DOI: 10.1002/anie.201005680

Expansion of the Genetic Code

In Vivo Incorporation of Multiple Noncanonical Amino Acids into Proteins

*Michael G. Hoesl and Nediljko Budisa**

genetic code · noncanonical amino acids · orthogonal ribosomes · protein engineering · reprogrammed translation

Expansion of the standard genetic code enables the design of recombinant proteins with novel and unusual properties. Traditionally, such proteins have contained only a single type of noncanonical amino acid (NCAA) in their amino acid sequence. However, recently reported initial efforts demonstrate that it is possible with suppression-based methods to translate two chemically distinct NCAAs into a single recombinant protein by combining the suppression of different termination codons and nontriplet coding units (such as quadruplets). The possibility of expanding the code with various NCAAs simultaneously further increases the toolkit for the generation of multifunctionalized proteins.

1. Introduction

Chemical methods that produce selectively modified proteins upon the external addition of reagents gave rise to the concept of a "chemical mutation"[1] as early as the 1960s. In this tradition, genetic-code engineering and expansion established in the 1990s enable genetically encoded chemical mutations by extending the flow of genetic information.[2] Nature also uses this strategy.[3] On the nucleic acid level, the flexibility associated with the interpretation of the genetic code has been unambiguously documented.[4] The most remarkable natural alterations in decoding are the incorporation of selenocysteine (Sec)[5] and pyrrolysine (Pyl)[6] into proteins in response to codons generally assigned to translation termination. Such recoding events are found in cells which possess several natural tRNAs that are capable of reading termination or nonsense codons in a specific sequence context.[7] In this way, the in vivo translation of UGA and UAG signals as sense codons is possible, and the protein-synthesis machinery can tolerate the addition of novel amino acids to the standard repertoire.

At the conceptual level, genetic-code engineering and expansion are concerned with the possibilities of adding either novel amino acids to the existing amino acid repertoire and/or novel base pairs to existing DNA/RNA. Such engineering and expansion requires reprogramming of the protein translation machinery by the assignment of new codons or reassignment of existing codons to noncanonical amino acids (NCAAs). Two general in vivo approaches are currently available for the addition of NCAAs or the replacement of canonical amino acids with NCAAs. "Genetic-code engineering" refers to the in vivo residue-specific incorporation of different NCAAs into target protein sequences by sense-codon reassignment.[8] In other words, code engineering comprises the cotranslational replacement of canonical amino acids with NCAAs in auxotrophic host cells.[9] In contrast, "genetic-code expansion" considers termination or nontriplet coding units as "blanks" for the site-specific addition of NCAAs to the existing amino acid repertoire.[10] Both approaches exploit the substrate tolerance of various cellular components, as shown in Figure 1.

Until recently, genetic-code engineering provided only one type of NCAA per target protein. Thereby, either the biophysical properties of a certain protein, for example, fluorescence[11] or folding behavior,[12] could be changed or a bioorthogonal reactive handle[13] for subsequent protein modification could be introduced. However, the combination of all these possibilities in one protein is highly desirable. It is

[*] M. G. Hoesl, Prof. Dr. N. Budisa
Molecular Biotechnology, Max Planck Institute of Biochemistry
Am Klopferspitz 18, 82152 Martinsried (Germany)
Fax: (+49) 89-8578-3557
Prof. Dr. N. Budisa
Department of Biocatalysis, Institute of Chemistry
Technical University Berlin
Franklinstrasse 29, 10587 Berlin (Germany)
E-mail: budisa@biocat.tu-berlin.de
Homepage: http://www.biocat.tu-berlin.de

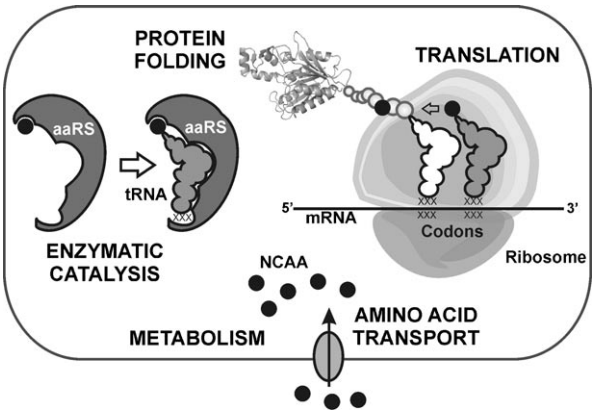

Figure 1. A general NCAA-incorporation experiment exploits the substrate tolerance of various cellular systems (uptake, metabolism, and components of the translation apparatus). For more details, see Ref. [8]. The relationship between code engineering and protein folding is discussed extensively in Ref. [19,20] as well as in textbooks.[2] Parts of Figure 1 were kindly provided by Dr. Birgit Wiltschi.

especially important to be able to make a variety of modifications if we take into account that many biological phenomena, such as conformational preferences,[14] enzymatic activity,[15] and dynamic behavior,[16] are based on collective effects of different amino acids at multiple positions in the protein sequence.

