Biosynthetic Incorporation of Non-natural Amino Acids into Proteins

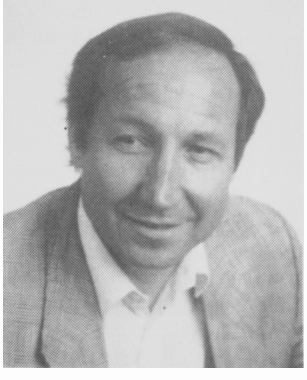
Josef Brunner
Department of Biochemistry, ETH-Zentrum, CH-8092 Zürich, Switzerland

1 Introduction

Since the elucidation of the structure of DNA in 1953 – the beginning of the modern era of molecular biology – many successful studies have been performed on the immensely complex and interrelated processes by which DNA is copied and then transcribed into RNA, and how the four letter alphabet of the latter is translated into the entirely different alphabet of polypeptides. One of the most fascinating aspects that has emerged from this work is the extraordinary accuracy with which these processes occur. Indeed, the maximal frequency with which wrong amino acids are inserted into a polypeptide chain is between only 1 in 1000 and 1 in 10000. To achieve such fidelity, the coordinated interplay of a large number of macromolecules is necessary.

Although a few synthetic, nearly isosteric analogues of amino acids, such as 3-fluorotyrosine or 6-fluorotryptophan, are known to be accepted by the biosynthetic machinery, it is quite clear that a general strategy to incorporate non-natural amino acids requires specific intervention at one or more steps of the translation. During recent years a general methodology has been developed which now permits a broad range of non-standard or non-natural amino acids to be incorporated at rationally selected sites and, hence, to generate polypeptides with novel structural and functional features. This offers a number of intriguing opportunities, with far-reaching implications for both basic research and for the engineering of proteins, in order to create novel types of catalysts or carriers of biologically active residues with pharmacologically interesting properties.

This review is intended to give a brief summary of some of the more recent developments and achievements. The first part is concerned with a discussion of some of the basic elements of translation, and of the manipulations that allow the bypassing of step(s) determining the high specificity of translation. In a subsequent section, the experimental approaches which have been developed to build up and assemble the molecules essential for these manipulations are outlined. The third part then summarizes some of the results that demonstrate the successful incorporation of non-natural amino acids into proteins. A final section will be devoted to some applications of this new technology. For reasons to be detailed there, we shall examine somewhat more closely the biosynthetic incorporation of photoactivatable amino acids into nascent polypeptides and the use of these polypeptides to identify biological targets.



Josef Brunner was born in 1943. After an apprenticeship as a laboratory technician, study of chemistry at the HTL Burgdorf (1964—67), and three years in industry, he began studying biochemistry at the ETH Zürich, from where he received his Ph.D. (with G. Semenza) in 1977. After postdoctoral work at Yale with Fred Richards (1977—79), he returned to the ETH where he has been an Oberassistant in the Department of Biochemistry. His principal research interest lies in the development of photocrosslinking methods for studying biological membranes.

2 The Role of Transfer RNAs in Protein Biosynthesis

As mentioned, the ribosome-catalysed synthesis of protein is an immensely complex process, and many important details still remain to be elucidated (an excellent overview is found in Stryer's textbook 'Biochemistry'[1]). To understand the manipulations necessary to incorporate non-natural amino acids into polypeptides, we must focus on the role of transfer RNAs (tRNAs) in the translation of the genetic information. The coded information is in the sequential triplets of messenger RNA (mRNA), each of which specifies a particular amino acid. In each cell, there is a set of tRNAs which is responsible for recognizing each of the 61 amino acid-specifying triplets (codons) and for carrying the cognate activated amino acid to the ribosome, where it participates in peptide synthesis. Thus, tRNAs act as the crossroads between the information-containing polynucleotide chains of the mRNA and those sites in the ribosome where this genetic information is expressed into the 20-letter alphabet of polypeptides. All cells are also endowed with a set of (at least) twenty enzymes, called aminoacyl-tRNA synthetases. Each aminoacyl-tRNA synthetase recognizes, in a highly specific manner, one of the 20 proteinogenic amino acids and catalyses the formation of an ester bond between the amino acid and the cognate tRNA molecule(s). In general, more than one species of tRNA can act as an acceptor for a particular amino acid, and because of the 'wobble' in base-pairing, some tRNAs can recognize more than one codon in the mRNA. Of the total 64 possible triplets, 61 code for the 20 amino acids, whereas the other 3 signal chain termination and, normally, are not read by tRNAs but rather recognized by proteins called release factors. Binding of a release factor to a termination codon somehow 'activates' peptidyl transferase so that it hydrolyses the peptidyl-tRNA ester bond allowing the polypeptide chain to leave the ribosome. As has become more and more clear during the past few years, in addition to this normal decoding, translation can involve non-standard decoding events such as ribosomal hopping, frameshifting, and reading through stop codons, all at unexpectedly high levels and using surprising mechanisms.[2] These are distinct from random translational errors and are largely determined by the structure of the mRNA being translated. It is important to refer to this fact, as it implies that the protein sequence cannot always be deduced from the nucleic acid sequence of the mRNA.

