Bull. Chem. Soc. Jpn., 72, 1409—1425 (1999)

Accounts

Extension of Protein Functions by the Incorporation of Nonnatural Amino Acids

Masahiko Sisido* and Takahiro Hohsaka

Department of Bioscience and Biotechnology, Faculty of Engineering, Okayama University, 3-1-1 Tsushimanaka, Okayama 700-8530

(Received February 25, 1999)

Site-specific mutagenesis of proteins was extended to incorporate nonnatural amino acids through in vitro cell-free translation or through in vivo translation in the living cells. To this end, a variety of nonnatural amino acids were synthesized and linked to tRNA by chemical aminoacylation. Further, unique codon/anticodon pairs were derived, such as the 4-base codon/anticodon pairs that are orthogonal to the conventional 3-base codon/anticodon pairs. By the extension of amino acids and the genetic codes, site-directed incorporation of specialty amino acids into a variety of proteins has been achieved. Amino acids that carry large aromatic side groups with straight configurations, such as 2-anthrylalanine and 2-anthraquinonylalanine were successfully introduced into proteins and the CGGG or AGGU 4-base codons were used to assign their positions.

Amino acids carrying photocleavable side groups, a spin probe, fluorescent probes, and other functional groups have been introduced into proteins. The specialty proteins were shown to be powerful tools for biochemical research and are expected to be novel molecules with designed arrangements of functional groups on the protein framework.

Site-directed incorporation of nonnatural amino acids into proteins is a versatile technique by which a variety of functional groups can be built into proteins with a designed spatial arrangement.[1—6] The nonnatural amino acids were incorporated through chemical misaminoacylation of tRNA, followed by the protein synthesis in the cell-free (in vitro) biosynthesizing system or in living cells (in vivo). When the basic strategy was first reported by Schultz[1,2] and Chamberlin[3] independently in 1989, two major problems remained to be solved. First, since only a limited number of nonnatural amino acids were available at that time, information on the amino acid structures that are allowed in the protein biosynthesizing system was lacking. Second, since the codon/anticodon pairs that are assignable to nonnatural amino acids were practically limited only to the amber nonsense codon (TAG), multiple incorporation of different nonnatural amino acids into a single protein was not possible. Furthermore, since the translation of the nonsense codon by the corresponding suppressor tRNA must compete with the termination by the binding of a release factor, the efficiency of the incorporation cannot be very high.

In the past ten years, the authors have been extending the technique by synthesizing a variety of aromatic nonnatural amino acids. We have clarified the type of nonnatural amino acids that can be incorporated into proteins through in vitro protein synthesizing systems from *E. coli* and other cells. Also, we have found that several 4-base codon/anticodon pairs work concomitantly with the conventional 3-base codon/anticodon pairs and direct nonnatural amino acids to specific positions of proteins. The details of the extension of amino acids and the extension of the genetic codes will be described in the first half of this account.

So far, the technique has been finding applications in probing biochemical functions, photoswitching enzyme activities, and building electron-transfer paths on proteins. These examples of the extension of protein functions are the second topic of this account. The incorporation of nonnatural amino acids into proteins which is sometimes referred as nonnatural amino acid mutagenesis or simply nonnatural mutagenesis, has been also conducted in the living cells. This has been made possible by co-injecting aminoacyl tRNAs carrying nonnatural amino acids and the target mRNA into a living cell in which the target protein is expressed on the membrane surface. The extension to the in vivo system is described as the last topic.

Brief Summary of the Technique. The nonnatural amino acid mutagenesis is based on the earlier finding by Chapeville[7] in 1962. During the translation of the genetic codes on mRNA to the amino acid sequence through ribosomal systems, the triplet codons discriminate only the anticodons of tRNA and do not recognize the amino acids that are linked to the tRNA. This has been elegantly demonstrated

by the chemical reduction of a cysteine unit on $tRNA^{Cys}$ to give an Ala unit on $tRNA^{Cys}$. The misaminoacylated tRNA, i.e., Ala-$tRNA^{Cys}$, when added to the in vitro biosynthesizing system, was found to incorporate Ala units in place of Cys units in the polypeptide chain produced. In this context, if a nonnatural amino acid was linked to a tRNA by any means, the amino acid will be introduced into proteins at the position directed by the anticodon of the tRNA. The key steps for the nonnatural mutagenesis are summarized as follows (Fig. 1).

(1) Synthesis of a nonnatural amino acid carrying a specialty side group.
(2) Chemical or biochemical misaminoacylation of the tRNA with the nonnatural amino acid.
(3) Assignment of a unique codon/anticodon pair to the nonnatural amino acid and introduction of the unique codon to the target DNA and insertion of the unique anticodon to the anticodon loop of the tRNA.
(4) In vitro protein biosynthesis in a cell-free extract of *E. coli* or other cell systems in the presence of the misaminoacylated tRNA and the target mRNA.
(5) Isolation and identification of the nonnatural mutant.