In the frame of genetic-code engineering, a simultaneous in vivo substitution of three different canonical amino acids by related NCAAs at up to 24 positions in a protein was established recently.[17,18] However, it would be extremely difficult to carry out equivalent experiments (i.e. simultaneous multiple in vivo addition at up to 24 positions) with currently available tools for genetic-code expansion. The efficient in vivo addition of novel amino acids to the standard repertoire is a challenging task because many translational and even other cellular components must be engineered or manipulated (see Figure 1). In this context, we present herein a critical overview of the most recent attempts at the multiple site-specific NCAA incorporation of two different NCAAs by the reassignment of termination and quadruplet codons. Finally, we sketch possible future developments in this field.

2. Orthogonal Pairs for Genetic-Code Expansion In Vivo

Genetic-code expansion as an experimental strategy aims to generate novel coding units for the addition of NCAAs to the standard cellular repertoire. In the simplest scenario, one or two stop codons can be considered as "blank" and used uniquely to encode specific NCAAs. The codons UAG (amber), UAA (ochre), or UGA (opal) can be used in this way because their termination function can be suppressed by a special class of adaptors known as suppressor tRNAs. In nature, these suppressor tRNAs can insert different canonical amino acids in response to either nonsense (one of the three terminator codons appears in the mRNA) or missense codons (alteration of the meaning of a sense codon) or a frameshift mutation in the parent gene.[21] In in vitro translation, chemically or enzymatically misacylated suppressor tRNAs are widely used as molecular tools to manipulate translation. For example, Hohsaka and Sisido[22] and Forster et al.[23] pioneered the use of misacylated suppressor tRNAs for the multiple, site-specific insertion of up to three amino acid analogues into a single protein by using sense, termination, or frameshifted (i.e. nontriplet) codons. Thus, it should be generally possible to manipulate the nature and lengths of the basic coding units in one gene in vivo, because the ribosomal machinery can handle codon/anticodon pairs greater than three nucleotides in length in vitro.[24] Excellent studies dealing with in vitro reassignments have been reported.[25]

The application of these strategies in living cells cannot make use of chemically acylated tRNAs but requires the evolution of novel aminoacyl-tRNA synthetases (aaRSs) capable of specifically charging a cognate tRNA with an NCAA. Such aaRS:tRNA pairs should be orthogonal; that is, there should not be any cross-reactivity between heterologous and endogenous aaRSs and tRNAs. The first systematic efforts towards the de novo design of a truly orthogonal aaRS:tRNA pair based on *Escherichia coli* glutaminyl-tRNA synthetase (GlnRS:$tRNA^{Gln}$) were not successful.[26] Therefore, Furter developed an alternative strategy that provided a first proof-of-principle for the site-specific in vivo incorporation of NCAAs in response to an in-frame UAG stop codon.[27] His approach was to import the *Saccharomyces cerevisiae* phenylalanyl-tRNA synthetase pair (PheRS:

Nediljko Budisa received his PhD in 1997 in the group of Robert Huber at the Max Planck Institute of Biochemistry in Martinsried (Germany). From 1997 to 2010 he continued there as a postdoctoral researcher and then as an assistant professor and research-group leader with financial support from the BioFuture funding program of the BMBF. He is now head of the Department of Biocatalysis at the Technische Universität Berlin. His research at the core of synthetic biology aims to provide a solid basis for the laboratory evolution of synthetic life forms with novel chemical possibilities.

Michael Hösl studied molecular biotechnology at the Technische Universität München (University of Technology, Munich), where he received his masters degree in 2007. Currently he is a PhD student in the research group of Prof. Nediljko Budisa, where he is working on method development for the incorporation of noncanonical amino acids into proteins and protein engineering.

$tRNA_{CUA}^{Phe}$; the natural anticodon GAA was mutated to CUA) into the *E. coli* expression host. The use of an aaRS:tRNA pair from an evolutionarily distant organism, such as *S. cerevisiae*, leads to highly reduced cross-reactivity between the introduced aaRS:tRNA pair and the endogenous *E. coli* elements, since tRNA identity elements evolved in a completely different manner. Subsequently, Schultz and co-workers identified and evolved the orthogonal tyrosyl-tRNA synthetase pair (TyrRS:$tRNA_{CUA}^{Tyr}$) from *Methanocaldococcus jannaschii* for use in *E. coli*.[28] In the following years, other orthogonal aaRS:tRNA pairs based on archaeal leucyl-tRNA synthetase (LeuRS:$tRNA^{Leu}$),[29] glutamyl-tRNA synthetase (GluRS:$tRNA^{Glu}$),[30] lysyl-tRNA synthetase (LysRS:$tRNA^{Lys}$),[31] and *S. cerevisiae* tryptophanyl-tRNA synthetase (TrpRS:$tRNA^{Trp}$)[32] were developed. However, until now mainly the orthogonal pairs based on *M. jannaschii* TyrRS:$tRNA_{CUA}^{Tyr}$ have been used frequently for the incorporation of mainly Tyr and Phe analogues or surrogates with extended aromatic groups.[10,33]