Transfer RNA molecules are small polynucleotides containing between 73 and 93 nucleotides. About half of their bases are paired and the molecules fold up into an L-shaped structure with two segments of double helices. As might be expected, tRNAs contain two functionally important elements, the amino acid attachment site which is at one end of the folded molecules and the codon recognition site – the so-called anticodon – located at the other end of the structure. During protein synthesis, the three anticodon bases interact with the codon bases of the mRNA, positioning the amino acylmoiety such that it can participate in the peptidyl transferase reaction. At no time does the side chain of the amino acid directly interact with the mRNA template. One of the constant features of tRNAs is the sequence of four non-paired residues at the amino acid acceptor (3') end which all terminate in -CpCpA. Recent evidence suggests that this part of the tRNA interacts with a distinct, highly conserved

region of 23S ribosomal RNA (rRNA), a finding which lends further support to the exciting hypothesis[3] that peptide bond formation is catalysed solely by rRNA This constant element of the amino acid acceptor end also provided a base for the development of a general methodology for the chemical misacylation of tRNAs All (natural) tRNA molecules contain a number of unusual bases formed by enzymatic modification of the tRNA precursor molecules The role of these modifications is not fully clear as yet, but, as will be shown in this article, they are not essential for effective participation in the peptidyl transferase reaction They may be important, however, for proper recognition by the aminoacyl tRNA synthetases, and also seem to stabilize the tRNA three-dimensional structure

More than thirty years ago, Chapeville *et al*[4] provided the first direct proof for the adapter function of the tRNA by showing that after the reductive desulfuration of cysteinyl-$tRNA^{cys}$ to alanyl-$tRNA^{cys}$ this alanyl residue is incorporated in response to a triplet coding for cysteine Using a related experimental approach, Johnson *et al*[5] later demonstrated that ribosomes also accept structurally modified amino acids A lysyl-tRNA acetylated at the N^{ϵ}-amino group was found to be incorporated into rabbit globin chains with nearly the same efficiency as the unmodified lysine

3 Experimental Approaches

To site-specifically incorporate non-natural amino acids into polypeptides by ribosome-based translation, a number of requirements must be fulfilled First, an mRNA must be generated in which the codon corresponding to the amino acid residue of interest is replaced by a codon not specifying one of the natural amino acids Secondly, a tRNA capable of recognizing this 'extra' codon must be constructed Thirdly, a procedure must be found to charge this tRNA properly with the non-natural amino acid And, fourthly, a translation system compatible with the afore mentioned manipulations is required

3.1 *In Vivo* and *in vitro* Translation

The most common and most efficient translation systems are living cells However, since all cells are surrounded by a membrane, which acts as an effective diffusion barrier, specific manipulation of an endogenous tRNA component in the manner outlined above is an extremely difficult task, not attainable at present However, it remains a great challenge for the future to overcome the various problems and to programme and manipulate *living* cells in a way such that they take up a non-natural amino acid and efficiently incorporate it into a protein component The reasons for this are that cells, especially bacteria, can grow on relatively cheap nutrients and produce high concentrations of protein, factors which are both important from a technological perspective of view Moreover, eukaryotic cells can carry out a number of post-translational modifications such as proteolysis, glycosylation, phosphorylation, sulfatation, fatty acylation, isoprenylation *etc*, many of which cannot be obtained, or with great difficulty only, in *in vitro* systems (see below) For the proper functioning of many proteins, post-translational modifications are essential

An attractive *in vivo* system is offered by the oocytes of *Xenopus laevis* These cells not only have an impressive diameter of about 1 mm, but a significant proportion of their content is represented by components required for the post-fertilization protein synthesis during early embryonic development These two facets have made oocytes a powerful system for studies of transcription, replication, assembly, and translation of micro-injected macromolecules[6] Although direct experimental evidence is still lacking, it is quite certain that oocytes would also utilize micro-injected, chemically misaminoacylated tRNA and produce very small amounts of mutant proteins containing a non-natural amino acid Although the amounts of protein generated in this way would hardly exceed those obtained from the more simple and economical *in vitro* translation systems (see below), the oocytes can correctly carry out post-translational modifications and direct proteins to specific intracellular compartments, and even export secretory proteins

Thus far, all successful approaches to incorporate non-natural amino acids into proteins have made use of suitably programmed *in vitro* translation systems supplemented with an exogenous, misacylated tRNA *In vitro* or cell-free translation is a well established and important technique to generate small amounts of protein Appropriate systems can be prepared easily from bacteria, yeast, as well as from cells of higher eukaryotes (plants, animals) Most frequently used are extracts from *Escherichia coli* and wheat germ, and cell lysates from rabbit reticulocytes (erythrocyte precursors that make large amounts of globin) Before use, these extracts are usually depleted of endogenous amino acids and energy sources and treated with ribonuclease to destroy their own mRNA Protein synthesis is then initiated by the addition of mRNA (obtained from natural sources or from *in vitro* transcription of plasmid DNA), amino acids, ATP, GTP, and an ATP-generating system consisting of creatine phosphate and creatine phosphokinase As mentioned previously, one of the limitations of a cell-free translation is the relatively low yield of protein Typically, only a few polypeptide chains are produced per mRNA chain, which corresponds to a few micrograms or a few tens of micrograms of protein per millilitre of cell lysate or extract It should be noted, however, that there may be ways to greatly increase translational efficiencies Thus, using continuous-flow systems, *in vitro* systems have been shown to be active for tens of hours and to produce up to several hundreds of polypeptide chains per copy of mRNA[7]

3.2 Choice of the tRNA

Two conceptually different approaches have been described for the site-specific incorporation of non-natural amino acids into proteins In the first approach,[8,9] advantage is taken of the fact that either of the three stop codons that normally signal chain termination can also function to direct incorporation of an amino acid provided a charged tRNA is available that can effectively compete with the release factor(s) in reading the 'stop' codon This principle, called suppression, is used by both bacteria and eukaryotic cells, for example for the translational insertion of the non-standard amino acid selenocysteine into formate dehydrogenase in *E coli* and into glutathione peroxidases of mammalian cells In the second approach, an extended genetic alphabet including a new base pair with a non-standard hydrogen bonding pattern is used to construct a 65th codon–anticodon pair[10,11]

