Incorporation of non-natural amino acids into proteins
Takahiro Hohsaka and Masahiko Sisido*

Chemical and biological diversity of protein structures and functions can be widely expanded by position-specific incorporation of non-natural amino acids carrying a variety of specialty side groups. After the pioneering works of Schultz’s group and Chamberlin’s group in 1989, noticeable progress has been made in expanding types of amino acids, in finding novel methods of tRNA aminocacylation and in extending genetic codes for directing the positions. Aminoacylation of tRNA with non-natural amino acids has been achieved by directed evolution of aminoacyl-tRNA synthetases or some ribozymes. Codons have been extended to include four-base codons or non-natural base pairs. Multiple incorporation of different non-natural amino acids has been achieved by the use of a different four-base codon for each tRNA. The combination of these novel techniques has opened the possibility of synthesising non-natural mutant proteins in living cells.

Addresses
Department of Bioscience and Biotechnology, Okayama University, 3-1-1 Tsushimanaka, Okayama 700-8530, Japan
*e-mail: sisido@cc.okayama-u.ac.jp

Current Opinion in Chemical Biology 2002, 6:809–815
1367-5931/02/$ — see front matter © 2002 Elsevier Science Ltd. All rights reserved.

Published online 18 October 2002

Abbreviations
aaRS aminoacyl-tRNA synthetase
aa-tRNA aminoacyl-tRNA
DHFR dihydrofolate reductase

Introduction
Amino acid substitution of proteins by genetic engineering had been restricted to the 20 naturally occurring amino acids. In 1989, the amber suppression method for introducing non-natural amino acids into specific positions of proteins was developed, opening the way to expanding protein functions [1,2]. In this method, amber suppressor tRNAs are aminoacylated with desired non-natural amino acids through chemical aminocacylation (i.e. an enzymatic ligation of an aminoacyl-dinucleotide with a tRNA lacking the terminal dinucleotide unit) [3]. The resulting aminoacyl-tRNAs (aa-tRNAs) are added to an in vitro translation system or Xenopus oocytes together with a mRNA or DNA of interest containing an amber stop codon (UAG) at a desired position. The amber codon is suppressed by the added aa-tRNA, resulting in the incorporation of the non-natural amino acid into the directed position of the target protein.

This method, however, suffers from several disadvantages. First, stop codons are under tough competition with the release factors that terminate peptide elongation. Thus, the release factors interfere with the incorporation of non-natural amino acids. Second, because among three stop codons, only the amber stop codon can be used for the suppression, multiple incorporation of non-natural amino acids into single proteins cannot be expected [4,5].

Preparation of aa-tRNAs is another issue to be improved, especially for a large-scale expression of non-natural proteins. Once an aa-tRNA is used for translation or is hydrolyzed at an aminoacyl bond, the deacylated tRNA will not be aminoacylated again in the cell-free translation systems or Xenopus oocytes. This disadvantage significantly decreases the yield of non-natural proteins.

This review focuses on recent improvements of non-natural amino acid mutagenesis addressed to these problems. Applications of non-natural mutagenesis to create proteins with specialty functions are also introduced.

Extension of codon-anticodon pairs
The genetic code consists of 61 codons for assigning 20 naturally occurring amino acids and 3 codons as stop signals. To fit non-natural amino acids into the existing genetic code system, it is essential to generate new codons that are specific to them. The use of an amber stop codon (Figure 1a) is one of the solutions without large alteration of the existing codon system. Alternatively, four-base codons have been developed (Figure 1b) [5,6]. In an Escherichia coli in vitro translation system, various kinds of four-base codons — AGGU, CCGU, CGAU, CCCU, CUCU, CUAA, GGGU, and their fourth-letter variants — were able to assign non-natural amino acids within the framework of the existing three-base codon system. Among them, GGGU gave the most efficient result, in which p-nitrophenylalanine (I, Figure 2) was introduced into streptavidin in 86% yield relative to a wild-type streptavidin. Because most of the four-base codons listed above are orthogonal to each other, double incorporation of two different non-natural amino acids into individual positions of streptavidin has been achieved by using efficient sets of the four-base codons such as GGGU and CCGG [5]. Surprisingly, codons could be further extended to five-base codons such as CGGN1N2, where N1 and N2 indicate one of four nucleotides. Indeed, the five-base codons were decoded by aa-tRNAs containing the corresponding five-base anticodons [7]. In addition, we have found that several kinds of four-base codons were also highly effective in a rabbit reticuloocyte lysate (Hohsaka T, Sisido M, unpublished data). Magliery et al. undertook an in vivo selection of efficient four-base codons by a library method [8,9], and four four-base codons, AGGA, CUAG, UAGU and CCCU, were identified [8].