In 2008 and 2009, the research groups of Yokoyama, Chin, and Carell succeeded in the development of an orthogonal pair based on the naturally occurring pyrrolysyl-tRNA synthetase (PylRS).[34–36] In methanogenic, anaerobic Archaea, such as *Methanosarcina mazei*, as well as in *E. coli*, Pyl incorporation takes place as a natural nonsense suppression event without the need for a specific sequence context.[37,38] Chemical manipulation of the aliphatic side chain of the Pyl structure can lead to numerous analogues with interesting functionalities for incorporation into proteins. Not surprisingly, engineered PylRS:$tRNA^{Pyl}$ pairs were made for the incorporation of versatile lysine analogues into proteins at single and multiple positions of recombinant target proteins.[34–36]

It was expected that both pyrrolysyl- and lysyl-tRNA synthetases should be especially amenable to manipulation of their enzyme specificity owing to their intrinsically high substrate tolerance. The specificity of the aminoacylation reaction (expressed as the discrimination or D factor) with regard to the 20 canonical amino acids varies considerably among different aaRSs in vitro.[39] The highest D factor (between 28000 and $>$ 500000) was found for TyrRS, whereas the lowest values, between 130 and 1700, were observed for LysRS.[40] The existence of "intrinsically relaxed" aaRSs, such as LysRS, is an excellent starting point for the design of new generations of orthogonal aaRSs with considerably improved specificity and catalytic performance.

Recently, Chin and co-workers generated two mutually orthogonal *M. jannaschii* TyrRS:$tRNA_{CUA}^{Tyr}$ pairs capable of suppressing UAG and AGGA.[41] This finding is remarkable, because it shows that evolutionary distance is not necessarily required for the combination of orthogonal pairs without the generation of cross-reactivities. On the basis of this result, it could be possible to make the orthogonal pairs already developed from *M. jannaschii* TyrRS combinable and thus available for multiple NCAA incorporation.

3. Genetic-Code Expansion with Two Different NCAAs

The site-specific incorporation of NCAAs into single recombinant proteins is highly relevant, as specific biological problems could be addressed in this way. As discussed earlier, the incorporation of multiple chemically distinct NCAAs into recombinant target proteins is also highly desirable. In this context, Schultz and co-workers established an orthogonal aaRS:tRNA pair derived from an archaeal LysRS:$tRNA^{Lys}$ pair for frameshift suppression of the quadruplet codon AGGA in vivo.[31] This pair was used with the amber-suppressing system based on *M. jannaschii* TyrRS and enabled the simultaneous incorporation of homoglutamine and *O*-methyltyrosine at different positions within myoglobin. The experiment provided an interesting proof-of-principle; however, the extremely low protein yield made it unattractive for practical applications.

A great improvement of this approach was recently reported by Liu and co-workers.[42] They succeeded in incorporating two chemically distinct NCAAs into green fluorescent protein (GFPuv) by combining two orthogonal pairs in a single expression experiment. However, unlike in the experiment of Schultz and co-workers, a UAG (amber) codon was used in combination with a UAA (ochre) codon. To enable efficient double labeling, they generated two orthogonal pairs without mutual cross-reactivity. A mutated *M. jannaschii* TyrRS for the charging of $tRNA_{CUA}^{Tyr}$ with *p*-azidophenylalanine (*p*-AzPhe) was combined with a modified PylRS, which loads N^{ε}-propargyloxycarbonyl-L-lysine (Pox-Lys) onto a $tRNA_{UAA}^{Pyl}$ species with a mutated anticodon loop (Figure 2). The modified TyrRS:$tRNA_{CUA}^{Tyr}$ pair was used in conjunction with a recently refined plasmid system for the improved expression of labeled proteins.[43] The authors reported that full-length, doubly labeled GFPuv was expressed "in good yields" and was sufficiently stable for conjugation by click chemistry with fluorescence tags containing complementary azide/alkyne functionalities.

It should be kept in mind that an anticodon switch in most tRNAs is tantamount to a modification of one of the three major identity elements (anticodon, discriminator base 73, and N1:N72 pair). For example, in *M. jannaschii* the interaction between the cognate $tRNA^{Tyr}$ anticodon and the anticodon-binding domain of TyrRS is one of the crucial identity determinants.[44] Therefore, it is difficult to manipulate complex tRNA identity elements to change tRNA specificity and retain or gain acceptor activity at the same time. For example, the engineered amber suppressor $tRNA^{Tyr}$ had about 300 times lower acceptor activity when compared to wild-type $tRNA^{Tyr}$.[45,46] As a result, the translation efficiency of the mRNAs with in-frame stop codons is dramatically decreased. To partially restore the amino acid loading, tRNA engineering[32,47] and adaptation of the aaRS anticodon-binding pocket were necessary.[46] *M. mazei* PylRS, however, seems to exhibit much weaker interactions with the anticodon loop of $tRNA^{Pyl}$. This feature might be extremely useful for anticodon manipulations (Figure 2). In the recent study, Liu and co-workers[42] exploited this feature and found that opal, ochre, and even quadruplet UAGA are very

