Beilstein-Institut

Bozen 2006, May 15th – 19th, 2006, Bozen, Italy

Basic Requirements for Reprogramming Intracellular Protein Translation

Birgit Wiltschi and Nediljko Budisa

Max Planck Institute of Biochemistry, BioFuture Independent Research Group Molecular Biotechnology, Am Klopferspitz 18, 82152 Martinsried, Germany

E-Mail: * budisa@biochem.mpg.de

Received: 7th December 2006 / Published: 5th November 2007

Abstract

Noncanonical amino acids as building blocks for the biosynthesis of tailor-made proteins represent a nearly infinite supply for the introduction of unusual functions, molecular scaffolds exerting conformational constraints or pharmacologically active entities into proteins. Exploitation of this supply for biotechnological or medical application is owed to the flexibility of the cellular systems involved in the incorporation of noncanonical amino acids into proteins. The broad substrate specificity of cellular amino acid transport systems allows for transmembrane passage of many noncanonical amino acid analogues. Subsequently, the intracellular amount of noncanonical amino acids can be tuned to levels high enough for efficient activation and tRNA charging by aminoacyl-tRNA synthetases (AARS). As a variety of noncanonical amino acids with different chemical properties are incorporated into polypeptide sequences, they are obviously metabolically stable. The indiscriminateness of the AARS towards many noncanonical substrates, i.e. their "catalytic promiscuity", is the central principle for expanding the scope of ribosomal protein synthesis. tRNAs charged with noncanonical amino acids mediate their efficient translation into nascent polypeptide chains by codon reassignment owing to the adaptability of the ribosome. Our contribution specifically highlights all theses principles as *sine qua non* for protein translation reprogramming with an expanded genetic code.

INTRODUCTION

In all living organisms, the genetic information is implemented by three processes, two of which – transcription and translation – are intimately linked (Fig. 1). The genetic information stored in the DNA molecule is passed on to progeny by replication, i.e. duplication of the DNA. It is transcribed to messenger RNA (mRNA), which is often processed, e.g., the mRNA may be edited, spliced or chemically modified and a 5'-cap and 3'-polyA tail may be added to produce mature mRNA. Finally, the genetic information now inherent to the mRNA is translated into the corresponding amino acid sequence of the protein at the ribosome [1]. The precision of the information transfer from nucleic acids to proteins is governed by a set of rules, the genetic code, that strictly links a linear tri-nucleotide sequence to a linear sequence of amino acids. The genetic code operates with tri-nucleotide units (codons) each of which codes for a single amino acid. There are 64 possible three-base combinations of the four nucleobases, enough to encode 64 amino acids. However, all life forms on Earth use only a standard set of 20 canonical amino acids to biosynthesize their proteins. Genes involved in replication, transcription and translation are almost universally conserved among all life kingdoms [2].

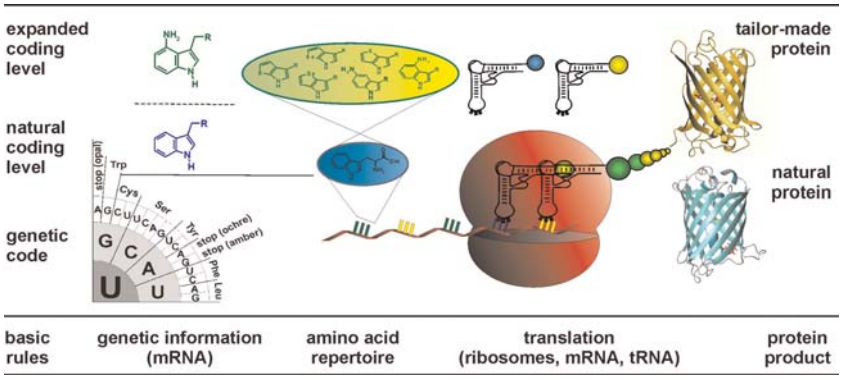

Figure **1.** Expanding the scope of protein biosynthesis. The genetic information stored in DNA is transcribed to mRNA and translated into a polypeptide sequence at the ribosome. The transfer RNA (tRNA) molecule acts as the "adaptor" that transfers the genetic information from the nucleic acid (mRNA) to the polypeptide. A set of rules, the genetic code, strictly links a linear nucleotide sequence to a linear sequence of amino acids. At the standard coding level, only 20 canonical amino acids are incorporated into the polypeptide chains of natural proteins. The flexibility of the protein translation machinery allows the incorporation of noncanonical amino acids into proteins at the expanded coding level. In this way, tailor-made proteins with novel, unnatural properties can be biosynthesized.

Once translation is accomplished, the newly synthesized polypeptide chains usually undergo further processing, i.e., they are post-translationally modified before the mature and active proteins can act as structural components of the cell or serve as functional biochem-

ical machines. Numerous nonstandard amino acid residues are known but they are all introduced into proteins by enzymatic modification after translation [3] or by non-ribosomal peptide synthesis [4].

Amino acids are provided for protein synthesis in the form of aminoacyl-tRNAs. The transfer RNA (tRNA) molecule acts as the "adaptor" that transfers the genetic information from the nucleic acid (mRNA) to the polypeptide (Fig. 1). The protein translation process encompasses two distinct molecular recognition events: the interaction of the aminoacyl-tRNA anticodon with the correct codon in the mRNA on the ribosome and the specific pairing of the tRNA anticodon with its pertinent amino acid by action of the aminoacyl-tRNA synthetase (AARS). Thus, the specific covalent attachment of an amino acid to its cognate tRNA (the aminoacylation identity), which interprets the genetic code, is controlled by the substrate specificity and selectivity of the AARS. That means that the quality of the translational process depends largely on the precision of aminoacylation and subsequent non-covalent interaction of the charged tRNA with the ribosome-bound mRNA (Fig. 1).