Another strategy for the extension of the genetic code has been proposed by using non-natural base pairs that are orthogonal to the naturally occurring four types. In 1991,
Bain et al. [10] showed that an isoC and isoG base pair is correctly formed when it is incorporated into codon and anticodon, respectively. Although the isoC–isoG pair works effectively in the translation process, the mRNA and tRNA containing the non-natural bases should be synthesized by chemical methods. Recently, novel non-natural base pairs that fit into the replication and transcription processes have been designed. Hirao et al. [11**,12] reported that a non-natural base, pyridine-2-one (y), was specifically introduced into a mRNA corresponding to a 2-amino-6-(2-thienyl)purine (s) in a DNA template. The non-natural codon yAG was successfully translated by a tRNA containing the corresponding non-natural anticodon CUs (Figure 1c) and aminoacylated with m-chlorotyrosine (2). Although the anticodon could not be introduced into the tRNA through transcription in this report, the result clearly demonstrates the feasibility of using a non-natural base pair to assign non-natural amino acids in in vitro and in vivo.

**Figure 1**

Non-natural amino acids discussed in the text.

**(a) Amber codon**

**(b) Four-base codon**

**(c) Non-natural codon**

**Figure 2 legend**

Codons that can be assigned to non-natural amino acids. **(a)** Amber codon. An amber codon UAG is decoded by aa-tRNA containing a CUA anticodon. **(b)** Four-base codon. CGGG is shown as an example of four-base codons. The CGGG codon is decoded by aa-tRNA containing the corresponding anticodon CCCG. **(c)** Non-natural codon. A recently developed non-natural codon-anticodon pair yAG-CUs and structures of y and s are shown. The yAG codon makes a pair with the CUs anticodon exclusively through a cell-free translation. The gray oval represents an amino acid.

**Aminoacylation of tRNAs with non-natural amino acids**

Because the chemical aminoacylation (Figure 3a) cannot be made in living cells, the non-natural mutagenesis had been carried out only in cell-free translation systems or in *Xenopus* oocytes through microinjection of the aa-tRNA. Exceptionally, amino acids that are analogous to natural ones are accepted as substrates by one of the aminoacyl-tRNA synthetases (aaRSs). For example, Kiick et al. [18*] found that a methionine analog, azidohomoalanine (3), was accepted by MetRS and incorporated into proteins in place of methionine in *E. coli*. The incorporated azido group was selectively modified by the Staudinger ligation with a phosphine-containing peptide.

Recently, non-natural aminoacylation has been achieved by mutated aaRS (Figure 3b) by Schultz’s group. A pair of an amber suppressor tRNA[ Tyr] and a TyrRS from *Methanococcus jannaschii* has been mutated to become orthogonal to any aaRS/tRNA pair from *E. coli* [19,20]. The latter TyrRS was further mutated not to accept tyrosine or any other amino acids, but to accept O-methyl-tyrosine (4) exclusively as the substrate [21**]. Five residues located in the active site of the TyrRS were randomly mutated. The library of the mutant TyrRSs was screened by two rounds of positive selection to accept O-methyl-tyrosine and subsequent negative selection to reject any natural amino acids. The final aaRS was found to accept O-methyl-tyrosine exclusively. Co-expression of the mutant TyrRS and amber suppressor tRNA in the presence of a dihydrofolate reductase (DHFR) gene containing an amber codon produced a mutant DHFR protein containing an O-methyl-tyrosine at the amber codon position in *E. coli* [21**]. An ion trap mass analysis demonstrated that O-methyl-tyrosine was incorporated at more than 95% fidelity.