3 2 1 Generation of Suppressor tRNA

For the purpose of incorporating a non-natural amino acid into a protein in response to a stop codon, a suppressor tRNA is needed that can effectively compete with the release factors Even though both bacteria and eukaryotic cells can, under selective pressure, make and use suppressor tRNAs, suppression is usually weak Were it otherwise, this would probably create a very serious problem as it would be generally difficult for the cell to terminate the translation of proteins correctly For this as well as for other reasons, efficient suppressor tRNAs cannot be (easily) obtained from natural sources

Efficient suppressor tRNAs can be constructed, however, by one of three procedures In the first, the anticodon-replacement procedure of Bruce and Uhlenbeck,[12] phenylalanine-specific tRNA from yeast ($tRNA^{Phe}$) serves as the starting material (Figure 1) The presence of a hypermodified, highly acid sensitive base (wyosine) within the anticodon loop of this tRNA provides an opportunity to open the loop chemically and remove the three adjacent anticodon bases by controlled digestion with ribonuclease This treatment also removes the two terminal nucleotides at the 3′-amino acid acceptor end The anticodon loop is then restored by enzymatically inserting a synthetic tetranucleotide containing the nucleotide sequence

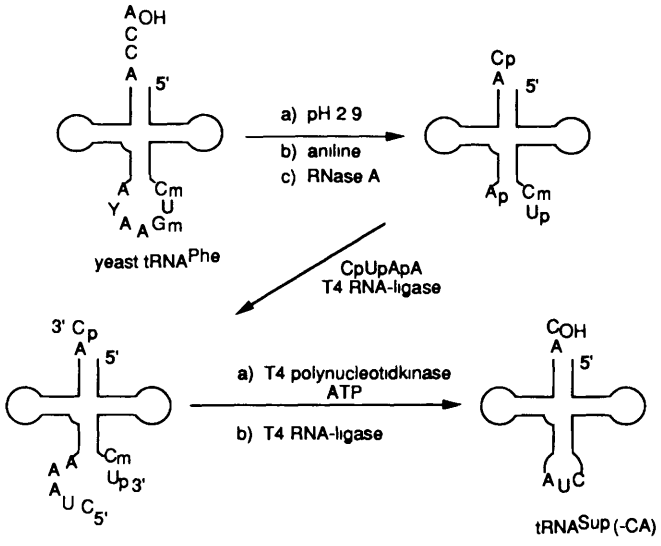


Figure 1 Steps involved in the construction of a 3′-abbreviated amber suppressor tRNA from yeast $tRNA^{Phe}$ by the anticodon replacement procedure

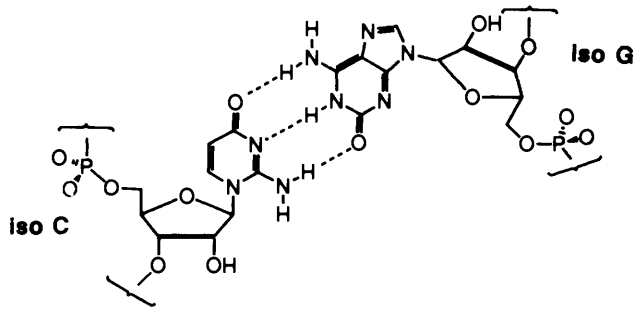


Figure 2 Structure of the iso-G–iso-C pair in forming a non-standard hydrogen bonding pattern

complementary to the stop codon The missing nucleotides at the 3′-terminus are replaced during chemical aminoacylation (see below)

The second procedure to prepare a suppressor tRNA is by run-off transcription of a DNA template which can be made by either of two related methods [13–15] The first utilizes plasmid amplification of a DNA template which contains both a T7 promoter and a suitable restriction site at the 3′-end of the tRNA gene Linearization of the overproduced plasmid followed by run-off transcription then affords the RNA with the desired sequence Alternatively, functional suppressor tRNA can be obtained by T7 polymerase-catalysed transcription of a synthetic linear DNA template featuring a double stranded promoter region and a long 5′-overhang corresponding to the transcribed region By either of these methods milligram amounts of tRNA can be produced Unlike natural tRNA, however, these run-off transcripts lack hypermodified bases, and there is a triphosphate rather than a monophosphate at their 5′-end The presence of this 5′-triphosphate was found to have no recognizable effect on the tRNA function during translation

Third, and finally, it may be pointed out that functional tRNAs have also been obtained by solid phase chemical synthesis using the phosphoramidite method for nucleotide coupling [16] However, despite the steady improvements of the corresponding methods, the preparation of a ribonucleic acid of the size of a tRNA still represents a considerable task Because of the necessity to protect the 2′-hydroxyl group of the individual building blocks, internucleotide phosphate bond formation is sterically hindered and slower than for the corresponding desoxyribonucleotides With the introduction and use of the *N,N*-diethylphosphoramidite ribonucleosides, which are more reactive than the older diisopropylphosphoramidites, and the use of optimized protocols for coupling and deprotection, a 74-mer suppressor tRNA could be synthesized, which after chemical aminoacylation displayed the same suppression efficiency as the corresponding tRNA (same sequence) obtained by run-off transcription [15]

3 2 2 Non-standard tRNA

As discussed by Benner and co-workers,[10] the geometry of the Watson–Crick base pairs can accommodate at least six mutually exclusive hydrogen bonding schemes of which, for whatever reason, only two are used by natural oligonucleotides One of the additional pairs is formed by iso-C and iso-G (Figure 2) In a joint effort, the groups of Benner and Chamberlin[11] succeeded in constructing by chemical synthesis a tRNA possessing the anticodon CU(iso-G) When chemically charged with an amino acid, this tRNA correctly reads the complementary codon (iso-C)AG introduced semisynthetically into an appropriate mRNA (see below)