In the above procedure, in vitro protein synthesizing systems have been employed for their versatility in preparing

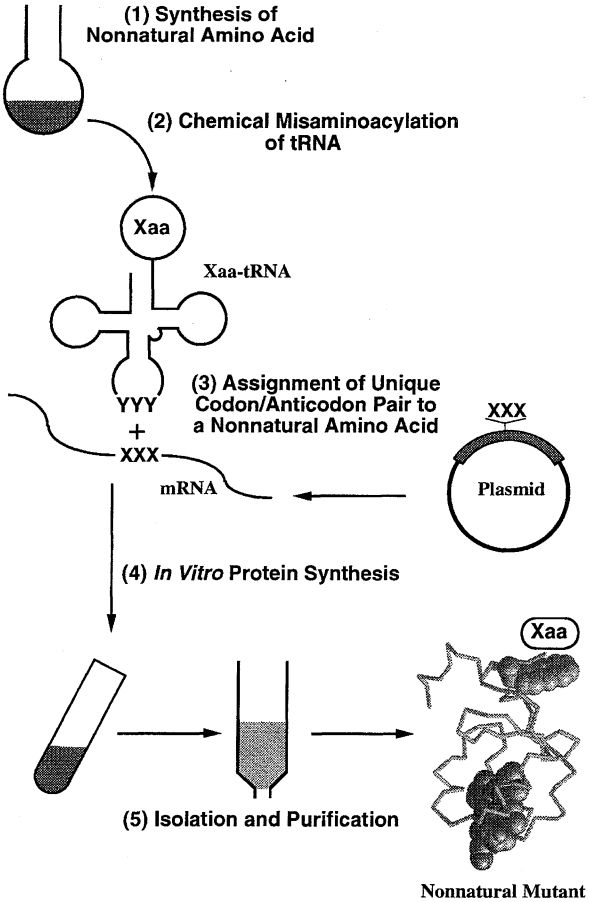


Fig. 1. Key steps for the site-directed incorporation of nonnatural amino acids into proteins through in vitro protein synthesizing system.

a wide variety of artificial proteins.[8,9] One of the problems, however, of the in vitro systems is very small protein yields. In a typical experiment of the authors' group, purified proteins are synthesized in one to a few ten μg quantity in a 10 to 100 μL-scale synthesis. At the present state of the art, a few milligram quantity of purified proteins may be the upper limit.[10] Although a variety of experiments including spectroscopic and enzymatic measurements can be done with the μg quantity of the mutants, further efforts to increase the protein production is clearly needed. These approaches have been reported recently.[11]

Nonnatural Amino Acids with Specialty Side Groups. Since the most advantageous point of the nonnatural mutagenesis over the conventional protein engineering is a potential extension of protein functions beyond the conventional biochemical framework, the availability of a wide variety of nonnatural amino acids is essential. The importance of nonnatural amino acids has been recognized earlier in the combinatorial library of synthetic peptides.[12] Combination of nonnatural amino acids together with the naturally-occurring ones in the peptide synthesis extends the chemical and biochemical diversity far beyond the conventional biological framework and thus provides new potentiality in synthetic peptide chemistry. A wide variety of nonnatural amino acids have been reported and many of those are already commercially available. A database for nonnatural amino acids that are commercially available has been provided by Konishi.[13]

We have been synthesizing nonnatural amino acids carrying aromatic side groups.[14] Typical examples are listed in Fig. 2. Those include fluorescent probes (naphthylalanines, pyrenylalanines, anthrylalanines), electron donors (*p*-dimethylaminophenylalanine, carbazolylalanines, dimethoxyphenylalanine), electron acceptors (*p*-nitrophenylalanine, dinitrophenylalanine, anthraquinonylalanine), electron mediators (anthraquinonylalanine, ferrocenylalanine), and metal ligands (bipyridylylalanines).

Throughout the amino acids listed in Fig. 2, the aromatic groups are linked by a single methylene unit to the α-carbon as in the case of phenylalanine, tyrosine, and tryptophan. The simple structure is adopted primarily because of the similarity to the naturally-occurring amino acids. The shortest spacer is also advantageous to keep the side-chain aromatic groups in a constrained orientation after the amino acids are incorporated into the protein main chain. Most of the amino acids have been synthesized from arylmethyl halides through coupling with diethyl acetamidomalonate, followed by ester hydrolysis and decarboxylation. The racemic *N*-acetylamino acids are usually deacetylated with acylase to give optically pure L-amino acids. In the case of 2-anthraquinonylalanine,[13f] however, the pure enantiomer was difficult to isolate, because of facile racemization due to the strong electron-withdrawing effect of the anthraquinonyl group. As far as the biosynthetic incorporation into proteins is concerned, the optical purity may not be important, since D-amino acids are usually rejected by ribosomal systems, probably because of the inaccessibility of D-amino acid-tRNA to the A site of ribosome. The anthraquinonylalanine unit after being incorporated into