Similar approaches were carried out for 2-naphthylalanine (5) [22], p-azidophenylalanine (6) [23], and p-benzoylphenylalanine (7) [24**]. In addition, mutant aaRSs for p-aminophenylalanine (8) and p-isoproplyphenylalanine (9) were also selected [25*]. Mutant aaRSs for non-natural amino acids with spin (10), fluorescent (11,12), and biotin labels (13), an aldehyde (14) or allyl (15) functionality, a metal binder (16), photocaged groups (17,18), a photoisomerizable group (19), and a sugar moiety (20) are being investigated [25*].

Hirao, Yokoyama and co-workers [26*] reported a mutated *E. coli* TyrRS that recognizes m-iodotyrosine (21) more efficiently than tyrosine. Three positions of the TyrRS were systematically mutated, and the aminoacylation activities of each mutant with iodotyrosine and with tyrosine were compared by use of an amber suppressor *E. coli* tRNA[ Tyr]. Then, two effective point mutations were combined together. One of these double mutants, V57C/I95, recognized iodotyrosine 10-fold more efficiently than it recognized tyrosine. Addition of the mutated enzyme and the amber suppressor tRNA to a wheat germ cell-free translation system expressing a c-Ha-Ras gene with an amber codon gave the mutant Ras protein containing an iodotyrosine at the directed position. LC-MS (liquid chromatography coupled with mass spectrometry) analysis demonstrated that iodotyrosine was incorporated in more than 95% fidelity. The wheat germ cell-free translation system has the advantage that eukaryotic proteins are liable to fold correctly. In addition, the *E. coli* TyrRS-tRNA[ Tyr] pair was orthogonal to wheat germ tRNA-aaRS pairs without any mutagenesis.

RajBhandary and co-workers [27,28] have developed a method for importing aminoacylated amber and ochre suppressor tRNAs into mammalian cells. They showed that both suppressor tRNAs aminoacylated with tyrosine were imported by using a transfecting reagent Effectene, and an active chloramphenicol acetyl transferase was expressed from co-transfected chloramphenicol acetyl transferase genes containing the amber or ochre codon at an internal position.

Suga and co-workers [29–31] have reported aminoacylation of tRNAs with ribozymes. They screened a ribozyme from random combinatorial pools and aminoacylated tRNA with specific N-biotinylated amino acids including e-biotinylated lysine (22, Figure 3c) [32**]. The aminoacylating ribozyme interacts with both the 3’ terminal and the anticodon loop of the target tRNA. The ribozyme strategy is not only useful for labeling of proteins with biotin, but is potentially applicable to customized ribozymes that aminoacylate desired non-natural amino acids to specific tRNAs.

**Extension of protein functions with non-natural amino acids**

By introducing non-natural amino acids with specialty functions, protein functions will be widely expanded as described below. Dougherty’s group [33] have been investigating ion channels by using the amber suppression method in *Xenopus* oocytes. Caged amino acids, S-0-nitrobenzyl-cystein (23) and O-0-nitrobenzyl-tyrosine (17) have been introduced into nicotinic acetylcholine receptor and KIR2.1 channel [34,35]. The resulting receptors recovered their activity when the o-nitrobenzyl group was removed by UV irradiation. The receptor activity could be controlled also by introducing agonist-tethered amino acids (24) [36*]. Four kinds of O-(trimethylammoniumalkyl)tyrosine have been synthesized and incorporated at three different positions of the acetylcholine receptor. Because the trimethylammoniumalkyl group works as an agonist for the receptor, the resulting receptor–agonist conjugates showed constitutive activity in the absence of acetylcholine.
Hecht and co-workers reported position-specific proteolysis by introducing allylglycine (25). Proteins containing allylglycine have been cleaved by treatment with iodine through a presumed iodolactone intermediate. Introduction of allylglycine to a trypsinogen [37*] or a trypsin inhibitor (ecotin) [38] successfully controlled the trypsin activity. Non-natural amino acids with sugar moieties were also investigated by their group [39].
Schultz and co-workers have been using their in vivo method to investigate protein–protein interactions in living cells. They incorporated p-benzoylphenylalanine (7) into a Phe52 position of dimeric glutathione S-transferase. Subsequent UV irradiation of the mutant resulted in forming a covalently linked homodimer [24**]. This photocrosslinking method will be useful for mapping protein–protein interactions in living cells.