The accuracy of the translation machinery is malleable with respect to catalytic promiscuity of the AARS and ribosome plasticity. Deviations from the rules of the genetic code allow protein translation to occur at an expanded level of genetic coding (Fig. 1) which opens up entirely new prospects for the biosynthesis of "unnatural" proteins. It is possible to incorporate non-proteinogenic, so-called noncanonical, amino acids into proteins [5, 6] that furnish the resulting alloproteins with novel properties. The unusual structural, chemical and functional qualities of such "tailored-made" proteins should be of great value for biotechnological or medical applications.

For successful *in vivo* protein biosynthesis with noncanonical amino acids the following conditions must be fulfilled: (1) Effective uptake/import of the noncanonical amino acid into the cell; (2) metabolic stability of the noncanonical amino acid in the cell; (3) its intracellular accumulation at levels high enough for efficient activation and tRNA acylation by AARS; and (4) translation of the noncanonical amino acid into nascent polypeptide chains in response to a sense or stop codon (codon reassignment and nonsense suppression, respectively). (5) Correct folding of the protein must not be obviated by incorporation of the noncanonical amino acid. The manipulation of the translation process involves diverse cellular functions, nonetheless, AARS and tRNAs represent the main targets for protein biosynthesis with an expanded amino acid repertoire [7].

Cellular Uptake of Noncanonical Amino Acids

Knowledge on the import of noncanonical amino acids into cells and their metabolic fate is scarce. Since a variety of amino acid analogues can be successfully incorporated into prokaryotic and eukaryotic proteins [8] the uptake is expected to occur efficiently. Moreover, the analogues appear to be metabolically stable such that they accumulate in quantities that promote their turn over by AARS.

During evolution, the organisms developed idiosyncratic ways of handling amino acids. In spite of that, amino acid biosynthesis pathways, intracellular amino acid transport, turnover and discharge mechanisms are conserved throughout the different kingdoms of life [9]. Apart from their central function in protein biosynthesis, amino acids play fundamental roles in many different processes including hormone metabolism, nerve transmission, cell growth, production of metabolic energy, nucleobase synthesis, nitrogen metabolism and urea biosynthesis [10]. For the efficient transport of these important substances from the environment into and between the cells, accordingly versatile amino acid transport systems have evolved that differ greatly in their substrate spectrum, transport mechanism and tissue specificity. In most cases, transporters recognize a broad spectrum of amino acids and their derivatives [10]. Such a lack of transport specificity can have adverse effects since many different kinds of small molecules penetrate the cell and form intracellular pools [9]. Once such substances are present in the cytosol, they may affect a variety of metabolic, synthetic, and other cellular functions. Particularly noncanonical amino acid analogues can (i) affect amino acid biosynthesis, (ii) interact with catabolic enzymes, (iii) act as (irreversible) enzyme inhibitors, (iv) interfere with amino acid transport and storage and (v) enter the translation process by serving as substrates for protein synthesis, i.e. they enter the genetic code [11, 12].

In the yeast *Saccharomyces cerevisiae*, most of the permeases mediating amino acid uptake are specific for one or a few related L-amino acids. Nevertheless, *S. cerevisiae* possess a broad-specificity, large-capacity, general amino acid permease that catalyses the uptake of most L- and D-amino acids, non-proteinogenic amino acids such as citrulline and ornithine and a number of toxic amino acid analogues and amino acid biosynthesis inhibitors [13]. “System L” transporters in mammalian cells are unique for their broad substrate selectivity. Their multispecific properties enable them to accept not only naturally occurring amino acids for transport but also amino-acid related, potentially toxic compounds [14, 15]. Some amino acid requiring mutants of *Escherichia coli* [16, 17] and *S. typhimurium* [18] can satisfy their auxotrophic demands by utilizing the corresponding D-amino acids for growth. Normal *E. coli* strains possess an efficient transport system for D-methionine [19, 20]. However, since all optically active amino acids found in proteins are of the L-configuration, D-methionine undergoes enzymatic conversion to the L-isomer [21] rather than being directly incorporated into proteins. A similar conversion is expected also to occur to the other D-amino acids.

Taken together, all evidence points to an amazing flexibility of the prokaryotic and eukaryotic amino acid uptake systems with regard to their substrate spectrum. The efficient uptake of noncanonical amino acids by these systems appears plausible. However, systematic studies on the cellular import of noncanonical amino acids are rare. Harrison *et al*. investigated the transport of L-4-azaleucine in *E. coli* and found that the aromatic and branched chain amino acid transporters serve for the accumulation of this leucine analogue [22].

Liu and Schultz were the first to systematically assess the cytotoxicity of 138 different noncanonical amino acids and α-hydroxy acids in *E. coli* [23]. Although they did not reveal intracellular levels of their tested amino acid analogues, they observed that substances that were very close analogues of natural amino acids or those that displayed reactive functionalities, demonstrated the highest toxicities [23]. In contrast, less related analogues or noncanonical amino acids with inert side-chains were only moderately toxic or not toxic at all. This observation suggests that the amino acid analogues exert their toxicity by incorporation into proteins, leading to aberrant protein structure and/or function. Similarly, chemical interaction of the reactive analogues with essential enzymes might cause the observed toxic effects. Thus, toxicity is not necessarily inherent to noncanonical amino acids, but they can interfere with the cellular metabolism and their toxicity be elicited only after their metabolic conversion [12].

In order to identify possible cellular import pathways for noncanonical amino acids, Liu and Schultz repeated the uptake assay with extremely toxic amino acid analogues in the presence of excess structurally similar natural amino acids. An excess of natural amino acid could "complement" the effects of 16 different toxins, either by outcompeting for the cellular uptake or for interference with essential cellular functions. Toxins tested in this way appeared to exploit the *E. coli* transporters for Ala, Glu, Lys, Leu, Met, Pro, Gln, Arg, Thr and Tyr. In similar competition assays, non- or moderately toxic Gln and Glu analogues were shown to complement the adverse cellular effects of Gln- or Glu-related toxins, respectively. This observation indicates that the non-toxic amino acid analogues were most probably taken up by the *E. coli* Gln and Glu transport systems, in an analogous manner to the corresponding toxins. These amino acid transporters appeared remarkably tolerant with respect to perturbations in amino acid structure as they accepted amino acid analogues with elongated side chains as well as ketone or methylene moieties at the γ-position and hydrazide and sulfoxide groups [23].

