, "Fluorinated coiled-coil proteins prepared *in vivo* display enhanced thermal and chemical stability," *Angew Chem Int Ed Engl*. **2001**, *40*:1494.

[41] Dessipri E, Tirrell DA and Atkins EDT, "Fluorinated poly(alpha, l-glutamate)s," *Macromolecules*. **1996**, *29*:3545.

[42] Yoshikawa E, Fournier MJ, Mason TL and Tirrell DA, "Genetically-engineered fluoropolymers — synthesis of repetitive polypeptides containing p-fluorophenylalalanine residues," *Macromolecules*. **1994**, *27*:5471.

[43] Datta D, Wang P, Carrico IS, Mayo SL and Tirrell DA, "A designed phenylalanyl-tRNA synthetase variant allows efficient *in vivo* incorporation of aryl ketone functionality into proteins," *J Am Chem Soc*. **2002**, *124*:5652.

[44] Hendrickson WA, Horton JR and LeMaster DM, "Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (mad): A vehicle for direct determination of three-dimensional structure," *EMBO J*. **1990**, *9*:1665.

[45] Bae JH, Alefelder S, Kaiser JT, Friedrich R, Moroder L, Huber R and Budisa N, "Incorporation of beta-selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein x-ray crystallography," *J Mol Biol*. **2001**, *309*:925.

[46] Eichler JF, Cramer JC, Kirk KL and Bann JG, "Biosynthetic incorporation of fluorohistidine into proteins in *E. coli*: A new probe of macromolecular structure," *Chembiochem*. **2005**, *6*:2170.

[47] Wang P, Tang Y and Tirrell DA, "Incorporation of trifluoroisoleucine into proteins *in vivo*," *J Am Chem Soc*. **2003**, *125*:6900.

[48] Tang Y and Tirrell DA, "Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host," *J Am Chem Soc*. **2001**, *123*:11089.

[49] Tang Y, Ghirlanda G, Vaidehi N, Kua J, Mainz DT, Goddard III WA, DeGrado WF and Tirrell DA, "Stabilization of coiled-coil peptide domains by introduction of trifluoroleucine," *Biochemistry*. **2001**, *40*:2790.

[50] Bilgicer B, Xing X and Kumar K, "Programmed self-sorting of coiled coils with leucine and hexafluoroleucine cores," *J Am Chem Soc*. **2001**, *123*:11815.

[51] Bilgicer B, Fichera A and Kumar K, "A coiled coil with a fluorous core," *J Am Chem Soc*. **2001**, *123*:4393.

[52] Son S, Tanrikulu IC and Tirrell DA, "Stabilization of bzip peptides through incorporation of fluorinated aliphatic residues," *Chembiochem*. **2006**, *7*:1251.

[53] Lee KH, Lee HY, Slutsky MM, Anderson JT and Marsh ENG, "Fluorous effect in proteins: De novo design and characterization of a four-alpha-helix bundle protein containing hexafluoroleucine," *Biochemistry*. **2004**, *43*:16277.

[54] Lee HY, Lee KH, Al-Hashimi HM and Marsh ENG, "Modulating protein structure with fluorous amino acids: Increased stability and native-like structure conferred on a 4-helix bundle protein by hexafluoroleucine," *J Am Chem Soc*. **2006**, *128*:337.

[55] Bae JH, Rubini M, Jung G, Wiegand G, Seifert MH, Azim MK, Kim JS, Zumbusch A, Holak TA, Moroder L, Huber R and Budisa N, "Expansion of the genetic code enables design of a novel "Gold" Class of green fluorescent proteins," *J Mol Biol*. **2003**, *328*:1071.

[56] Zhang K, Diehl MR and Tirrell DA, "Artificial polypeptide scaffold for protein immobilization," *J Am Chem Soc*. **2005**, *127*:10136.

[57] Carrico IS, Maskarinec SA, Heilshorn SC, Mock ML, Liu JC, Nowatzki PJ, Franck C, Ravichandran G and Tirrell DA, "Lithographic patterning of photoreactive cell-adhesive proteins," *J Am Chem Soc*. **2007**, *129*:4874.

[58] Beatty KE, Liu JC, Xie F, Dieterich DC, Schuman EM, Wang Q and Tirrell DA, "Fluorescence visualization of newly synthesized proteins in mammalian cells," *Angew Chem Int Ed Engl*. **2006**, *45*:7364.

[59] Beatty KE, Xie F, Wang Q and Tirrell DA, "Selective dye-labeling of newly synthesized proteins in bacterial cells," *J Am Chem Soc*. **2005**, *127*:14150.

[60] Dieterich DC, Link AJ, Graumann J, Tirrell DA and Schuman EM, "Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT)," *Proc Natl Acad Sci U S A*. **2006**, *103*:9482.

[61] Kirshenbaum K, Carrico IS and Tirrell DA, "Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues," *Chembiochem*. **2002**, *3*:235.

[62] Sharma N, Furter R, Kast P and Tirrell DA, "Efficient introduction of aryl bromide functionality into proteins *in vivo*," *FEBS Lett*. **2000**, *467*:37.

[63] Baskin JM and Bertozzi CR, "Bioorthogonal click chemistry: Covalent labeling in living systems," *Qsar Comb Sci*. **2007**, *26*:1211.

[64] Agard NJ, Prescher JA and Bertozzi CR, "A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of blomolecules in living systems," *J Am Chem Soc*. **2004**, *126*:15046.

[65] Kiick KL, van Hest JCM and Tirrell DA, "Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host," *Angew Chem Int Ed Engl*. **2000**, *39*:2148.

[66] Doring V, Mootz HD, Nangle LA, Hendrickson TL, de Crecy-Lagard V, Schimmel P and Marliere P, "Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway," *Science*. **2001**, *292*:501.