Reversible photoregulation of proteins has been achieved by incorporating a photosensitive group such as azobenzene into specific positions. We have introduced p-phenylazophenylalanine (26) into various positions of a horseradish peroxidase [40]. The F179-mutant and F162-mutant showed reversible photoregulation of the enzyme activity. Incorporation of phenylazophenylalanine into a camel anti-lysozyme antibody or into a DNA-binding protein also induced reversible photosresponse of ligand binding (Hohsaka T, Sisido M, unpublished data).

Position-specific incorporation of fluorescent groups is an important step toward comprehensive proteome analysis. Because of the size limitation of non-natural amino acids that are accepted by the translation machinery, relatively small but long-wavelength emitting fluorescent amino acids have been searched for. Amino acids carrying coumarin (27) [41] and anthraniloyl (28) [42] group are in such class of amino acids. They can be incorporated into proteins in relatively high yields and their fluorescence intensity and wavelength are sensitive to the ligand binding. Larger fluorescent groups such as fluorescein had never been introduced into proteins. Very recently, however, a green-fluorescent BODIPY FL-labeled p-aminophenylalanine (29) was designed, synthesized and successfully incorporated into proteins through an E. coli cell-free translation system (Hohsaka T, Sisido M, unpublished data). This finding will open a way for practical application of the non-natural mutagenesis for comprehensive analysis of protein–protein and protein–nucleic acid interactions.

Conclusions

Introduction of non-natural amino acids into proteins is a potential tool that is widely applicable to genomic, proteomic and cellular biological researches. Position-specific incorporation of probes such as biotin and fluorescent groups will be useful for high-throughput analyses of protein–protein and protein–nucleic-acid interactions. The possibility is further expanded when multiple non-natural amino acids are incorporated by using the four-base codon strategy. For these purposes, cell-free protein synthesis is advantageous for its rapid and simple processing. In vivo synthesis of non-natural mutants using artificial aaRS is essential for large-scale expression and for analyses of protein–protein interactions in living cells. Large-scale synthesis is particularly required for utilization of the non-natural mutants as nano-sized chemical devices, such as biosensors, biophotoelectronic devices and protein microarrays.

Acknowledgements

Works on four-base codons was supported by the Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Science, Sports and Culture, Japan (No. 11 102 003) and by Industrial Technology Research Grant Program in ‘00 from the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest

Conclusions

Introduction of non-natural amino acids into proteins is a potential tool that is widely applicable to genomic, proteomic and cellular biological researches. Position-specific incorporation of probes such as biotin and fluorescent groups will be useful for high-throughput analyses of protein–protein and protein–nucleic-acid interactions. The possibility is further expanded when multiple non-natural amino acids are incorporated by using the four-base codon strategy. For these purposes, cell-free protein synthesis is advantageous for its rapid and simple processing. In vivo synthesis of non-natural mutants using artificial aaRS is essential for large-scale expression and for analyses of protein–protein interactions in living cells. Large-scale synthesis is particularly required for utilization of the non-natural mutants as nano-sized chemical devices, such as biosensors, biophotoelectronic devices and protein microarrays.