To conclude, there is good evidence that noncanonical amino acids are imported into cells by the same pathways as natural amino acids. Due to the remarkable substrate tolerance of the amino acid transporters, amino acid analogues with a plethora of different structural and/or functional properties can be imported into living cells. However, conclusive data on the intracellular accumulation levels of noncanonical amino acids still remain to be collected.

Unnatural Amino Acid Turnover By AARS: Catalytic Promiscuity

Amino acids are provided for protein synthesis in the form of aminoacyl-tRNAs. The transfer RNA (tRNA) molecule acts as the "adaptor" that transfers the genetic information from the nucleic acid (mRNA) to the polypeptide. Sophisticated proofreading mechanisms control the quality of genetic message transmission during translation [24]. The protein translation process encompasses two distinct molecular recognition events: the interaction of the aminoacyl-tRNA anticodon with the correct codon in the mRNA on the ribosome and the specific pairing of the tRNA anticodon with its pertinent amino acid by action of

the aminoacyl-tRNA synthetase. Thus, the specific covalent attachment of an amino acid to its cognate tRNA, which interprets the genetic code, is controlled by the substrate specificity and selectivity of the AARS. That means that the accuracy of the translational process depends largely on the precision of aminoacylation and subsequent non-covalent interaction of the charged tRNA with the ribosome-bound mRNA.

The selection of amino acids for protein synthesis takes place essentially at the aminoacylation of tRNA under the control of AARS [1]. In this way, the amino acids are associated with the coding triplets and once a tRNA is misacylated with a noncognate amino acid, the latter amino acid will be directly incorporated into proteins [24]. A battery of twenty cellular enzymes, one for each canonical amino acid, operates at the interface between nucleic acids and proteins, utilizing ATP as an energy source and Mg^{2+} ions as cofactors. They are multi-domain enzymes ([25] that fish their pertinent amino acids from a cellular pool and join them with their cognate tRNAs (Fig. 2A). The AARS catalyses a two-step reaction, where the amino acid is first condensed with ATP to form a highly reactive aminoacyl adenylate (activation reaction). In the second step, also known as the transfer reaction, the activated amino acid is transferred to the 3' end of its cognate tRNA [1, 5]. However, recent functional genomics studies in bacterial and archaeal systems have shown that numerous organisms do not use the full complement of 20 canonical AARS to synthesize their aminoacyl-tRNAs [26].

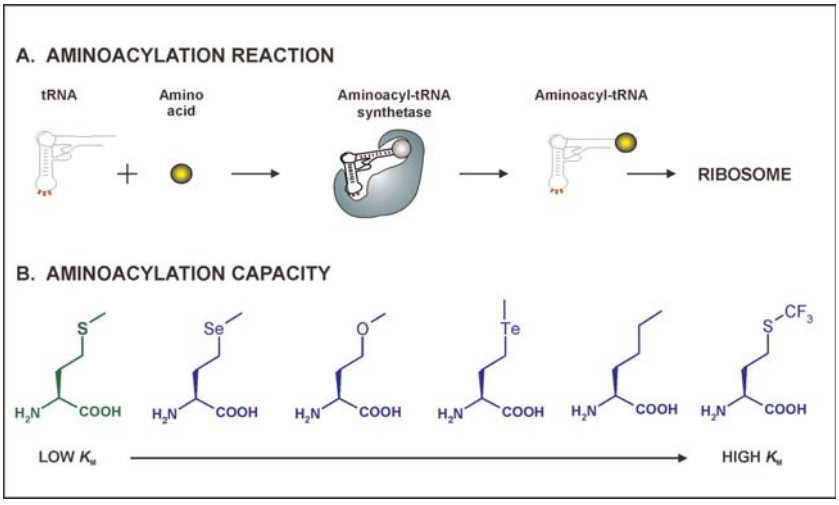

Figure **2.** Aminoacyl-tRNA synthesis. (A) AARS are multi-domain enzymes that fish their pertinent amino acids from a cellular pool and join them with their cognate tRNAs. The AARS must be able to specifically discriminate between different canonical amino acids as well as between canonical and noncanonical ones in the intracellular pool. The aminoacyl-tRNAs can then participate in protein translation at the ribosome. (B) Methionine (green) and its analogues (blue) are substrates for methionyl-tRNA synthetase (MetRS). These analogues are quite bad substrates for MetRS as indicated by their higher K_M and lower k_{cat} values in comparison with the native substrate Met. Nonetheless, MetRS exhibits a remarkable promiscuity in substrate

binding and tRNA charging – a fact extensively exploited in the last decade in order to co-translationally incorporate rather large number of Met analogues and surrogates into protein sequences.

Each AARS must be able to specifically discriminate between different canonical amino acids as well as between canonical and noncanonical ones in the intracellular pool. The overall error rates of translation are in the range of about 2×10^{-3} to 2×10^{-4} in *Escherichia coli* cells [27, 28], while the frequency of errors in the aminoacylation reaction is probably even lower than that. The high accuracy of this process is ensured by editing mechanisms inherent to many AARS [1, 24, 27]. At the level of the aminoacylation reaction the error frequency is influenced by two factors, namely the affinity of the enzyme for different amino acids as substrates (Fig. 2B), and the relative concentration of competing substrates, i.e. amino acids or noncanonical analogues, in the cytosol. AARS discriminate especially restrictively against: i) metabolic precursors and intermediates, ii) structurally and/or chemically similar amino acids (and related compounds) that are relatively abundant in the cytosol, e.g. leucine and valine, which differ only by one methylene group, iii) amino acid analogues that sterically and chemically differ significantly from their canonical counterparts. For example, homocysteine, a metabolic precursor of Met is a substrate for activation by MetRS but is efficiently edited on the basis of size and chemistry [29]. However, isosteric methionine analogues, such as norleucine and SeMet, are accepted for activation by AARS and subsequently loaded onto $tRNA^{Met}$ [30]. Thus, AARS lack absolute substrate specificity and display catalytic promiscuity that allows noncanonical amino acids to enter the translation process. In general, the term catalytic promiscuity describes the capacity of the enzyme to catalyse an adventitious secondary activity at the active site responsible for the primary activity [31]. The persistence of the catalytic promiscuity in modern enzymes in general in and in AARS in particular, can be explained by the fact that such secondary activities usually do not affect the fitness of the organism and therefore there is no selective pressure to eliminate them. This may be, e.g., with respect to the noncanonical amino acids, because the organism does usually not encounter the substrate for the promisicuous reaction.