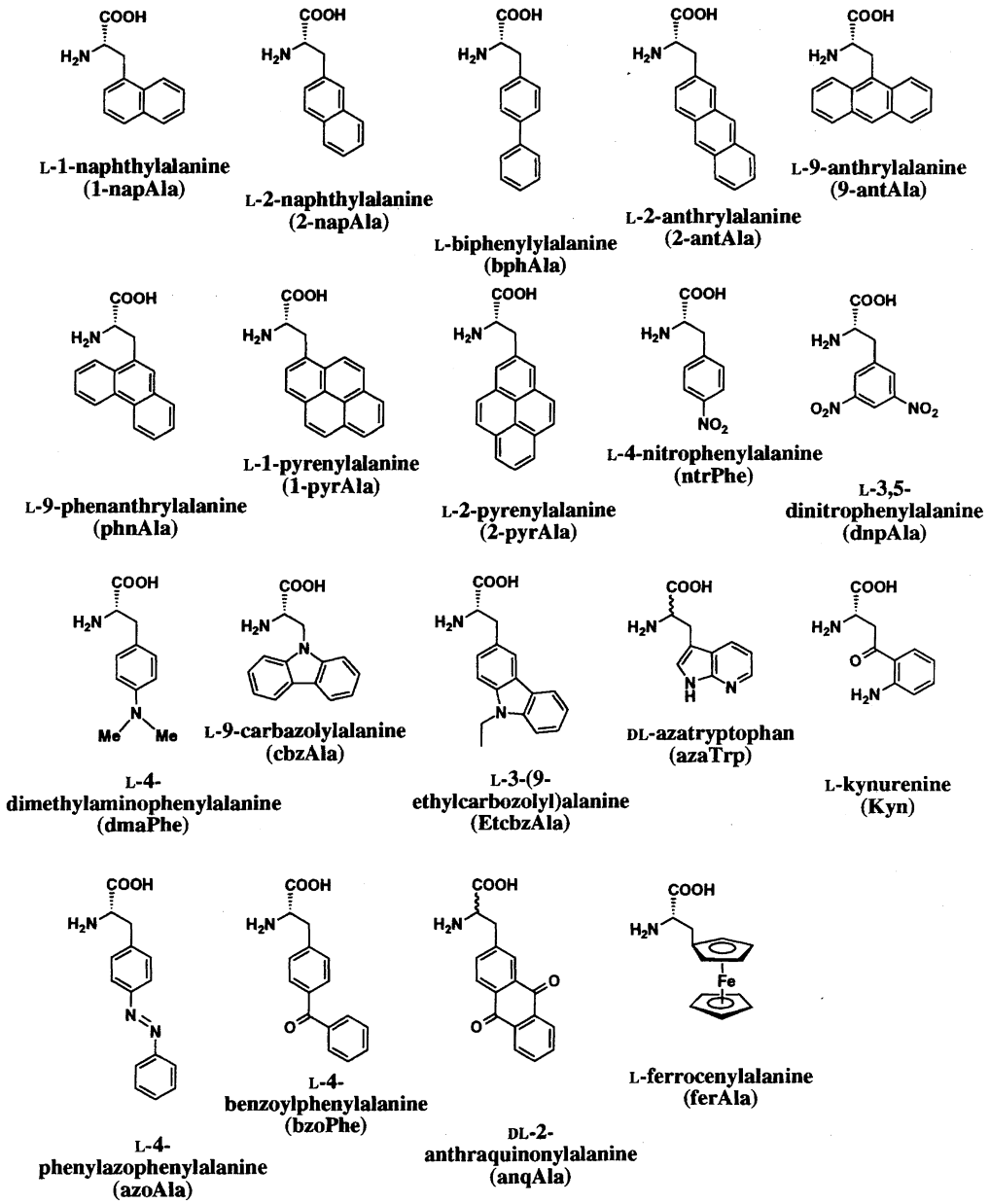


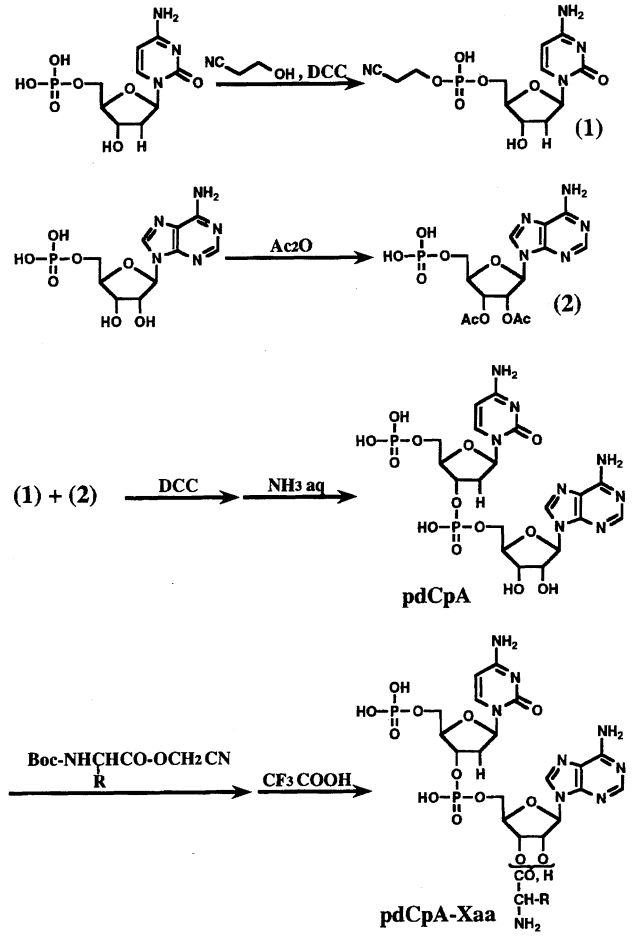
Fig. 2. Nonnatural amino acids with aromatic side groups.

a polypeptide chain does not racemize even after several redox cycles.

Other types of nonnatural amino acids, such as those with long alkyl groups, nucleobases, a crown ether, and even a porphyrin ring on the side chain, have been reported. Most of them can be linked to the 3′-terminal of tRNA, but only a few of them were incorporated into proteins as will be described later.

Chemical and Biochemical Misaminoacylation of tRNAs. Aminoacylation of tRNA with nonnatural amino acids is a key step of the nonnatural mutagenesis. Since direct aminoacylation of tRNA at the 3′-*O* position of the terminal adenosine unit seems intractable, an alternative pathway using enzymic ligation of 3′-*O*-aminoacylated pCpA to a truncated tRNA that lacks a terminal pCpA unit, has been proposed by Hecht and co-workers.[15,16] Chamberlin[6] and Schultz[17] showed that the C unit of the dinucleotide pCpA, could be replaced by a deoxy C (dC) unit without reducing the ligation efficiency. Although this made the synthesis of the dinucleotide much simpler, the synthesis of pdCpA and its regioselective aminoacylation is still a difficult step, particularly for biochemists. The synthetic route we are currently employing is shown in Scheme 1.

The route consists of the coupling of minimum-protected pA and pdC units followed by the deprotection and the HPLC purification to give free pdCpA. The latter was linked with Boc-derivative of nonnatural amino acids and the Boc group was deprotected with trifluoroacetic acid. The above route may be most convenient to obtain a few ten to hundred μg quantity. For synthesis of larger amounts of aminoacyl-pdCpA, however, other routes will be more appropriate. Hecht and co-workers[18] recently proposed an improved route using



Scheme 1. Synthetic route for small-scale synthesis of aminoacyl-pdCpA carrying a nonnatural amino acid.

a allyl-protected amino acid followed by the deprotection with iodide.