3.3 Aminoacylation of tRNA

Although a variety of conditions (for example, in heterologous systems and in the presence of organic solvents) have been reported to lead to increased levels of misacylation of tRNA, natural aminoacyl-tRNA synthetases, because of their specificity, are unlikely to provide generally useful catalysts for charging tRNA with non-natural amino acids More promising in this respect might be synthetases genetically engineered to display altered substrate specificities In this way it was possible to activate a number of phenylalanine analogues [17] In the longer term, it may even be possible to replace synthetases by antibodies that catalyse the aminoacyl transfer reaction Initial work towards exploring this possibility was reported recently by Schultz and colleagues [18] They examined antibodies raised against a phosphonate diester which contains elements of the acyl donor, the acyl acceptor, and the leaving group, in a tetrahedral geometry mimicking that of the probable transition state for the transesterification reaction One of the antibodies was indeed found to be a remarkably efficient catalyst that provided further important insight into the requirements for efficient aminoacylation catalysts

Another approach to generate misaminoacylated tRNAs is by chemical modification of the amino acid side chain of natural aminoacyl-tRNAs However, because of the high degree of selectivity required to modify the amino acid side chain of an entire aminoacyl-tRNA molecule, the scope of this approach is very limited Examples where this approach has been applied successfully are the above mentioned conversion of cysteinyl-$tRNA^{cys}$ into alanyl-$tRNA^{cys}$ and the acylation of the ϵ-amino group of lysyl-tRNA (see also below)

The only truly general method for the aminoacylation of tRNA is that developed by Hecht and his group [19–21] Key elements of this method are the chemical (amino)acylation of the 3′(2′) hydroxyl group of the dinucleotide pCpA (or pdCpA) and the subsequent T4 RNA ligase-mediated coupling of this aminoacylated pCpA (pdCpA) to an abbreviated tRNA missing the 3′-terminal nucleotides pCpA (Figure 3)

Several procedures, employing different protection/deprotection schemes, have been developed for the (amino)acylation of the 3′-terminal dinucleotide (discussed in reference 22) To simplify the procedure and to improve the stability of the intermediates, the original procedures have been modified in two important respects First, the pCpA was replaced by pdCpA, a DNA/RNA hybrid The absence of a free hydroxyl group within the cytosine ribose leads to fewer side-products during acylation and, equally important, eliminates the risk of H^+-catalysed isomerization of the 3′–5′-phosphodiester bond during deprotection and purification of the (amino)acylated pCpA [23] Second, for protection of the α-amino group of the

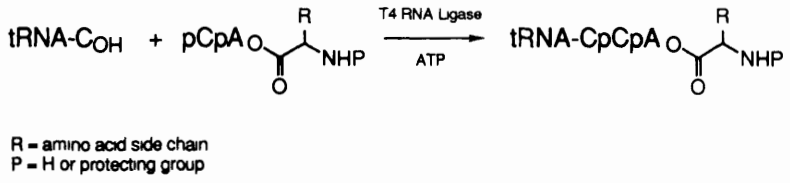


Figure 3 The key step in the chemical aminoacylation of tRNA

amino acid, protecting groups are employed that can be removed after the ligation reaction, *i e* on the intact tRNA Since the ester bond of aminoacyl-p(d)CpA and aminoacyl-tRNA are far more sensitive to hydrolysis than that of the α-amino group-protected counterparts, both the preparation of acyl-p(d)CpA and the ligation reaction are greatly facilitated Among the protecting groups found to satisfy most needs are pyroglutamyl- and the nitroveratryl group which can be removed enzymatically and photolytically, respectively (Figure 4) [24 25] Other amino protecting groups also reported to be compatible with this general scheme are the nitrophenylsulfenyl (reductive removal) and the 2-(4-biphenyl)-prop-2-yloxycarbonyl (Bpoc) groups (acidic removal) [25]

Much progress has been made also in defining conditions favouring mono *O*-acylation of unprotected p(d)CpA Presumably the most simple and efficient method currently available is that developed by Schultz and co-workers [22] In this procedure, unprotected pdCpA is acylated with the cyanomethylester-activated N^{α}-(nitroveratryl)-blocked amino acid to give high yields (up to 80%) of the 3′(2′)-*O*-acyl-pdCpA which, without further purification, could be used for the ligation reaction It may be of interest that preferential, or selective, *O*-acylation of unprotected pCpA can also be accomplished with *N*,*N*′-carbonyldiimidazole-activated Boc-protected amino acids using an acetonitrile–tetrahydrofuran–water solvent system [26]

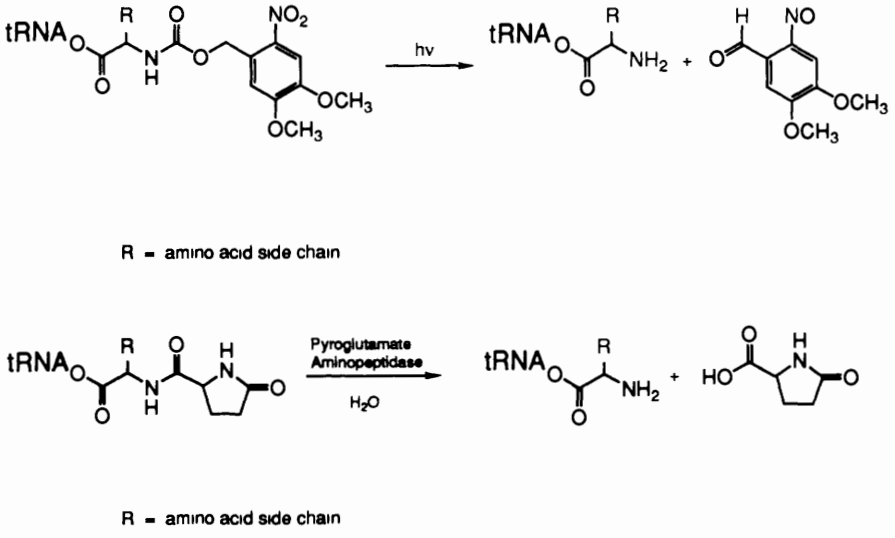
Although the procedure of Schultz has allowed the preparation of a broad variety of aminoacyl-tRNAs, it should be kept in mind that amino acids containing reactive side-chain functions require additional or alternative protection For example, the nitroveratryl group cannot be used to protect the amino group of the light-sensitive 4-(trifluoromethyldiazirinyl)-phenylalanine, a precursor of a highly reactive carbene used for photocrosslinking purposes [23] The final sections of this article contain a brief presentation and discussion of work aimed at acylating tRNA with an amino acid carrying additional sensitive functions, including a radioisotope of high specific radioactivity