The lack of absolute substrate specificity has long been known for AARS from many different species [11] and was recognized very early as an invaluable tool to expand the natural amino acid repertoire *in vivo* [32]. Exploitation of the relaxed substrate specificity of wild-type AARS [11,33] and isolation of mutant AARS with expanded substrate specificity [34] or impaired edition functions [35] yielded successful incorporation of more than 100 noncanonical amino acids into proteins during the last decade [8]. Most recently, Szostak and co-workers tested all 20 AARS from *E. coli* for their substrate specificities in an *in vitro* assay [36]. They demonstrated that in addition to the 20 canonical amino acids, 59 previously unknown substances were recognized as substrates by AARS and were charged onto cognate tRNAs.

The naturally occurring catalytic promiscuity of the AARS paved the way to the expansion of the genetic code and an impressive number of noncanonical amino acids have been successfully translated into proteins. As our knowledge of the AARS increases, it is likely that novel mutant enzymes will be developed to increase this number even further.

Ribosome Flexibility

Successful incorporation of noncanonical amino acids with bulky side chains such as glycosyl-, cumarin- or azophenyl-moieties into proteins [8] once more highlights the remarkable substrate tolerance of the cellular amino acid uptake systems and the catalytic promiscuity of the (mutant) AARS. Even more interestingly, it points to an impressive flexibility of the ribosome that must be able to accommodate tRNAs charged with these bulky analogues. However, size is not the only thing that matters since D-analogues of the naturally occurring amino acids are exclusively rejected from protein biosynthesis [37]. Additional mechanisms such as peptide bond formation in the ribosome are expected to discriminate unnatural amino acids from canonical ones [38].

During the translation of the genetic information from an mRNA molecule into a protein, the ribosome accepts the aminoacyl-tRNAs in the form of ternary complexes with elongation factors and GTP. Various RNA and protein components of the ribosome influence the fidelity of the translation process. The selection of the cognate aminoacyl-tRNA in response to a certain codon from a pool containing also noncognate aminoacyl-tRNAs, occurs mainly on the basis of matching codon and anticodon triplets. In fact, discrimination of cognate ternary complexes from noncognate ones is achieved by the ribosome in two consecutive steps, initial selection and subsequent proofreading. This mechanism operates on the basis of both, substrate stabilities and induced fit and noncognate ternary complexes are efficiently excluded from translation in the initial selection step [37].

In the early 1950 s, several investigators showed that ribosome-mediated protein synthesis does not require the integrity of the cell and can continue after cell disruption. Ribosomes in crude cell extracts could be programmed with endogenous DNA templates to synthesize encoded proteins. Such test tube transcription/translation systems proved to be invaluable tools for the elucidation of the molecular mechanisms of protein biosynthesis, its biochemical requirements and the deciphering of the genetic code. Recently, reconstituted translation systems such as PURE have been introduced that consist of highly purified *E. coli* ribosomes and tRNA, plus recombinant translation factors and AARS [39]. These reconstituted systems are capable of translating mRNAs with protein yields similar to those of crude extracts. The essential advantage of reconstituted translation systems over crude extracts is that they allow tight control of the components that are critical for the manipulation of the genetic code, that is, tRNA, AARS, amino acids, and release factors. Hence, reconstituted translation systems can be provided with either all naturally occurring tRNAs along with 20 amino acids or, more importantly, also with enzymatically or chemically charged tRNAs. Chemical aminoacylation of tRNAs with (noncanonical) amino acids is usually carried out using the method developed by Hecht and coworkers [40].The system supplemented with chemically charged tRNAs offers an ideal platform to examine not only

the ribosome's tolerance towards unnatural amino-acid related substances but also the malleability of the genetic code. By omitting certain amino acids and the corresponding tRNAs and AARS, Forster and co-workers were among the first who used a reconstituted translation system to simultaneously reassign three sense codons to noncanonical amino acids [41]. Employing three different tRNAs, chemically charged with noncanonical amino acids, it was possible to synthesize a small peptide that coincidentally displayed the features of the three amino acids analogues.

Efforts on understanding ribosomal selectivity are predominantly focused on the discrimination of cognate from noncognate aminoacyl-tRNAs [37]. As outlined above, experiments on *in vivo* translation of noncanonical amino acids document the remarkable tractability of the ribosome in accepting tRNAs charged with noncanonical amino acids of most diverse structures [8]. Fifteen years ago, Sisido and co-workers studied the adaptability of noncanonical aromatic amino acids to the active centre of ribosome A-site. By using mRNA programmed *E. coli* S 30 extract they identified the determinants that dictate adaptability of amino acid analogues carrying large aromatic groups and their efficient translational incorporation [38]. They postulated that accommodation of analogues with bulky aromatic side chains in the ribosome A site is dependent not only on the size but also on the geometry of the aromatic ring system (Fig. 3A). The authors found that, e. g., 2-naphtylalanine adapted efficiently to the active centre of ribosomal A site and was successfully incorporated into polypeptides, whereas 2-anthrylalanine adapted less efficiently and 9-anthrylalanine was neither adapted nor incorporated (Fig. 3B). Thus, the incorporation efficiency of a certain noncanonical amino acid was directly correlated to its adaptability to the A site of the ribosome.