The aminoacyl-pdCpA is ligated with the truncated tRNA that lacks the 3′ terminal pCpA unit by using T4 RNA ligase. The truncated tRNA can be obtained through T7 RNA polymerase-catalyzed transcription of the corresponding DNA. The full-length tRNA aminoacylated with a nonnatural amino acid at the terminal 3′-*O* position is usually used without further purification in the next in vitro synthesis.

In the biochemical system, a specific amino acid is linked to the cognate tRNA by an enzyme, aminoacyl tRNA synthetase (ARS). The fidelity of ARS in linking proper amino acid to the corresponding tRNA is extremely high. Furthermore, the ARS does proof-reading after charging amino acids. If an incorrect amino acid is charged to the tRNA, the misaminoacylated tRNA will be decomposed by the ARS. However, the high fidelity may be expected only for the 20 naturally-occurring amino acids. Indeed, some nonnatural amino acids have been incorporated into proteins in an *E. coli* strain cultured on the growth medium that contained nonnatural amino acids. For example, Tirrell and co-workers[19] cultured a strain of *E. coli* that requires methionine for growth (Met-axotroph), in the presence of selenomethionine in place of methionine in the growth medium. They found that selenomethionine is successfully incorporated into their designed polypeptides in place of methionine (Fig. 3). The successful incorporation indicates that selenomethionine is charged to $tRNA^{Met}$ by MetRS. *p*-Fluorophenylalanine and thienylalanine have been also incorporated into the polypeptide in place of phenylalanine by using Phe-axotroph.[20,21] Similar experiments on the misaminoacylation with amino acids that are close analogs of one of the naturally-occurring ones, have been also reported before.[22]

The above experiments indicate that the fidelity of ARS is limited only to the 20 naturally-occurring amino acids; if close analogs are present, some tRNA may be misaminoacylated by the endogenous ARSs. Evidently, the misaminoacylation may take place only for nonnatural amino acids that have close structural similarity to one of the 20 naturally-occurring ones. In our previous experiments,[23] for example, the incorporation efficiencies of L-1-pyrenylalanine and L-9-anthrylalanine into β-galactosidase through *E. coli* in vivo system were very small.