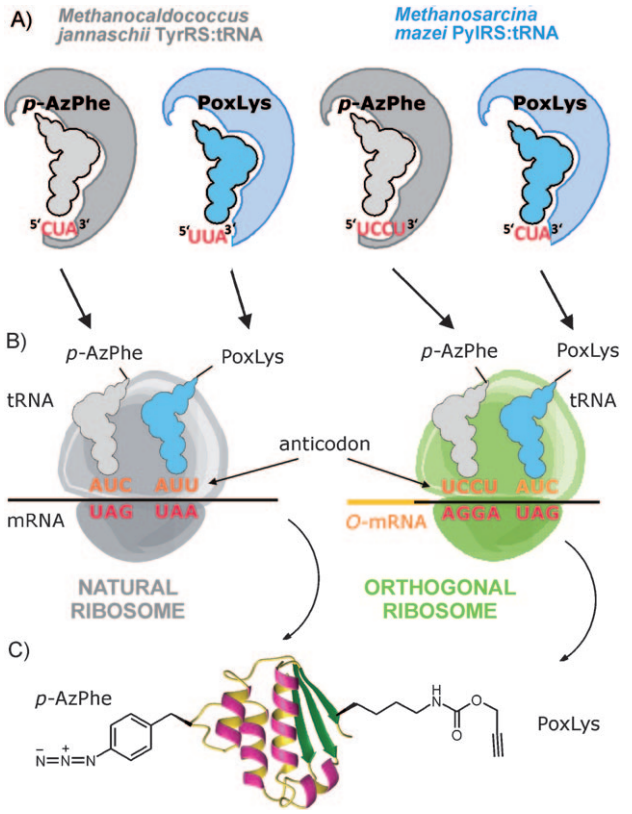

Figure 2. Two approaches for the double incorporation of NCAAs into recombinant proteins on the basis of read-through methods. A) Mutated TyrRS and PylRS were used to load *p*-AzPhe and PoxLys onto cognate suppressor tRNAs with different anticodons. B) Liu and co-workers[42] used natural ribosomes to translate GFPuv mRNA containing in-frame amber (UAG) and ochre (UAA) stop codons. Chin and co-workers[50] used an orthogonal ribosome that is thought to be more efficient at surpressing amber codons and quadruplet codons for the expression of doubly labeled recombinant GST-calmodulin. C) Both research groups could analytically demonstrate NCAA incorporation and the chemical reactivity of the delivered functional groups (terminal azides and alkynes). The efficiency of both approaches should be analyzed in one integrated experiment, as described in the text. The proof-of-concept for the in vivo site-specific double incorporation of NCAA by Schultz and co-workers[31] was not incorporated in the figure (to enable a clearer arrangement). Note that the orthogonal PllRS:tRNAPyl pair shows remarkably high natural tolerance for different anticodons: CUA, UUA, UCA, and UCUA. Parts of Figure 2 were kindly provided by Dr. Birgit Wiltschi.

efficiently suppressed by the $tRNA^{Pyl}$ species. Thus, aaRS:tRNA pairs such as the modified PylRS:$tRNA_{CUA}^{Pyl}$ pair, which seems to have a certain plasticity with respect to the anticodon as an identity element, might exhibit a greater efficiency and versatility for the expression of recombinant proteins with single, double, and triple in-frame amber stop codons as well as with a combination of amber and ochre codons.

When Liu and co-workers attempted to use a mutant of *M. jannaschii* TyrRS capable of recognizing *O*-sulfo-L-tyrosine instead of *p*-AzPhe in combination with *M. mazei* PylRS:$tRNA_{UAA}^{Pyl}$ for PoxLys incorporation, poor yields of the doubly labeled GFPuv were observed.[42] These findings highlight an important issue that has not been critically evaluated with due care in the literature to date. Namely, it is evident that a number of reported *M. jannaschii* derived aaRS:$tRNA_{CUA}^{Tyr}$ pairs are characterized by poor catalytic performance, which urgently requires improvement.[43] In general, the suppression efficiency does not exceed 50% and is highly dependent on the position of the stop codon within the gene sequence (context effects).[48] There are positions in gene sequences which are difficult or even impossible to read-through with suppressor tRNAs.[49] Therefore, it is necessary to first search for sites suitable for suppression in the gene sequence of interest. Finally, the efficiency of the methodology is highly dependent on the protein used. "Model proteins", such as dihydrofolate reductase, glutathione S-transferase, myoglobin, luciferase, β-lactamase, lysozyme, or GFP, are doubtlessly well-suited for this labeling technology. However, it remains to be determined to which extent "difficult" proteins, such as single-chain antibodies or proteins with repetitive sequences, such as collagens, are amenable for such experiments.