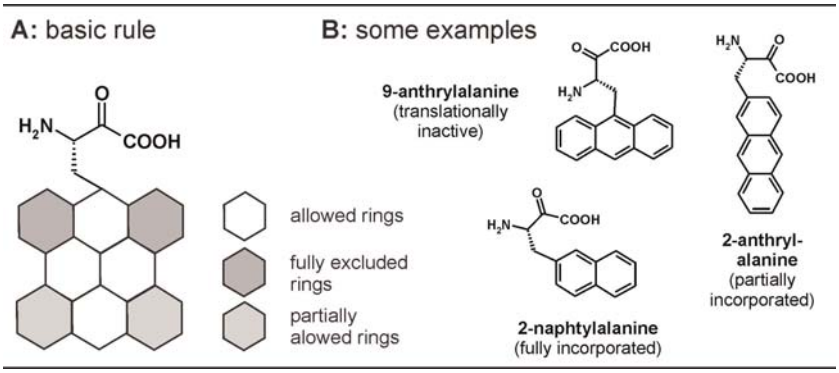

Figure **3.** Ribosomal adaptability of polyaromatic amino acids. (A) Schematic representation of the Sisido-Hohsaka rules for adaptability of different polyaromatic amino acids [53]; (B) Some classic examples to which the Sisido-Hohsaka rules can be applied.

Aromatic amino acids with benzene rings in the grey positions will be rejected by the ribosome A site (e. g. 9-anthrylalanine). In contrast, when the benzene rings occur at the white positions of the structure, the amino acids are accepted by the ribosome and

are incorporated into proteins at higher yields (e.g. 2-naphtylalanine). Structures with partially allowed rings (light grey) are usually translated at very low yields (less than 1%; e.g. 2-anthrylalanine).

Besides size and geometry, chirality of amino acids is also fundamental to ribosomal proofreading. For example, D-amino acids are widely distributed in living organisms and their active utilization in many cell types has been demonstrated [17, 18, 20, 42, 43]. However, all D-amino acids occurring at specific positions in peptides or proteins result from post-translational modification or from non-ribosomal synthesis [4] and never form ribosomal protein translation. It should be kept in mind that erroneous incorporation of D-amino acids in a protein is likely to disrupt secondary structural elements like α-helices. Stringent quality control mechanisms at different levels ensure effective translation of only L-amino acids, namely during charging of the tRNA and complex formation of the aminoacyl-tRNAs with elongation factors, and at the ribosome itself [24]. For example, erroneously formed D-aminoacyl-tRNAs of some amino acids like Tyr are usually subjected to D-amino acid deacylases capable of recycling such D-aminoacyl-tRNA molecules into free tRNA and D-amino acid [44]. Not surprisingly, previous attempts to incorporate D-amino acids into proteins using chemically misacylated tRNAs in cell-free translation systems were unsuccessful. However, it was demonstrated that certain mutations in 23S rRNA induce alterations in the ribosomal peptidyltransferase centre, yielding ribosomes that discriminate less stringently against D-aminoacyl-tRNAs in the ribosomal A-site [45]. This approach was exploited by Hecht and co-workers in order to devise an *in vitro* system for enhanced incorporation of D-amino acids into protein sequences [46].

The substrate tolerance of the ribosome is not restricted to amino acid side chains. Recently, Forster and co-workers successfully synthesized peptidomimetics using a reconstituted transcription/translation system [41]. Peptidomimetics are small peptide-like chains that are designed to mimic peptides. Often, they contain backbone modifications that differ substantially from the peptide amide structure and, thus, can affect local conformational geometries. While β-amino acids are not translationally active [47], analogues with unusual backbones such as α-hydroxy amino acids can be used as substitutes for normal α-amino acids in ribosomal synthesis.

Owing to the ability of the ribosome to accommodate amino acids and amino acid analogues with remarkably diverse structures, geometries and sterical properties (Fig. 4), a number of unnatural amino acids [48] could be successfully incorporated into polypeptides with cell-free protein expression platforms as well as living cells. Among them are non-canonical amino acid analogues with charged, polar, uncharged, aromatic, nonpolar or α,α-disubstituted side chains. Of special interest are fluorescent amino acids with cumarinyl-, anthraniloyl-, benzooxadiazolyl- and dansyl fluorophores - which are usually rather bulky - since they are very sensitive to their microenvironment and are, thus, well suited for the detection, e.g., of ligand binding. Equally important are reactive groups amenable to specific post- translational derivatization. The scope of protein engineering will be further

extended with noncanonical amino acids carrying redox-sensitive moieties, such as nitrophenylalanines and anthraquinonylalanines, spin-probes and photo-switchable or metal chelating residues.

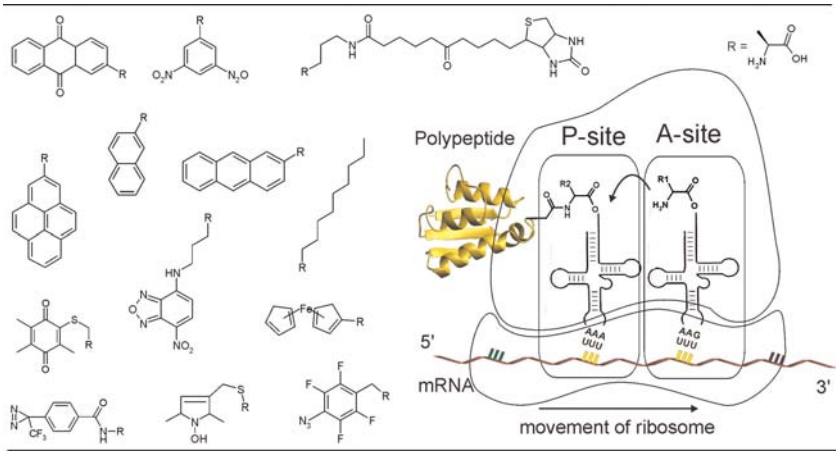

Figure **4.** Ribosomal tolerance towards diverse noncanonical amino acid side chains. This sketch shows only a small part of the chemical diversity that can be accommodated by the ribosome A site and subsequently inserted into target protein sequences. The majority of these substances can only be chemically charged onto tRNA molecules, hence, translation into polypeptide sequences in response to sense, termination or extended codons is achieved only in various *in vitro* expression systems.