In the above in vivo syntheses, nonnatural amino acids, when they are mischarged to tRNAs, will be incorporated into multiple positions that are assigned by the anticodon of the tRNA. In this sense, site-directed incorporation in the in vivo synthesis is not possible. The large-scale production of nonnatural mutants in the in vivo system is, however, definitely advantageous over the in vitro system. Tirrell and coworkers[19–21] have been doing the large-scale production of their designed polypeptides as new biomolecular materials.

Biochemical misaminoacylation of tRNAs with nonnat-

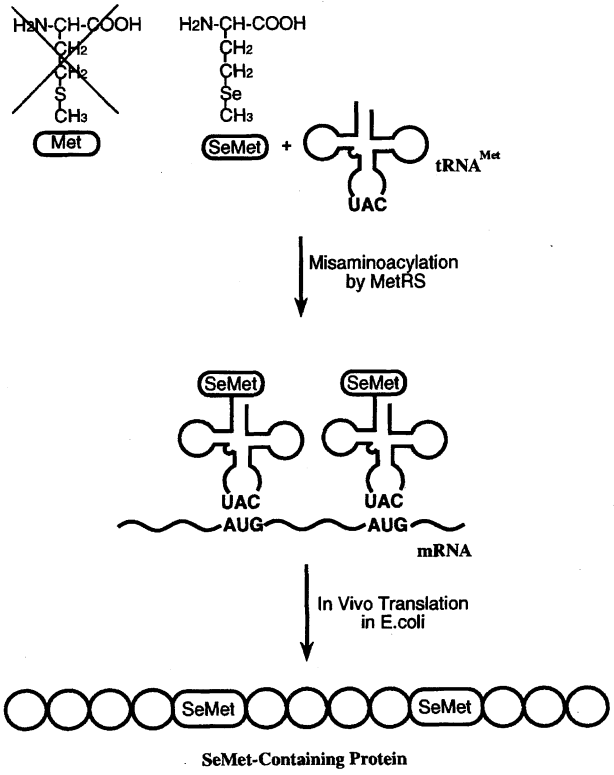


Fig. 3. In vivo incorporation of selenomethionine into proteins by using *E. coli* methionine axotroph.

ural amino acids will become possible if one has mutant ARSs that are active to nonnatural amino acids. Hennecke and co-workers[24—26] prepared a mutant of *E. coli* phenylalanine-tRNA synthetase (PheRS) replacing 294Ala by Gly. The point mutation opened a pocket at the *para*-position of the phenylalanine substrate. They observed that the mutant ARS actually linked *p*-chlorophenylalanine to the $tRNA^{Phe}$ and the nonnatural amino acid was incorporated into luciferase. Since their mutant ARS is also active to phenylalanine, it cannot be used for the unique incorporation of *p*-chlorophenylalanine in the in vivo synthesis. Schultz and co-workers[27] made an attempt to use a catalytic antibody for the aminoacylation. They raised an antibody against a neutral phosphonate diester transition-state analog for the 3′-terminal of aminoacyl tRNA. Although the antibody could not recognize any unique nonnatural amino acid, this is the first attempt to obtain artificial ARSs.

If an artificial ARS that recognizes a nonnatural amino acid, but is orthogonal to the naturally-occurring ones, were obtained, it will serve both the in vitro and in vivo biosynthesis of nonnatural mutants. In particular, the in vivo protein synthesis with the artificial ARS is a challenging idea, because it will mean to create a new living organism that is outside the paradigm of the present biochemical system. The extension of the in vivo system toward this direction will be described later in this account.

In Vitro Biosynthesis in the Presence of Chemically Misaminoacylated Yeast $tRNA^{Phe}_{CCCG}$. Typical steps for the preparation of nonnatural mutants currently performed in the authors laboratory are as follows. Yeast $tRNA^{Phe}$ was chemically misaminoacylated with a nonnatural amino acid, typically *p*-nitrophenylalanine, and it was added to the S30 *E. coli* in vitro biosynthesizing system together with the target mRNA. The S30 lysate is obtained commercially. The plasmid that encodes the target DNA is shown in Fig. 4.[28,29] The DNA contains a T7 promoter, a T7 tag, a streptavidin that includes a CGGG 4-base codon at the position of interest, and the His tag at the C-terminal. The T7 tag is a dodecapeptide that specifically binds to the T7-tag antibody, and works to report the yield of the protein by the antibody binding, irrespective of the chain length. The His tag is a histidine hexamer that shows special affinity to a Ni complex-linked gel, and serves to purify the protein that was successfully elongated down to the C-terminal.