The discussed problems of stop-codon-suppression systems should also be considered in the light of the dynamics and kinetics of ribosomal translation with misacylated tRNAs. A recent study of Cornish and co-workers[51] demonstrates that the translational machinery does not distinguish between correctly charged and misacetylated tRNAs. However, small but reproducible differences in the dynamics of the ribosome cycles with misacetylated tRNAs were observed. This observation might indicate that each misacylation slightly slows down the related translation cycle. If this hypothesis is true, the design of methodologies for the incorporation of three or more NCAAs into proteins should take into account these issues as well. Thus, a lot of optimization work can be anticipated. For example, it was recently demonstrated that the suppression-based incorporation of single NCAAs can be further improved by the use of mutated ribosomal proteins,[52] elongation factors,[53] and even mutated ribosomal RNAs.[54] Furthermore, the optimization of the interaction between the elongation factor Tu and NCAA-tRNA bears a lot of potential for suppression enhancement.[55–57] A complete discussion of these and related issues (e.g. amino acid uptake,[58] protein folding,[19] tRNA modifications,[59] wobble-free tRNA design,[60] mutagenesis of the anticodon-binding domain,[61] and ribosomal kinetic proofreading[51]) is out of the scope of this Minireview.

4. Orthogonal Ribosomes—A Critical View

The basic weakness of incorporation methods based on stop-codon suppression is that suppressor tRNAs have to compete with release factors (RFs) at the ribosomal A site (where RFs catalyze the termination reaction). This competition lowers the overall yield of labeled full-length protein significantly. Higher suppression efficiency would be especially desirable when several in-frame stop codons need to be suppressed. However, a reduction in the release-factor activity in vivo can lead to enhanced cellular toxicity.[62] Chin and co-workers succeeded in increasing the efficiency of

NCAA incorporation by using an orthogonal ribosome mutated for the improved read-through of termination signals in mRNAs.[54] In brief, a library of 16S rRNAs was constructed with mutations in the ribosomal A site. In the next step, the library was screened for mutant ribosomes which retained the translational fidelity of native ribosomes but exhibited a substantial increase in read-through efficiency of in-frame amber stop codons. This increased read-through of UAG codons most probably results from a decreased affinity of the ribosomal A site for release factor 1 (whereas release factors 2 and 3 recognize the other two stop codons, UAG and UAA). Furthermore, these ribosomes are specialized to translate only target mRNAs, which carry mutations in the Shine–Dalgarno (SD) region; thus, through modified SD/anti-SD interactions, target mRNA is exclusively translated. In this way, the synthesis of the proteome is performed by natural ribosomes, whereas mutated ribosomes are used for the expression of recombinant target proteins (Figure 2).

In their most recent study, Chin and co-workers[50] revisited the earlier ideas and in vivo experiments of the Schultz research group on combinations of triplet and quadruplet codons for the double incorporation of NCAAs.[31] By using the previously developed screening systems, they generated a mutant ribosome (named "riboQ1") capable of the efficient decoding of quadruplet codons, which are normally less efficiently decoded by the natural ribosome. At the same time, the mutant ribosome maintained the enhanced in-frame amber-decoding capacity. To check whether riboQ1 is capable of performing efficient double labeling, Chin and co-workers generated two orthogonal pairs without mutual cross-reactivity. However, unlike Liu and co-workers, they used the *p*-AzPhe-loading mutant of *M. jannaschii* TyrRS with a $tRNA_{UCCU}^{Tyr}$ species (instead of $tRNA_{CUA}^{Tyr}$) in combination with a mutated *M. mazei* PylRS, which loads PoxLys onto its cognate $tRNA_{CUA}^{Pyl}$. The model protein GST-calmodulin, with AGGA at position 1 and UAG at position 40, was fully translated, although the authors did not provide information about specific protein yields. The labeled protein was subjected to click chemistry in the form of an intramolecular azide/alkyne Cu^{+}-catalyzed cycloaddition to demonstrate that such a reaction is generally possible in GST-calmodulin.

Liu and co-workers[42] found that quadruplet anticodons are well-tolerated by natural ribosomes when introduced into *M. mazei* $tRNA^{Pyl}$, whereas Chin and co-workers[50] reported that only riboQ1 was efficient for the reading of quadruplets introduced into $tRNA^{Tyr}$ derived from *M. jannaschii* (Figure 2). However, the efficiency of riboQ1 can only be realistically estimated if both combinations of orthogonal pairs are tested in the presence of normal and mutated ribosomes. Such a comparison should take into account that $tRNA_{CUA}^{Pyl}$ seems much more amenable to anticodon manipulations than $tRNA_{CUA}^{Tyr}$. An additional critical point, as recently pointed out by Suga and co-workers,[63] is the general problem of the efficiency of riboQ1 in frameshift-suppression reactions. A careful inspection of the supplementary information of the original manuscript[50] indicates the presence of a significant amount of truncated protein as a by-product of the reprogrammed translation.

At this stage of development, the design of orthogonal (mutated) ribosomes undoubtedly represents an interesting proof-of-principle for the reassignment of nontriplet codons. This approach might even lead to interesting theoretical considerations ranging from the design of "alternative protein universes" with a quadruplet code to fundamental questions regarding the origin of life.[64] However, reprogrammed protein translation without ribosome mutations could be an equally good basis for such theoretical studies.[2,65] In the meantime, it remains to be determined whether reprogrammed protein translation with mutated ribosomes can be extended to eukaryotic (yeast or mammalian) cells and especially to industrial microbial strains designed for applications in biotechnology.