Protein Plasticity

The formation of stable (secondary) structure elements in proteins such as α-helices or β-sheets is usually not confined to a defined amino acid sequence but numerous different polypeptide sequences are able to bring forth similar structural features. Many proteins are able to accept mutations with no or only a limited loss of activity and essentially unchanged stability. This quality is termed protein plasticity and most likely results from evolutionary optimization complying with "function defines form". Thus, characteristic folds and stable conformations in proteins have been conserved in spite of their great sequence variations.

Protein plasticity obeys the rules of the standard genetic code. Evolution of the genetic code in living cells limited the number of amino acids to only 20 canonical compounds. The evolutionary conservation of this repertoire is a consequence of at least two facts. First, certain classes of amino acids were excluded from the code through negative selection since they exhibit adverse effects on the folding and structural integrity of target proteins. Second, the exclusion of other amino acid types that do not distort protein structures can be explained by their unavailability at the time when the genetic code appeared in its present day form.

The incorporation of unnatural amino acids into proteins expands the genetic code. However, it may overexpand protein plasticity that was evolutionarily optimized for only 20 canonical amino acids. In fact, the core of globular proteins is probably the most tightly packed form of organic matter in nature that tolerates only little if any structural alteration (Fig. 5). It consists mainly of amino acids with chemically similar or even uniform, hydrophobic side chains ("convergent types"; Fig. 5) leaving few possibilities for substantial structural or functional variations. In fact, overpacking the core with large either canonical or noncanonical side-chains causes a loss of native-like structure [49]. The "ambivalent (or amphipathic) type" rare canonical amino acids Met and Trp along with Cys, Thr, Tyr, Ala and the imino acid Pro are distributed in the protein core, at surfaces and in minicores (Fig. 5). Their noncanonical analogues are expected to be incorporated into proteins in an identical manner. Surface-exposed canonical amino acids are of the "divergent type" because they display vast chemical and sterical diversity (Fig. 5). Consequently, the most reasonable topological context for incorporation of noncanonical amino acids with chemically and sterically divergent side chains into proteins is their surface.

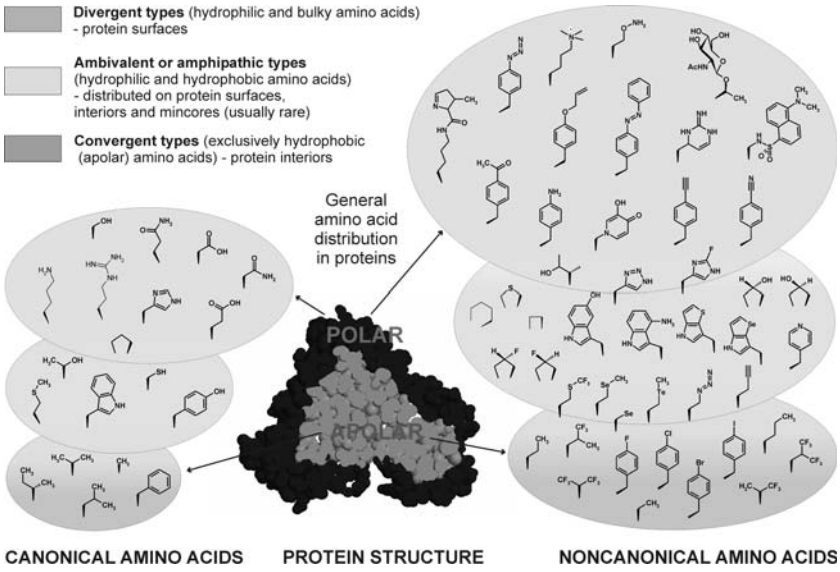

Figure **5.** Distribution of amino acid types and protein topology. The interior of globular proteins is tightly packed and consists mainly of amino acids with chemically similar or even uniform, hydrophobic side chains ("convergent types"). Only few noncanonical fulfil the stringent chemical and sterical criteria for successful incorporation into the protein core. The "ambivalent or amphipathic" amino acid types, the rare canonical amino acids Met and Trp along with Cys, Thr, Tyr, Ala and the imino acid Pro are distributed in the protein core, at the surface and in minicores. Their noncanonical analogues are expected to be incorporated into proteins in an identical manner. Surface-exposed canonical amino acids are often hydrophobic and bulky. They are of the "divergent type" because they display vast chemical and sterical diversity. Noncanonical amino acids with equally divergent side chains can be successfully accommodated at the protein surface.

Depending on the degree of evolutionary optimization of a given amino acid sequence and the plasticity of the structure it forms, translation of noncanonical amino acids into proteins is either neutral or disruptive. In most cases, however, the noncanonically substituted alloproteins are expected to be functionally impaired in comparison to their natural counterparts. There are indeed a few cases where the insertion of noncanonical amino acids into recombinant proteins yielded protein structures capable of accomplishing new functions [50]. One of the most prominent examples includes the design of a golden class of tailor-made autofluorescent proteins. In particular, the incorporation of 4-aminotryptophan into green fluorescence protein (GFP) from *Aequorea victoria* substantially shifts its fluorescence emission to longer wavelengths. Gold fluorescent protein (GdFP) features a unique principle of photophysics that causes a Stokes shift of about 100 nm in its 4-aminotryptophan substituted chromophore [51].

Taken together, protein plasticity often does not leave enough space for systematic amino acid side chain variations. In order to circumvent these difficulties an adaptation of polypeptide sequences for insertion of noncanonical amino acids with novel chemical and/or sterical properties is necessary. This means, some kind of "artificial evolution" of protein plasticity with an expanded genetic code has to be achieved.