Figure 4 Schemes for the deprotection of N^{α}-nitroveratryl- and pyroglutamoyl-protected aminoacyl-tRNAs

4 Participation of Misacylated tRNA in Protein Synthesis

Having outlined the experimental strategies to prepare misacylated (acyl)aminoacyl-tRNAs, this section focuses on the behaviour of these species during ribosome-catalysed protein-synthesis The key step in polypeptide chain elongation is peptide bond formation, a reaction catalysed by peptidyltransferase The substrates for this process are *peptidyl*-tRNA and *aminoacyl*-tRNA which are bound to distinct sites on the ribosome, referred to as the P site (the peptidyl donor) and the A site (the peptidyl acceptor), respectively To further refine our understanding of this reaction and to define the structural and spatial parameters requisite for effective participation of aminoacyl- and peptidyl-tRNAs, Hecht and colleagues prepared a variety of structurally modified peptidyl-tRNAs and examined their ability to bind to the P-site and to function as peptidyl donors for phenylalanyl-tRNA in the ribosomal A site [27 28] A surprising variety of the modified peptidyl-tRNAs were found to participate in ribosome-mediated peptide formation, or even in formation of products with altered connectivity As an example, *N*-(chloroacetyl)phenylalanyl-tRNA^{Phe} gave rise to two 'peptides', the expected *N*-chloroacetyl phe-phe, and the product resulting from reaction of the N^{α}-amino group of the phenylalanyl-tRNA with the peptidyl-tRNA at an electrophilic position different from the peptidyl-tRNA carboxylate ester (Figure 5) Thus, there seems to be considerable flexibility in the peptidyl transferase reaction, which may have to do with the fact that the ribosomal RNA plays a more active role in the peptidyl transferase reaction and that peptide bond formation may actually be catalysed by this component and not by ribosomal proteins

With the accessibility of chemically misacylated *amino*acyl-tRNAs, the above studies could be extended to include also experiments with structurally altered substrates (peptidyl-tRNA acceptors) for the ribosomal A site Most of this work was done by Schultz and his group, who demonstrated that a range of backbone and side-chain modified amino acids can be incorporated into growing polypeptides [25] The successful replacements include α,α-disubstituted amino acids, *N*-alkyl amino acids, and lactic acid, an isoelectronic analogue of alanine D-Amino acids were not incorporated It is possible, however, that the discrimination against D-amino acids arises in part at the level of formation of the tRNA–EF–Tu–GTP ternary complex,



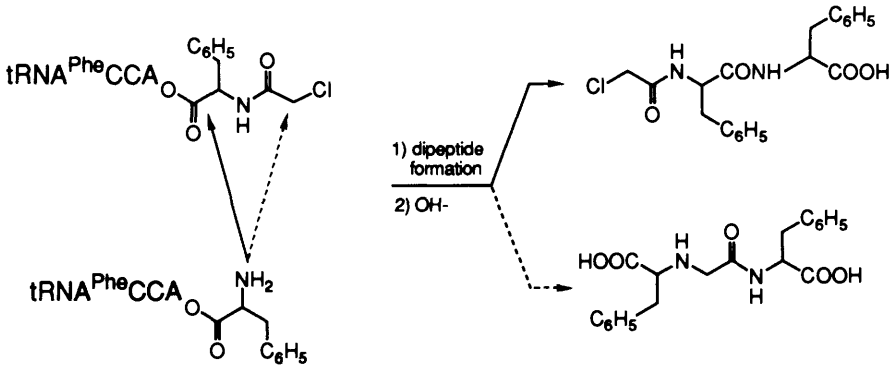


Figure 5 Reactions catalysed by ribosomal peptidyl transferase using a chemically charged, structurally modified peptidyl-tRNA as a P-site substrate.

which carries the aminoacylated tRNA to the ribosome, rather than in the peptidyltransferase reaction itself. Ribosomes seem also to tolerate amino acids with abnormally bulky side chains, as is the case for a number of N^{ϵ}-modified lysines, including N^{ϵ}-biotinyllysine and the fluoroescent probe N^{ϵ}-fluoresceinthiocarbomoyllysine.

A detailed analysis of read-through, suppression, and site-specificity of the incorporation of a non-natural amino acid into a short model peptide was reported by Chamberlain and colleagues.[15] Translations were done with reticulocyte lysate in the presence of tRNA suppressors that had been obtained by run-off transcription of a synthetic DNA template and chemically misacylated with [^{125}I]iodotyrosine. Under optimal conditions, the efficiency of suppression of the UAG (amber) was 63%, and the [^{125}I]iodotyrosine was incorporated exclusively at the position of the UGA stop codon. Consistent with observations also made by Schultz's group, suppression with aminoacylated tRNAs lacking hypermodified bases was strongly dependent upon the Mg^{2+} concentration in the translation system. Chamberlain and colleagues also compared amber (UAG), opal (UGA), and ochre (UAA) suppressors. Interestingly, the opal suppressor gave the highest level of suppression for incorporation of [^{125}I]iodotyrosine (81%), while the ochre suppressor gave the lowest level (48%), compared with 63% for the amber suppressor.