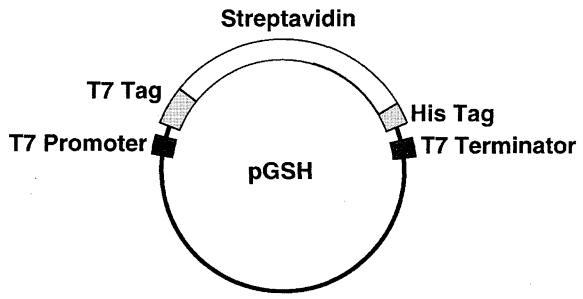


Fig. 4. The plasmid encoding mutant streptavidin incorporated with a nonnatural amino acid at the 83rd position. A T7 tag (Met–Ala–Ser–Met–Thr–Gly–Gly–Gln–Gln–Met–Gly–Thr) is linked at the N-terminal and a histidine hexamer is linked at the C-terminal.

The codon/anticodon pair that directs the position of the nonnatural amino acid was a CGGG/CCCG pair. The CGGG codon is introduced at the 83rd position, for example, of the streptavidin mRNA and the CCCG anticodon at the anticodon loop of the tRNA, respectively. The details of the 4-base codon/anticodon pair will be described in the next section. Figure 5 shows results of Western blotting of the reaction mixture of *E. coli* in vitro translation system that contained *p*-nitrophenylalanyl $tRNA_{CCCG}$ and mRNA encoding streptavidin including a CGGG 4-base codon at the 83rd position.

The full-length streptavidin was synthesized only in the presence of the aminoacyl $tRNA_{CCCG}$, and the protein is produced neither in the absence of the $tRNA_{CCCG}$ nor in the presence of non-acylated $tRNA_{CCCG}$. The former result indicates that the protein synthesis is really directed by the 4-base codon/anticodon pair and the latter result indicates that the yeast tRNA cannot be aminoacylated by the endogenous ARS's in the S30 system. If the aminoacylation of yeast $tRNA^{Phe}$ took place, phenylalanine or some other amino acids would be incorporated into the 83rd position and the full-length streptavidin would be obtained. No full-length proteins in the absence of nitrophenylalanyl-tRNA indicates that these processes did not take place in the biosynthesizing system. The production of full-length proteins in the presence of nitrophenylalanyl-tRNA indicates that the incorporation of the nonnatural amino acid is really achieved in the *E. coli* S30 in vitro system.

Extension of the Genetic Codes. One of the essential steps for the incorporation of nonnatural amino acids is to find special codon/anticodon pairs that can be assigned exclusively to a nonnatural amino acid. The codon/anticodon pairs must be orthogonal to the conventional triplet codons for the

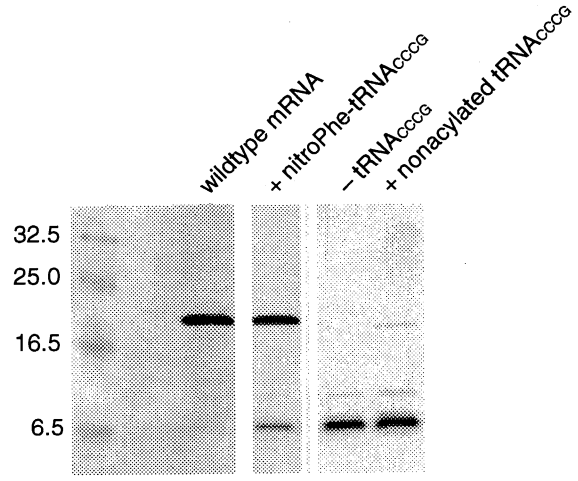


Fig. 5. Western blot analysis of in vitro translation in the *E. coli* S30 system. The reaction mixture was applied to the SDS-PAGE and transformed to PVDF membrane, followed by incubation with *anti*-T7 tag antibody. The bands were visualized with alkaliphosphatase-labeled *anti*-mouse IgG and NBT/BCIP.

naturally-occurring amino acids. The amber nonsense codon (UAG), has been employed by Schultz and Chamberlin.[1–6] In this case, amber suppressor tRNA that has a CUA anticodon and is aminoacylated with a nonnatural amino acid, is used to deliver the amino acid into proteins. When the amber codon appeared on the mRNA, it will be translated by the amber suppressor tRNA to the nonnatural amino acid. Otherwise the amber codon is bound to a protein called a releasing factor, resulting in the termination of the protein synthesis.