5. Combination of Code Engineering and Expansion for the Incorporation of Multiple Amino Acids

Whereas methodologies based on expansion of the genetic code consider termination triplets or quadruplets as blank codons for NCAA incorporation, genetic-code engineering is based on sense-codon reassignment by exploiting the substrate tolerance of endogenous aaRSs. With this method, we recently demonstrated the simultaneous in vivo incorporation of three different synthetic amino acids in a single expression experiment with polyauxotrophic *E. coli* strains. On the one hand, we explored the possibility of generating a "Teflon" protein with novel properties by introducing fluorinated amino acids at up to 24 positions in the thermostable lipase from *Thermoanaerobacter thermohydrosulfuricus*.[18] On the other hand, we tailored barstar from *Bacillus amyloliquefaciens* with a reactive handle for click chemistry (homopropagylglycine), a nonperturbing fluorescent tag (4-azatryptophan), and *cis*-4-fluoroproline, which had been reported to have a stabilizing effect on barstar. The simultaneous and efficient incorporation of the three different NCAAs yielded a fluorescence-detectable and stabilized barstar derivative that can undergo click-chemistry reactions.[17]

To further expand the possibilities of protein tailoring, we recently combined code engineering and code expansion.[49] Thereby, the orthogonal pair enabled us to add NCAAs at permissive sites in a position-specific manner, whereas the use of the auxotrophy-based method enabled the multiple residue-specific incorporation of isostructural NCAAs in response to sense codons. It is indeed important to combine the advantages of both methodological approaches, since, for example, the currently available aaRS mutants for orthogonal pairs are in principle not capable of generating enzymes that discriminate efficiently between subtle structural variations (small functional groups and atoms such as -H/-F, -H/-CH_3, -S-/-CH_2-, -CH=/-N=, or -H/-OH). In general, the careful combination of the possibilities of both approaches, including the possible use of orthogonal ribosomes as well as genome-engineered cells (see Section 6), will enable not only an efficient addition/substitution of NCAAs into the standard amino acid repertoire prescribed by the genetic code, but will

also offer protein-engineering toolkits with unprecedented possibilities.

6. Organismal Chemistry (Synthetic Cells)

Genetic-code expansion assumes the generation of novel coding units in the genetic code and the subsequent expansion of the amino acid repertoire in living cells. One day, this approach should yield semisynthetic living cells with inheritable and specific chemical alterations in the proteome and emergent properties/behaviors not found in the natural biological realm.[2] However, due to the low cellular tolerance of chemically altered proteomes, to date the introduction of novel amino acids has generally only been possible at the level of single recombinant proteins. Thus, all experimental efforts to change or expand coding capacities are still only useful extensions of recombinant DNA technology. In this context, the double and triple incorporation of NCAAs are certainly among the most advanced tools for single-protein engineering.

However, with the recent studies of Venter and coworkers,[66] a realistic prospect for the proteome-wide usage of such methodologies can be envisaged. Venter and coworkers reported a series of experiments on the synthesis and transplantation of a whole bacterial genome with the goal of designing a "minimal cell".[67] The described techniques make it conceivable that bacterial strains could be produced with termination codons or related aaRS/tRNA genes deleted from the whole genome (e.g. from *Mycoplasma*, which naturally possesses an alternative version of the genetic code[4]). Novel orthogonal pairs or even nontriplet coding units with novel assignments (i.e. codon capture) could then be reintroduced into the engineered genome. Such an extension to the amino acid repertoire could perhaps even lead to an evolutionary advantage for engineered cells. Thus, we would have a fairly good chance to generate entirely new organism functions through the systematic proteome-wide introduction of NCAAs. In this way, genetic-code engineering and expansion would deal with the chemical composition of whole proteomes instead of single proteins. Such organismal chemistry would render unprecedented benefits and opportunities in the coming age of synthetic biotechnology.

Addendum

The quality of orthogonal pairs: Tippmann and co-workers very recently provided solid genetic and structural evidence for the lack of capacity of orthogonal pairs evolved to specifically recognize—and therefore translate with high fidelity—certain NCAAs.[68] They clearly demonstrated that certain "gate-keeping" residues in the related aaRSs are of high importance for the design of enzymes with higher preference toward NCAAs than canonical amino acids. Their study will facilitate the refinement of existing and design of new orthogonal pairs in the future.

Organismal chemistry: Very recently, Sakamoto and coworkers succeeded in engineering an *E. coli* strain capable of no longer using the amber (UAG) stop codon for translation termination.[69] This was achieved by eliminating the UAG-recognizing release factor 1, in combination with only a few genetic modifications and complementations. It was not necessary to exchange all the amber stop codons in the whole genome. With these engineered cells, the group was able to incorporate *p*-iodophenylalanine in response to six in-frame amber codons in a model protein. If these results are reproducible by other research groups, the study should be regarded as one of the most important advances in the whole field in the last few years, since the "competition problem" between the amber suppressor-tRNA and release factor 1 is totally eliminated.