Summary and Outlook

For efficient and robust reprogramming of the cellular translation machinery some aspects deserve detailed experimental attention:

As outlined above, the cellular amino acid uptake systems accept an extraordinarily versatile substrate spectrum for transport. According to preliminary noncanonical amino acid uptake experiments performed by Liu and Schultz, the cytotoxicity of noncanonical amino acids varies greatly [23]. Their uptake assay provides useful information on the applicability of a specific noncanonical amino acid for the expansion of the natural amino acid repertoire which would strongly disfavour toxic analogues. However, for efficient incorporation of a noncanonical amino acid into a protein the analogue must accumulate in the cell at high levels, especially because most noncanonical amino acids are bad substrates for AARS [11, 23, 33]. Therefore, it would be of outstanding importance to evaluate the intracellular amounts of at least some noncanonical amino acids by classical uptake assays with radioactively labelled analogues.

The naturally occurring relaxed substrate specificity of AARS is a good starting point for the rational design of mutant enzymes with unusual substrate specificities. These are especially interesting as part of orthogonal tRNA/AARS pairs that are mutually compatible but resistant to challenge by competing interactions with a natural amino acid, and non-cognate AARS or tRNA. Currently available orthogonal pairs allow site specific incorporation of noncanonical amino acids *in vivo* in response to termination codons [52]. Appropriate mutant AARS with tailored substrate specificities should activate and charge even exotic noncanonical amino acids for translation that cannot be incorporated into proteins with the existing techniques.

Ribosomes have proved their ability to accept a vast variety of different noncanonical amino acids. However, protein plasticity can be readily overexpanded by translation of unnatural amino acids into sequences that are evolutionarily optimized for the standard genetic code. Inevitably, it will be necessary to adapt the polypeptide sequences for the insertion of noncanonical amino acids with novel chemical and/or sterical properties.

References

[1] Ibba, M., Söll, D. (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69:**617 – 650.

[2] Osawa, S., Jukes, T.H., Watanabe, K., Muto, A. (1992) Recent evidence for evolution of the genetic code *Microbiol.Rev.* **56:**229 – 264.

[3] Uy, R., Wold, F. (1977) Posttranslational covalent modification of proteins. *Science* **198:**890 – 896.

[4] Finking, R., Marahiel, M.A. (2004) Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **58:**453 – 488.

[5] Hendrickson, T.L., de Crecy-Lagard, V., Schimmel, P. (2004) Incorporation of nonnatural amino acids into proteins. *Annu. Rev. Biochem.* **73:**147 – 176.

[6] Hahn, U., Palm, G.J., Hinrichs, W. (2004) Old codons, new amino acids. *Angew. Chem. Intl Ed. Engl.* **43:**1190 – 1193.

[7] Budisa, N. (2005) Reprogramming the cellular translation machinery. In: *Engineering the Genetic Code*, pp. 90 – 184. Wiley, Weinheim.

[8] Xie, J., Schultz, P.G. (2006) A chemical toolkit for proteins – an expanded genetic code. *Nature Rev. Mol. Cell Biol.* **7:**775 – 782.

[9] Wheatley, D.N., Inglis, M.S., Malone, P.C. (1986) The concept of the intracellular amino acid pool and its relevance in the regulation of protein metabolism, with particular reference to mammalian cells. *Curr. Top. Cell. Regul.* **28:**107 – 182.

[10] Wipf, D., Ludewig, U., Tegeder, M., Rentsch, D., Koch, W., Frommer, W.B. (2002) Conservation of amino acid transporters in fungi, plants and animals. *Trends Biochem. Sci.* **27:**139 – 147.

[11] Budisa, N. (2004) Prolegomena to future efforts on genetic code engineering by expanding its amino acid repertoire. *Angew. Chem. Intl Ed. Engl.* **43:**3387 – 3428.

[12] Kirk, K.L. (1991) *Biochemistry of Halogenated OrganicCompounds*. Plenum, New York.

[13] Sophianopoulou, V., Diallinas, G. (1995) Amino acid transporters of lower eukaryotes: regulation, structure and topogenesis. *FEMS Microbiol. Rev.* **16:**53 – 75.

[14] Kanai, Y., Endou, H. (2003) Functional properties of multispecific amino acid transporters and their implications to transporter-mediated toxicity. *J. Toxicol. Sci.* **28:**1 – 17.

[15] Uchino, H., Kanai, Y., Kim, D.K., Wempe, M.F., Chairoungdua, A., Morimoto, E., Anders, M.W., Endou, H. (2002) Transport of amino acid-related compounds mediated by L-type amino acid transporter 1 (LAT1): insights into the mechanisms of substrate recognition. *Mol. Pharmacol.* **61:**729 – 737.

[16] Kuhn, J., Somerville, R.L. (1974) Uptake and utilization of aromatic D-amino acids in Escherichia coli K12. *Biochim. Biophys. Acta* **332:**298 – 312.

[17] Kuhn, J., Somerville, R.L. (1971) Mutant strains of Escherichia coli K12 that use D-amino acids. *Proc. Natl Acad. Sci. U S A* **68:**2484 – 2487.

[18] Krajewska-Grynkiewicz, K., Walczak, W., Klopotowski, T. (1971) Mutants of Salmonella typhimurium able to utilize D-histidine as a source of L-histidine. *J. Bacteriol.* **105:**28 – 37.

[19] Gal, J., Szvetnik, A., Schnell, R., Kalman, M. (2002) The metD D-methionine transporter locus of Escherichia coli is an ABC transporter gene cluster. *J. Bacteriol.* **184:**4930 – 4932.

[20] Kadner, R.J. (1977) Transport and utilization of D-methionine and other methionine sources in Escherichia coli. *J. Bacteriol.* **129:**207 – 216.

[21] Cooper, S. (1966) Utilization of D-methionine by Escherichia coli. *J. Bacteriol.* **92:**328 – 332.

[22] Harrison, L.I., Christensen, H.N., Handlogten, M.E., Oxender, D.L., Quay, S.C. (1975) Transport of L-4-azaleucine in Escherichia coli. *J. Bacteriol.* **122:**957 – 965.

[23] Liu, D.R., Schultz, P.G. (1999) Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl Acad. Sci. U S A* **96:**4780 – 4785.