Thus far, the highest level of *in vitro* incorporation of a non-natural amino acid into a polypeptide chain was achieved by employing the non-standard tRNA possessing the anticodon CU(iso-dG) ($tRNA_{CU(iso\text{-}dG)}$) and an mRNA containing the complementary codon (iso-C)AG.[11] While, as noted above, the level of suppression of stop codons was between 48 and 81%, read-through of the (iso-C)AG codon was 90%. The specificity of translation of this 65th codon was also high. Neither semi-synthetic suppressor $tRNA_{CUA}$ nor any natural tRNA allowed reading of the (iso-C)AG codon. When the ribosome encountered the (iso-C)AG codon in the absence of charged $tRNA_{CU(iso\text{-}dG)}$, the primary outcome was continued translation following a frameshift that skipped the iso-C base, a finding strongly supporting the view that termination of translation depends upon binding of release factors to stop codons to prevent frameshifting.

5 Areas of Application

With the accessibility of proteins containing 'designer' amino acids at selected positions, new avenues for research in several areas have been opened. The remainder of this article summarizes some recent biochemical and biophysical analyses of mutated proteins carried out recently by Schultz's group, as well as efforts in our laboratory to utilize the technology for the preparation of a new generation of macromolecular photoaffinity- and photocrosslinking probes.

5.1 Biochemical and Biophysical Analyses of Mutated Proteins

In a first study[8] reported by Schultz's group, large scale *in vitro* translations were used to prepare truncated versions of β-lactamases, wherein Phe-66, a residue near the active site and conserved in four class A β-lactamases, was replaced by several analogues of Phe. Yields of mutated β-lactamase obtained in this *E. coli* translation system were estimated to be 5.5 to 7.5 μg per millilitre, which represents 15—20% suppression efficiency (note that this suppression efficiency is much lower than those referred to above obtained with the reticulocyte lysate). While several of these mutant proteins could be purified to near homogeneity and characterized by their K_m and k_{cat} values, those mutants containing substitutions corresponding to the greatest steric perturbations were unstable and presumably were proteolysed during purification attempts. This latter observation is consistent with earlier mutagenesis experiments which demonstrated that Phe-66 plays an important structural role. The site-specific incorporation of amino acids with side chains that meet specific steric, electronic, and other constraints can greatly extend the scope of mutagenesis in studying the folding, packing, and stabilities of proteins, as was demonstrated by another study.[29] There, specific replacements of an interior residue in T4 lysozyme were used to evaluate the effects on protein stability of the stepwise removal of methyl groups from the hydrophobic core, of side-chain solvation, of packing density, and of side-chain conformational entropy. Perhaps not surprisingly, all the mutations examined significantly influenced protein stability, a point to consider in any mutational study.

Proteins containing isotopically substituted amino acids at defined positions offer new opportunities for studying proteins by NMR spectroscopy.[30] To illustrate this, a single Ala residue (Ala-82) in T4 lysozyme was replaced by [^{13}C]alanine. Using ^{13}C-filtered ^{1}HNMR spectroscopy, the α-H and the CH_3 of this single alanyl residue were clearly visible, implying that NMR spectroscopy of mutant proteins may be successfully used to determine chemical shifts, pK_a values, and relaxation parameters for individual amino acids in both native and denatured proteins. *In vitro* translation may also allow incorporation of amino acids containing nitroxide free radicals into proteins. These could serve then as electron spin resonance probes for a wide range of structural and dynamic studies.[31] Notably the fact that spectra can be collected from 10μl-samples containing micromolar or sub-micromolar concentrations of protein (< 10 pmoles) would make such applications particularly attractive. An important question still to be answered is whether conditions for translation can be found that preserve the sensitive nitroxide functionality.

5.2 *In vitro* Synthesis of Photocrosslinking Probes

The terms photocrosslinking and photoaffinity labelling refer to a variety of techniques that very often serve to identify the receptor for a given polypeptide or protein.[32] In the conventional approaches, this is usually pursued by covalently crosslinking the two interacting species through the action of homo- and heterobifunctional reagents. Efforts are now being made to extend this popular technique by directly incorporating a pho-

toactivatable amino acid at a single, predetermined site into the ligand and, hence, to generate photoaffinity probes of defined structure and reactivity

The potential of this approach is already evident from a respectable number of studies in which photoactivatable amino acids were translationally incorporated into proteins destined to be targeted to and into membrane vesicles derived from the endoplasmic reticulum (reviewed in reference 32) Every protein destined for a specific cellular membrane, or sub-cellular compartment, contains a corresponding 'address label' In the case of the endoplasmic reticulum this is the so-called signal sequence, a predominantly hydrophobic stretch of 15—25 amino acids at or near the N-terminus of the newly synthesized protein When this signal emerges from the ribosome, it is first recognized by a cytosolic factor, the signal recognition particle (SRP), whose binding slows down or even stops further translation until the ribosome–SRP–nascent chain complex docks to a receptor in the membrane where the signal is transferred to integral components to initiate membrane insertion/translocation By means of proteins containing photoactivatable amino acids within or near the signal sequence, three components could thus far be identified which transiently contact the signal and are functionally engaged in the processes of protein targeting to and translocation across membranes There is hope that basically the same approach can be applied to identify the 'receptors' for other signals, especially for those whose recognition is directed by sequence-independent, colligative properties As discussed by Rothman,[33] this principle of recognition is relatively new in biology and may be important in other arenas such as in the distinction between folded and unfolded proteins by heat shock proteins, or the targeting of transport vesicles for membrane fusion Since a photoactivatable residue may be positioned within or near such a signal without seriously affecting its interaction with the target, photoactivatable proteins could emerge as powerful tools to characterize the interactions underlying cellular recognition, folding, and sorting events However, before the potential of this methodology can be fully exploited, a difficult technical problem has to be solved This has to do with the very limited capacities of the *in vitro* systems for studying these interactions, with typical yields of crosslinking products being in the femtomole or sub-femtomole range Unfortunately, no general and efficient method is thus far available to identify and characterize the crosslinked target components