Acknowledgements

Works on four-base codons was supported by the Grant-in-Aid for Specially
Promoted Research from the Ministry of Education, Science, Sports and
Culture, Japan (No. 11 102 003) and by Industrial Technology Research
Grant Program in ‘00 from the New Energy and Industrial Technology
Development Organization (NEDO) of Japan.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest

Conclusions

Introduction of non-natural amino acids into proteins is a potential tool that is widely applicable to genomic, proteomic and cellular biological researches. Position-specific incorporation of probes such as biotin and fluorescent groups will be useful for high-throughput analyses of protein–protein and protein–nucleic-acid interactions. The possibility is further expanded when multiple non-natural amino acids are incorporated by using the four-base codon strategy. For these purposes, cell-free protein synthesis is advantageous for its rapid and simple processing. In vivo synthesis of non-natural mutants using artificial aaRS is essential for large-scale expression and for analyses of protein–protein interactions in living cells. Large-scale synthesis is particularly required for utilization of the non-natural mutants as nano-sized chemical devices, such as biosensors, biophotoelectronic devices and protein microarrays.

Acknowledgements

Works on four-base codons was supported by the Grant-in-Aid for Specially
Promoted Research from the Ministry of Education, Science, Sports and
Culture, Japan (No. 11 102 003) and by Industrial Technology Research
Grant Program in ‘00 from the New Energy and Industrial Technology
Development Organization (NEDO) of Japan.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest

Conclusions

Introduction of non-natural amino acids into proteins is a potential tool that is widely applicable to genomic, proteomic and cellular biological researches. Position-specific incorporation of probes such as biotin and fluorescent groups will be useful for high-throughput analyses of protein–protein and protein–nucleic-acid interactions. The possibility is further expanded when multiple non-natural amino acids are incorporated by using the four-base codon strategy. For these purposes, cell-free protein synthesis is advantageous for its rapid and simple processing. In vivo synthesis of non-natural mutants using artificial aaRS is essential for large-scale expression and for analyses of protein–protein interactions in living cells. Large-scale synthesis is particularly required for utilization of the non-natural mutants as nano-sized chemical devices, such as biosensors, biophotoelectronic devices and protein microarrays.

Acknowledgements

Works on four-base codons was supported by the Grant-in-Aid for Specially
Promoted Research from the Ministry of Education, Science, Sports and
Culture, Japan (No. 11 102 003) and by Industrial Technology Research
Grant Program in ‘00 from the New Energy and Industrial Technology
Development Organization (NEDO) of Japan.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:
• of special interest
•• of outstanding interest


An azidohomolalnine is incorporated into a protein as a methionine analogue in E. coli. The resulting protein is ligated with phosphine-labeled peptides.


A breakthrough report describing the incorporation of a 21st amino acid into a protein in a growing E. coli. A mutated TyrRS that exclusively recognizes an amber suppressor tRNA and O-methyl-tyrosine is developed, and expressed in an E. coli together with a DHFR gene containing an amber codon. The produced DHFR contains O-methyl-tyrosine at the directed position.


This report describes a novel mutated aARS that catalyzes an aminoacylation of tRNA with p-azido-phenylalanine. The result raises the possibility that non-natural amino acids with large side chains may be accepted by ARSs after appropriate mutations.


This is a review of expanded translation systems for introducing non-natural amino acids into proteins. Recent work of author's group is described.


A mutated TyRS catalyzes an aminoacylation of a nonsense suppressor tRNA with a non-natural amino acid, m-iodotyrosine. The resulting aminoacyl-tRNA introduces iodotyrosine into a desired position of proteins in a wheat germ translation system. This system, enabling the introduction of iodine into specific positions of proteins, will be useful for X-ray structure analyses.


This paper describes the ribozyme-mediated aminoaacylation of tRNAs. Addition of a tRNA recognition domain selected from a random pool to a previously developed acyl-transfer ribozyme gives a novel ribozyme that aminoaacylates specific tRNAs with α- or β-biotinylated amino acids.


This paper describes a useful strategy to cleave a polypeptide chain internally through the ionization of allylglycine incorporated at desired positions of proteins.


Incorporation of agonist-linked amino acids into an acetylcholine receptor gives a constitutively active receptor. The result provides a useful tool for probing structures of ligand-binding sites of proteins.