The nonsense codon/nonsense suppressor tRNA strategy is basically a unique method to incorporate nonnatural amino acids at specific positions. However, some drawbacks are reported. First, the nonsense codon that can be assigned to nonnatural amino acids is practically limited only to the amber codon. Therefore, only a single type of nonnatural amino acid can be incorporated into a protein. Further, since the amber suppression must compete with the releasing factor, the incorporation efficiency cannot reach 100%. Finally, depending on the codon context, a readthrough of the amber codon, i.e., the continuation of protein synthesis without incorporating nonnatural amino acid, may occur.

We reported that some 4-base codon/anticodon pairs, such as AGGU/ACCU and CGGG/CCCG, can be used to assign nonnatural amino acids in the *E. coli* in vitro system.[28,29] The principle of the 4-base codon/anticodon strategy is illustrated in Fig. 6. When the sequence CGGG is successfully read as the 4-base codon by the $tRNA_{CCCG}$, the protein synthesis continues to the end and a full-length protein containing a nonnatural amino acid will be obtained. If the CGGG sequence is read as the CGG 3-base codon by the endogenous $tRNA^{Arg}_{CCG}$, the reading frame is shifted backward by 1 base, resulting in the encounter of a nonsense codon UAA. Thus, the unsuccessful translation as the 3-base codon results in the truncation of the protein synthesis. The actual performance of the 4-base codon/anticodon strategy is demonstrated in Fig. 5.

When the $tRNA_{CCCG}$ with *p*-nitrophenylalanine is added, the major product is a protein of 19 kDa, that corresponds to the full-length streptavidin containing *p*-nitrophenylalanine. The minor product at 6.5 kDa corresponds to the truncated protein just after the CGGG 4-base codon. The ratio of yields of the full-length protein to the truncated one depends on the codons, structure of nonnatural amino acids, and other reaction conditions. The incorporation of nonnatural amino acid at the specified position has been directly confirmed by the HPLC identification of trypsin-digested peptide fragments of the mutant streptavidin that contained a fluorescent amino acid, ε-nitrobenzoxadiazolyl-L-lysine.[28b]

Other 4-base codons are also used to assign nonnatural amino acids, i.e., CGGN and AGGN (N indicates any of the four bases). They are derived from the minor codons of arginine (AGG and CGG) that appear less frequently in the *E. coli* system. Therefore, the competition between the 4-base codon and the 3-base codon is in favor of the former, especially when an excess amount of $tRNA_{CCCG}$ is added. Among the CGGN- and AGGN-type 4-base codons, CGGG was most efficient. The highest translation efficiency of the CGGG codon in the *E. coli* in vitro system was 70% when 2-naphthylalanine or biphenylylalanine was linked to the $tRNA_{CCCG}$. Higher efficiency was achieved in the rabbit reticulocyte in vitro system. In the latter system, nearly 100% translation efficiency has been observed for 2-naphthylalanine.

Another potential advantage of the 4-base codon/anticodon strategy is the possibility of multiple incorporation of different nonnatural amino acids into a single protein. The AGGU codon does not recognize $tRNA_{CCCG}$ and the CGGG codon does not recognize $tRNA_{ACCU}$. Further, both are practically orthogonal to the conventional 3-base codons. The absence of the cross reactivity ensures that two different nonnatural amino acids can be directed to two independent positions, respectively, in a single protein. The potential multiple incorporation of different amino acids will expand protein