At least a part of this problem might be solved by adopting a strategy ('label-transfer crosslinking') successfully employed in conventional crosslinking [32] Accordingly, the side chain of the photoactivatable amino acid would have to be made cleavable and equipped within the photoactivatable moiety with a reporter group (Figure 6)

Following translational incorporation of such an amino acid and its crosslinking to a target molecule, cleavage of the susceptible bond would result in the formal transfer of the reporter group onto the target, thereby yielding simply labelled targets which should now be far more easy to characterize than the original crosslinking product While it is not difficult to select the individual functional elements, their joining together into a compact structure, representing the side chain of a tRNA-activated amino acid, is not a trivial task The difficulty arises mainly from the fact that only a radioisotope reporter group of high specific radioactivity (*e g* ^{32}P, ^{35}S, or ^{125}I) would give the necessary detection sensitivity and would also be small enough to be compatible with this approach While ^{32}P and ^{35}S are not available in synthetically convenient forms, manipulation of ^{125}I, for safety reasons, is usually restricted to millicurie or, chemically speaking, to nanomole quantities This poses further severe constraints on the design and synthesis of such a multifunctional amino acid

Work in our laboratory is directed towards realizing this task [26] The basic approach chosen is summarized in Figure 7 In this approach, the side chain sulfhydryl function of either cysteinyl-pCpA (prior to ligation to abbreviated $tRNA^{sup}$) or cysteinyl-tRNAsup (after ligation) is alkylated with a reagent

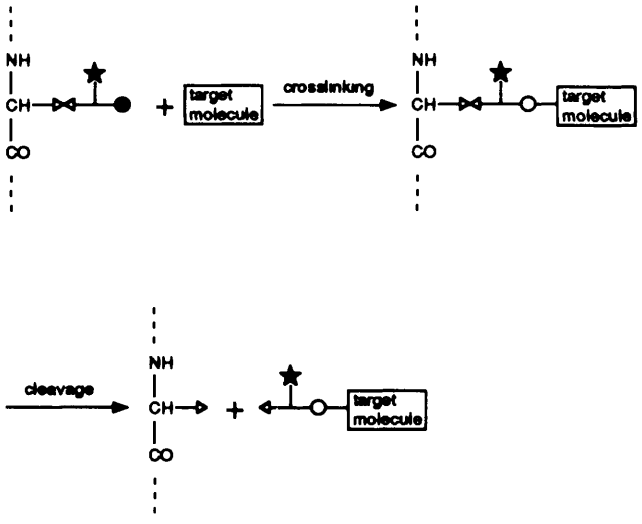


Figure 6 Schematic representation of 'Label Transfer Crosslinking' The side chain of the crosslinking amino acid contains a photoactivatable group (filled circle), a radioisotope reporter group of high specific radioactivity (asterisk), and a cleavable element (a pair of triangles) Following covalent photocrosslinking to a target molecule, the susceptible element is cleaved whereby the reporter group is formally transferred onto the target component

combining the necessary elements for label-transfer crosslinking The rate at which cysteine and other thiols react with *N*-alkyllmaleimides suggested that the reaction would be fast and specific enough to successfully join the two reactants, even though both might be present initially in sub-millimolar concentrations and the reaction time would have to be kept short (minutes) to minimize hydrolysis of the aminoacyl-tRNA (aminoacyl-pCpA) ester bond

2′(3′)-*O*-cysteinyl-pCpA could be obtained in approximately 20% yield by acylation of pCpA with *N*-Boc-dimethylthiazolidine and subsequent *N,S*-deprotection of the mono *O*-acyl product by successive treatments with trifluoroacetic acid and water Ligation of cysteinyl-pCpA to abbreviated $tRNA^{sup}$ then

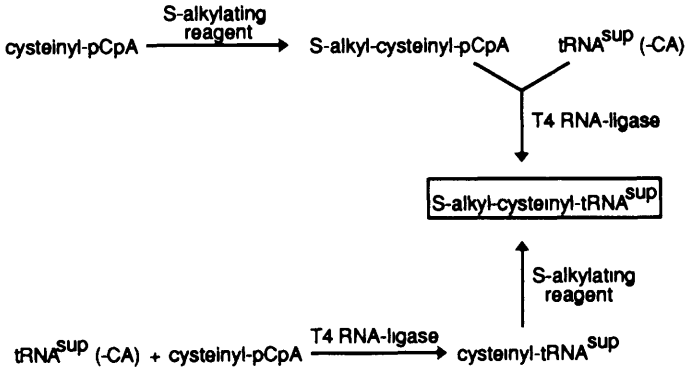
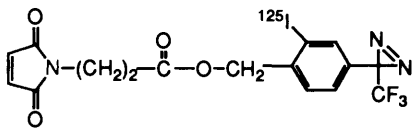


Figure 7 Scheme for the preparation of multifunctional, tRNA-activated amino acids by *S*-alkylation of a cysteinyl side chain

afforded the desired cysteinyl-$tRNA^{sup}$ We could show also that the thiol group of cysteinyl-pCpA is fully available for *S*-alkylation by various *N*-alkylmaleimides including [^{125}I]TIDM-3, a reagent which, at least formally, meets the criteria for a label transfer photocrosslinker (Figure 8)