[67] Tang Y and Tirrell DA, "Attenuation of the editing activity of the *Escherichia coli* leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins *in vivo*," *Biochemistry*. **2002**, *41*:10635.

[68] Ibba M, Kast P and Hennecke H, "Substrate specificity is determined by amino acid binding pocket size in *Escherichia coli* phenylalanyl transfer RNA synthetase," *Biochemistry*. **1994**, *33*:7107.

[69] Hamano-Takaku F, Iwama T, Saito-Yano S, Takaku K, Monden Y, Kitabatake M, Soll D and Nishimura S, "A mutant *Escherichia coli* tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine," *J Biol Chem*. **2000**, *275*:40324.

[70] Ibba M and Hennecke H, "Relaxing the substrate specificity of an aminoacyl transfer RNA synthetase allows in vitro and *in vivo* synthesis of proteins containing unnatural amino acids," *Febs Letters*. **1995**, *364*:272.

[71] Bentin T, Hamzavi R, Salomonsson J, Roy H, Ibba M and Nielsen PE, "Photoreactive bicyclic amino acids as substrates for mutant *Escherichia coli* phenylalanyl-tRNA synthetases," *J Biol Chem*. **2004**, *279*:19839.

[72] Kwon I and Tirrell DA, "Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells," *J Am Chem Soc*. **2007**, *129*:10431.

[73] Kim W, George A, Evans M and Conticello VP, "Cotranslational incorporation of a structurally diverse series of proline analogues in an *Escherichia coli* expression system," *Chembiochem*. **2004**, *5*:928.

[74] Wang L and Schultz PG, "Expanding the genetic code," *Chem. Comm*. **2002**, 1.

[75] Wang L, Brock A, Herberich B and Schultz PG, "Expanding the genetic code of *Escherichia coli*," *Science*. **2001**, *292*:498.

[76] Xie J and Schultz PG, "A chemical toolkit for proteins — an expanded genetic code," *Nat Rev Mol Cell Biol*. **2006**, *7*:775.

[77] Wang L and Schultz PG, "Expanding the genetic code," *Angew Chem Int Ed Engl*. **2005**, *44*:34.

[78] Wang L, Xie J and Schultz PG, "Expanding the genetic code," *Annu Rev Biophys Biomol Struct*. **2006**, *35*:225.

[79] Zhang ZW, Gildersleeve J, Yang YY, Xu R, Loo JA, Uryu S, Wong CH and Schultz PG, "A new strategy for the synthesis of glycoproteins," *Science*. **2004**, *303*:371.

[80] Wu N, Deiters A, Cropp TA, King D and Schultz PG, "A genetically encoded photocaged amino acid," *J Am Chem Soc*. **2004**, *126*:14306.

[81] Tippmann EM and Schultz PG, "A genetically encoded metallocene containing amino acid," *Tetrahedron*. **2007**, *63*:6182.

[82] Lemke EA, Summerer D, Geierstanger BH, Brittain SM and Schultz PG, "Control of protein phosphorylation with a genetically encoded photocaged amino acid," *Nat Chem Biol*. **2007**, *3*:769.

[83] Summerer D, Chen S, Wu N, Deiters A, Chin JW and Schultz PG, "A genetically encoded fluorescent amino acid," *Proc Natl Acad Sci U S A*. **2006**, *103*:9785.

[84] Link AJ and Tirrell DA, "Cell surface labeling of *Escherichia coli* via copper(I)-catalyzed [3+2] cycloaddition," *J Am Chem Soc*. **2003**, *125*:11164.

[85] Link AJ, Vink MK, Agard NJ, Prescher JA, Bertozzi CR and Tirrell DA, "Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids," *Proc Natl Acad Sci U S A*. **2006**, *103*:10180.

[86] Yoo TH and Tirrell DA, "High-throughput screening for methionyl-tRNA synthetases that enable residue-specific incorporation of noncanonical amino acids into recombinant proteins in bacterial cells," *Angew Chem Int Ed Engl*. **2007**, *46*:5340.

[87] Kekenes-Huskey PM, Vaidehi N, Floriano WB and Goddard WA, "Fidelity of phenylalanyl-tRNA synthetase in binding the natural amino acids," *J Phys Chem B*. **2003**, *107*:11549.

[88] Datta D, Vaidehi N, Zhang DQ and Goddard WA, "Selectivity and specificity of substrate binding in methionyl-tRNA synthetase," *Protein Sci*. **2004**, *13*:2693.

[89] McClendon CL, Vaidehi N, Kam VWT, Zhang DQ and Goddard WA, "Fidelity of seryl-tRNA synthetase to binding of natural amino acids from hierdock first principles computations," *Protein Eng Des Sel*. **2006**, *19*:195.

[90] Wang P, Vaidehi N, Tirrell DA and Goddard WA, "Virtual screening for binding of phenylalanine analogues to phenylalanyl-tRNA synthetase," *J Am Chem Soc*. **2002**, *124*:14442.

[91] Dahiyat BI and Mayo SL, "De novo protein design: Fully automated sequence selection," *Science*. **1997**, *278*:82.

[92] Street AG and Mayo SL, "Computational protein design," *Structure Fold Des*. **1999**, *7*:R105.

[93] Zhang DQ, Vaidehi N, Goddard WA, Danzer JF and Debe D, "Structure-based design of mutant methanococcus jannaschii tyrosyl-tRNA synthetase for incorporation of o-methyl-l-tyrosine," *Proc Natl Acad Sci U S A*. **2002**, *99*:6579.

[94] Zhang DQ, "Structure-based design of mutant proteins," *PhD Thesis,* California Institute of Technology (Pasadena, CA). **2002**.