Received: September 10, 2010
Published online: February 25, 2011

[1] K. E. Neet, D. E. Koshland, *Proc. Natl. Acad. Sci. USA* **1966**, *56*, 1606 – 1611.
[2] N. Budisa, *Engineering the Genetic Code*, Wiley-VCH, Weinheim, **2005**.
[3] A. Ambrogelly, S. Palioura, D. Soll, *Nat. Chem. Biol.* **2007**, *3*, 29 – 35.
[4] S. Osawa, T. H. Jukes, K. Watanabe, A. Muto, *Microbiol. Rev.* **1992**, *56*, 229 – 264.
[5] F. Zinoni, A. Birkmann, W. Leinfelder, A. Bock, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 3156 – 3160.
[6] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459 – 1462.
[7] R. F. Gesteland, J. F. Atkins, *Annu. Rev. Biochem.* **1996**, *65*, 741 – 768.
[8] N. Budisa, *Angew. Chem.* **2004**, *116*, 6586 – 6624; *Angew. Chem. Int. Ed.* **2004**, *43*, 6426 – 6463.
[9] C. Minks, S. Alefelder, L. Moroder, R. Huber, N. Budisa, *Tetrahedron* **2000**, *56*, 9431 – 9442.
[10] L. Wang, P. G. Schultz, *Angew. Chem.* **2005**, *117*, 34 – 68; *Angew. Chem. Int. Ed.* **2005**, *44*, 34 – 66.
[11] S. Lepthien, M. G. Hoesl, L. Merkel, N. Budisa, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16095 – 16100.
[12] C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber, L. Moroder, *Angew. Chem.* **2001**, *113*, 949 – 951; *Angew. Chem. Int. Ed.* **2001**, *40*, 923 – 925.
[13] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19 – 24.
[14] C. Wolschner, A. Giese, H. A. Kretzschmar, R. Huber, L. Moroder, N. Budisa, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7756 – 7761.
[15] M. G. Hoesl, C. G. Acevedo-Rocha, S. Nehring, M. Royter, C. Wolschner, B. Wiltschi, N. Budisa, G. Antranikian, *ChemCatChem* **2011**, *3*, 213 – 221.
[16] T. Steiner, P. Hess, J. H. Bae, B. Wiltschi, L. Moroder, N. Budisa, *PLoS ONE* **2008**, *3*, No. e1680.
[17] S. Lepthien, L. Merkel, N. Budisa, *Angew. Chem.* **2010**, *122*, 5576 – 5581; *Angew. Chem. Int. Ed.* **2010**, *49*, 5446 – 5450.
[18] L. Merkel, M. Schauer, G. Antranikian, N. Budisa, *ChemBioChem* **2010**, *11*, 1505 – 1507.
[19] L. Moroder, N. Budisa, *ChemPhysChem* **2010**, *11*, 1181 – 1187.
[20] M. G. Hoesl, M. Larregola, H. Cui, N. Budisa, *J. Pept. Sci.* **2010**, *16*, 589 – 595.
[21] S. Brenner, A. O. Stretton, S. Kaplan, *Nature* **1965**, *206*, 994-998.
[22] T. Hohsaka, M. Sisido, *Curr. Opin. Chem. Biol.* **2002**, *6*, 809 – 815.
[23] A. C. Forster, Z. P. Tan, M. N. L. Nalam, H. N. Lin, H. Qu, V. W. Cornish, S. C. Blacklow, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6353 – 6357.