[24] Ibba, M., Söll, D. (1999) Quality control mechanisms during translation. *Science* **286:**1893 – 1897.

[25] Wolfson, A., Knight, R. (2005) Occurrance of the aminoacyl-tRNA synthetases in high-molecular weight complexes correlates with the size of substrate amino acids. *FEBS Lett.* **579:**3467 – 3472.

[26] Ibba, M., Becker, H.D., Stathopoulos, C., Tumbula, D.L., Soll, D. (2000) The adaptor hypothesis revisited. *Trends Biochem. Sci.* **25:**311 – 316.

[27] Jakubowski, H. (2003) Accuracy of Aminacyl-tRNA Synthetases: Proofreading of amino acids. In: *Aminoacyl-tRNA Synthetases*. (Ibba, M., Francklyn, C., Cusack, S., Eds). Landes Bioscience, Austin, Texas.

[28] Francklyn, C., Perona, J.J., Puetz, J., Hou, Y.M. (2002) Aminoacyl-tRNA synthetases: versatile players in the changing theater of translation. *RNA* **8:**1363 – 1372.

[29] Fersht, A.R., Dingwall, C. (1979) Editing mechanism for the methionyl-tRNA synthetase in the selection of amino acids in protein synthesis. *Biochemistry* **18:**1250 – 1256.

[30] Cowie, D.B., Cohen, G.N., Bolton, E.T., DeRobichon-Szulmajster, R.H. (1959) Amino acid analog incorporation into bacterial proteins. *Biochim. Biophys. Acta* **34:**39 – 46.

[31] Copley, S.D. (2003) Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr. Opin. Chem. Biol.* **7:**265 – 272.

[32] Richmond, M.H. (1962) Effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol. Rev.* **26:**398..

[33] Link, A.J., Mock, M.L., Tirrell, D.A. (2003) Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* **14:**603 – 609.

[34] Kast, P., Hennecke, H. (1991) Amino acid substrate specificity of *Escherichia coli* phenylalanyl-tRNA synthetase altered by distinct mutations. *J. Mol. Biol.* **222:**99 – 124.

[35] Döring, V., Mootz, H.D., Nangle, L.A., Hendrickson, T.L., de Crecy-Lagard, V., Schimmel, P., Marliere, P. (2001) Enlarging the amino acid set of Escherichia coli by infiltration of the valine coding pathway. *Science* **292:**501 – 504.

[36] Hartman, M.C.T., Josephson, K., Szostak, J.W. (2006) Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc. Natl Acad. Sci. U S A* **103:**4356 – 4361.

[37] Rodnina, M.V., Wintermeyer, W. (2001) Ribosome fidelity: tRNA discrimination, proofreading and induced fit. *Trends Biochem. Sci.* **26:**124 – 130.

[38] Hohsaka, T., Sato, K., Sisido, M., Takai, K., Yokoyama, S. (1993) Adaptability of nonnatural aromatic amino acids to the active center of the E. coli ribosomal A site. *FEBS Lett.* **335:**47 – 50.

[39] Shimizu, Y., Kanamori, T., Ueda, T. (2005) Protein synthesis by pure translation systems. *Methods* **36:**299 – 304.

[40] Hecht, S.M., Alford, B.L., Kuroda, Y., Kitano, S. (1978) "Chemical aminoacylation" of tRNA's. *J. Biol. Chem.* **253:**4517 – 4520.

[41] Forster, A.C., Tan, Z.P., Nalam, M.N.L., Lin, H.N., Qu, H., Cornish, V.W., Blacklow, S.C. (2003) Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proc. Natl Acad. Sci. U S A* **100:**6353 – 6357.

[42] Yamada, R.-H., Kera, Y., Takahashi, S. (2006) Occurrence and functions of free D-aspartate and its metabolizing enzymes. *Chem. Rec.* **6:**259 – 266.

[43] Yoshimura, T., Esak, N. (2003) Amino acid racemases: functions and mechanisms. *J. Biosci. Bioeng.* **96:**103 – 109.

[44] Ferri-Fioni, M.-L., Fromant, M., Bouin, A.-P., Aubard, C., Lazennec, C., Plateau, P., Blanquet, S. (2006) Identification in Archaea of a novel D-Tyr-tRNATyr deacylase. *J. Biol. Chem.* **281:**27575 – 27585.

[45] O'Connor, M., Lee, W.-C.M., Mankad, A., Squires, C.L., Dahlberg, A.E. (2001) Mutagenesis of the peptidyltransferase center of 23S rRNA: the invariant U2449 is dispensable. *Nucleic Acids Res.* **29:**710 – 715.

[46] Dedkova, L.M., Fahmi, N.E., Golovine, S.Y., Hecht, S.M. (2003) Enhanced D-amino acid incorporation into protein by modified ribosomes. *J. Am. Chem. Soc.* **125:**6616 – 6617.

[47] Dougherty, D.A. (2002) Unnatural amino acids as probes of protein structure and function. *Curr. Opin. Chem. Biol.* **4:**645 – 452.

[48] Hohsaka, T., Sisido, M. (2002) Incorporation of non-natural amino acids into proteins. *Curr. Opin. Chem. Biol.* **6:**809 – 815.

[49] Marsh, E.N.G., Neil, E. (2000) Towards the non-stick egg: designing fluorous proteins. *Chem. Biol.* **7:**R153-R157.

[50] Wiltschi, B., Budisa, N. (2007) Natural history and experimental evolution of the genetic code. *Appl. Microbiol. Biotechnol.* In press.

[51] 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. (2003) Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. *J. Mol. Biol.* **328:**1071 – 1081.

[52] Wang, L., Brock, A., Herberich, B., Schultz, P.G. (2001) Expanding the genetic code of Escherichia coli. *Science* **292:**498 – 500.

[53] Sisido, M., Hohsaka, T. (2001) Introduction of specialty functions by the position-specific incorporation of nonnatural amino acids into proteins through four-base codon/anticodon pairs. *Appl. Microbiol. Biotechnol.* **57:**274 – 281.