In contrast, *S*-modification of cysteinyl-$tRNA^{sup}$ proved more difficult, and yields ranging from 20—90% have been achieved Presumably during work-up and isolation of the cysteinyl-$tRNA^{sup}$, a fraction of the thiol underwent oxidation Current efforts are aimed at preparing cysteinyl-$tRNA^{sup}$ alkylated with [^{125}I]TIDM-3 of high specific radioactivity (> 100 Ci $mmol^{-1}$) and at assessing translational incorporation of the



[^{125}I]TIDM-3

Figure 8 Structure of [^{125}I]TIDM-3, a thiol-specific (*N*-alkylmaleimide), cleavable (carboxylate ester) label-transfer photocrosslinker that can be prepared with an extremely high specific radioactivity (> 2000 Ci mmol^{-1})

modified cysteine into polypeptides Although this modified cysteine is an abnormally bulky residue, initial data indicate that it is accepted by the ribosomes Moreover, when placed within the signal sequence of a nascent preprolactin chain, this chain can be photocrosslinked to the signal sequence-binding subunit of SRP, strongly supporting the view that the amino acid replacement did not abolish proper recognition

In conclusion, proteins containing non-natural or modified natural amino acids at rationally selected sites provide novel and unique tools that will provide further exciting new insights into the structure and function of individual proteins, as well as into their interactions with other proteins or components of such complex systems as biological membranes, organelles, or even whole cells However, before the full potential of this technology can be realized, a number of challenging chemical and biochemical problems remain to be solved

Acknowledgements The author thanks Drs N Mantei and B Martoglio, and T Weber and C Gahwiler for assistance in the preparation of the manuscript It is also a pleasure to acknowledge the financial support by The Swiss National Science Foundation The work on label-transfer photocrosslinking was carried out in the laboratory of Prof Semenza

6 References

1 L Stryer, 'Biochemistry', 3rd Edition, Feeman & Company, New York, 1988
2 J F Atkins, R B Weiss, and R F Gesterland, *Cell*, 1990, **62**, 413
3 H F Noller, V Hoffarth, and L Zimniak, *Science*, 1992, **256**, 1416
4 F Chapeville, F Lippmann, G von Ehrenstein, B Weisblum, W J Ray, and S Benzer, *Proc Natl Acad Sci , USA*, 1962, **48**, 1086
5 A E Johnson, W R Woodward, E Herbert, and J R Menninger, *Biochemistry*, 1976, **15**, 569
6 A Colman, in 'Transcription and Translation', ed B D Hames and S J Higgins, IRL Press, 1984, Oxford, p 271
7 A S Spirin, V I Baranov, L A Ryabova, S Y Ovodov, and Y B Alkahov, *Nature*, 1988, **242**, 1162
8 C J Noren, S J Anthony-Cahill, M C Griffith, and P G Schultz, *Science*, 1989, **244**, 182
9 J D Bain, C G Glabe, T A Dix, and A R Chamberlin, *J Am Chem Soc* , 1989, **111**, 8013
10 J A Piccirilli, T Krauch, S E Moroney, and S A Benner, *Nature*, 1990, **343**, 33
11 J D Bain, C Switzer, A R Chamberlin, and S A Benner, *Nature*, 1992, **356**, 537
12 A G Bruce and O C Uhlenbeck, *Biochemistry*, 1982, **21**, 855
13 C J Noren, S J Anthony-Cahill, D J Suich, K A Noren, M C Griffith, and P G Schultz, *Nucleic Acid Res* , 1990, **18**, 83
14 J D Bain, D A Wacker, E E Kuo, M H Lyttle, and A R Chamberlin, *J Org Chem* , 1991, **56**, 4615
15 J D Bain, E S Diala, C G Glabe, D A Wacker, M H Lyttle, T A Dix, and A R Chamberlin, *Biochemistry*, 1991, **30**, 5411
16 M H Lyttle, P B Wright, N D Sinha, J D Bain, and A R Chamberlain, *J Org Chem* , 1991, **56**, 4608
17 P Kast, Ph D Thesis, 1991, ETH, Zurich
18 J R Jacobsen, J R Prudent, L Kochersperger, S Yonkovich, and P G Schultz, *Science*, 1992, **256**, 365
19 J M Pezzuto and S M Hecht, *J Biol Chem* , 1980, **255**, 865
20 T G Heckler, Y Zanma, T Naka, and S M Hecht, *J Biol Chem* , 1983, **258**, 4492
21 T G Heckler, L -H Chang, Y Zama, T Naka, M S Chorghade, and S M Hecht, *Biochemistry*, 1984, **23**, 1468
22 S A Robertson, J A Ellman, and P G Schultz, *J Am Chem Soc* , 1991, **113**, 2722
23 G Baldini, B Martoglio, A Schachenmann, C Zugliani, and J Brunner, *Biochemistry*, 1988, **27**, 7951
24 J R Roesser, C Xu, R C Payne, C K Surratt, and S M Hecht, *Biochemistry*, 1989, **28**, 5185
25 J Ellman, D Mendel, S Anthony-Cahill, C J Noren, and P G Schultz, *Methods Enzymol* , 1991, **202**, 301
26 B Martoglio, Ph D Thesis, 1992, ETH, Zurich
27 J R Roesser, M S Chorghade, and S M Hecht, *Biochemistry*, 1986, **25**, 6361
28 T G Heckler, J R Roesser, C Xu, P -I Chang, and S M Hecht, *Biochemistry*, 1988, **27**, 7254
29 D Mendell, J A Ellman, Z Chang, D L Veenstra, P A Kollman, and P G Schultz, *Science*, 1992, **256**, 1798
30 J A Ellman, B F Volkman, D Mendel, P G Schultz, and D E Wemmer, *J Am Chem Soc* , 1992, **114**, 7959
31 G L Millhauser, *Trends Biochem Sci* , 1992, **17**, 448
32 J Brunner, *Annu Rev Biochem* , 1993, **62**, 483
33 J E Rothman, *Nature*, 1989, **340**, 433