[24] J. C. Anderson, T. J. Magliery, P. G. Schultz, *Chem. Biol.* **2002**, *9*, 237 – 244.
[25] Z. P. Tan, S. C. Blacklow, V. W. Cornish, A. C. Forster, *Methods* **2005**, *36*, 279 – 290.
[26] D. R. Liu, T. J. Magliery, M. Pasternak, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10092 – 10097.
[27] R. Furter, *Protein Sci.* **1998**, *7*, 419 – 426.
[28] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498 – 500.
[29] J. C. Anderson, P. G. Schultz, *Biochemistry* **2003**, *42*, 9598 – 9608.
[30] S. W. Santoro, J. C. Anderson, V. Lakshman, P. G. Schultz, *Nucleic Acids Res.* **2003**, *31*, 6700 – 6709.
[31] J. C. Anderson, N. Wu, S. W. Santoro, V. Lakshman, D. S. King, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7566 – 7571.
[32] R. A. Hughes, A. D. Ellington, *Nucl. Acids Res.* **2010**, *38*, 6813 – 6830.
[33] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413 – 444.
[34] E. Kaya, K. Gutsmiedl, M. Vrabel, M. Müller, P. Thumbs, T. Carell, *ChemBioChem* **2009**, *10*, 2858 – 2861.
[35] T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818 – 822.
[36] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232 – 234.
[37] O. Namy, Y. Zhou, S. Gundllapalli, C. R. Polycarpo, A. Denise, J. P. Rousset, D. Soll, A. Ambrogelly, *FEBS Lett.* **2007**, *581*, 5282 – 5288.
[38] J. Yuan, P. O'Donoghue, A. Ambrogelly, S. Gundllapalli, R. L. Sherrer, S. Palioura, M. Simonovic, D. Soell, *FEBS Lett.* **2010**, *584*, 342 – 349.
[39] H. Jakubowski, E. Goldman, *Microbiol. Rev.* **1992**, *56*, 412 – 429.
[40] W. Freist, D. H. Gauss, *Biol. Chem. Hoppe Seyler* **1995**, *376*, 451 – 472.
[41] H. Neumann, A. L. Slusarczyk, J. W. Chin, *J. Am. Chem. Soc.* **2010**, *132*, 2142 – 2144.
[42] W. Wan, Y. Huang, Z. Y. Wang, W. K. Russell, P. J. Pai, D. H. Russell, W. R. Liu, *Angew. Chem.* **2010**, *122*, 3279 – 3282; *Angew. Chem. Int. Ed.* **2010**, *49*, 3211 – 3214.
[43] T. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, *J. Mol. Biol.* **2010**, *395*, 361 – 374.
[44] P. Fechter, J. Rudinger-Thirion, M. Tukalo, R. Giege, *Eur. J. Biochem.* **2001**, *268*, 761 – 767.
[45] R. Giege, *Nat. Struct. Biol.* **2003**, *10*, 414 – 416.
[46] T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto, S. Yokoyama, *Nat. Struct. Biol.* **2003**, *10*, 425 – 432.
[47] L. Wang, P. G. Schultz, *Chem. Biol.* **2001**, *8*, 883 – 890.
[48] A. K. Kowal, C. Kohrer, U. L. RajBhandary, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2268 – 2273.
[49] M. G. Hoesl, N. Budisa, *ChemBioChem* **2010**, DOI: 10.1002/cbic.201000586.
[50] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, *Nature* **2010**, *464*, 441 – 444.
[51] P. R. Effraim, J. Wang, M. T. Englander, J. Avins, T. S. Leyh, R. L. Gonzalez, Jr., V. W. Cornish, *Nat. Chem. Biol.* **2009**, *5*, 947 – 953.
[52] Y. Huang, W. K. Russell, W. Wan, P.-J. Pai, D. H. Russell, W. Liu, *Mol. BioSyst.* **2010**, *6*, 683 – 686.
[53] Y. Doi, T. Ohtsuki, Y. Shimizu, T. Ueda, M. Sisido, *J. Am. Chem. Soc.* **2007**, *129*, 14458 – 14462.
[54] K. H. Wang, H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Biotechnol.* **2007**, *25*, 770 – 777.
[55] F. J. LaRiviere, A. D. Wolfson, O. C. Uhlenbeck, *Science* **2001**, *294*, 165 – 168.
[56] A. C. Forster, *Nucleic Acids Res.* **2009**, *37*, 3747 – 3755.
[57] J. Guo, E. M. Charles III, H. S. Lee, D. Groff, P. G. Schultz, *Angew. Chem.* **2009**, *121*, 9312 – 9315; *Angew. Chem. Int. Ed.* **2009**, *48*, 9148 – 9151.
[58] C. Giese, S. Lepthien, L. Metzner, M. Brandsch, N. Budisa, H. Lilie, *ChemMedChem* **2008**, *3*, 1449 – 1456.
[59] A. Weixlbaumer, F. V. Murphy, A. Dziergowska, A. Malkiewicz, F. A. P. Vendeix, P. F. Agris, V. Ramakrishnan, *Nat. Struct. Mol. Biol.* **2007**, *14*, 498 – 502.
[60] I. Kwon, K. Kirshenbaum, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 7512 – 7513.
[61] J. K. Takimoto, K. L. Adams, Z. Xiang, L. Wang, *Mol. BioSyst.* **2009**, *5*, 931 – 934.
[62] R. E. Doerig, B. Suter, M. Gray, E. Kubli, *EMBO J.* **1988**, *7*, 2579 – 2584.
[63] G. Hayashi, Y. Goto, H. Suga, *Chem. Biol.* **2010**, *17*, 320 – 321.
[64] I. A. Chen, M. Schindlinger, *Bioessays* **2010**, *32*, 650 – 654.
[65] N. Budisa, L. Moroder, R. Huber, *Cell. Mol. Life Sci.* **1999**, *55*, 1626 – 1635.
[66] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, H. O. Smith, J. C. Venter, *Science* **2010**, *329*, 52 – 56.
[67] A. C. Forster, G. M. Church, *Mol. Syst. Biol.* **2006**, *2*, 45.
[68] A. K. Antonczak, Z. Simova, I. T. Yonemoto, M. Bochtler, A. Piasecka, H. Czapińska, A. Brancale, E. M. Tippmann, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, DOI: 10.1073/pnas.1012276108.
[69] T. Mukai, A. Hayashi, F. Iraha, A. Sato, K. Ohtake, S. Yokoyama, K. Sakamoto, *Nucl. Acids Res.* **2010**, *38*, 8188 – 8195.