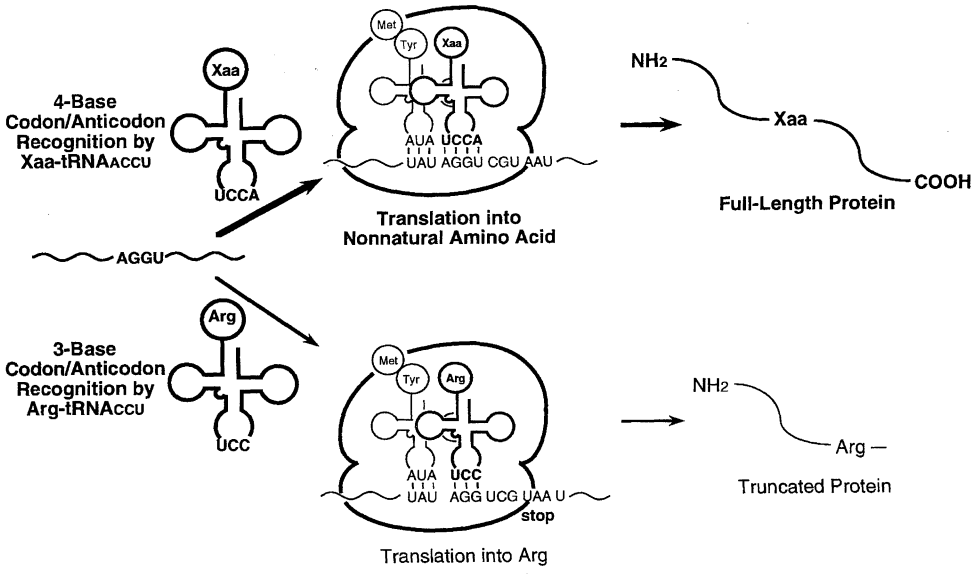


Fig. 6. Principle of the 4-base codon/anticodon strategy. See text for detail.

functions, for example, for the charge separations between electron donors and acceptors and for the arrangements of nucleic bases with a designed order.

An alternative way to find special codons that are assignable to nonnatural amino acids is to incorporate unusual nucleobases into RNAs. Bain et al.[30] proposed that *iso*G and *iso*C may work as orthogonal nucleobases to four other bases (A, U, G, and C). They showed that an unusual codon *iso*C–A–G was read by the anticodon C–U–*iso*G and successfully translated to iodotyrosine. Although the use of unusual nucleobases is a fascinating idea, the unavailability of DNA and RNA polymerase restricts its practical application.

The fourth approach to find codons for nonnatural amino acids is to inactivate some of minor tRNAs in the in vitro system. Kanda et al.[31] reported that a specific tRNA can be inactivated by an antisense oligoDNA in the *E. coli* system. Further, they showed that the tRNA hybridized with the antisense oligoDNA was specifically digested by RNase H. Since some of the conventional codons are less frequently used and the amounts of the corresponding tRNAs are marginal in the *E. coli* systems, the antisense inactivation or the "knocking-out" of the minor tRNA will be a promising approach to make a space in the codon table.

Structure of Amino Acids That are Accepted by *E. coli* Ribosome. Not all the nonnatural amino acids shown in Fig. 1, for example, can be incorporated into proteins. The types of nonnatural amino acids that can be incorporated into proteins may be determined mainly by the accessibility of the corresponding aminoacyl tRNAs to the ribosome A site, as illustrated in Fig. 7. This has been suggested by experiments using puromycin analogs carrying a variety of nonnatural amino acids in place of *O*-methyltyrosine at the 3′ amino group.[32] The inhibitory effects of various puromycin analogs in the *E. coli* in vitro protein synthesizing system are compared in Fig. 8.

The puromycin analog carrying L-2-naphthylalanine moderately inhibited the protein synthesis. This indicates that the analog is bound to the A site of ribosome. In contrast, the

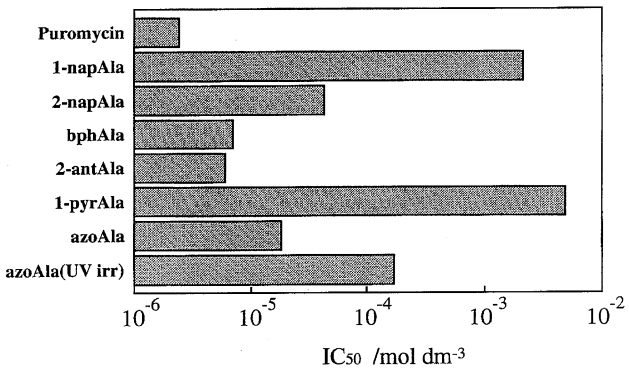


Fig. 8. Efficiencies of puromycin analogs as evaluated by the concentrations required to 50% inhibition of protein synthesis in *E. coli* in vitro system.

analog carrying L-1-naphthylalanine showed little inhibitory effect, indicating the inaccessibility to the A site. The results of the inhibitory effects for various puromycin analogs may be summarized as in Fig. 9. The benzene rings indicated as X block the binding of the puromycin analogs to the A site of ribosome, whereas those indicated as A do not. It is interesting to note that the analog with a *trans*-azobenzene moiety does not contain X ring, but the analog with a *cis*

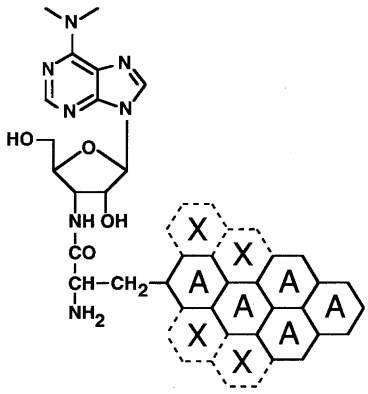


Fig. 9. Allowed (A) and precluded (X) regions of aromatic amino acids for the accessibility of puromycin analogs to *E. coli* ribosomal A site.

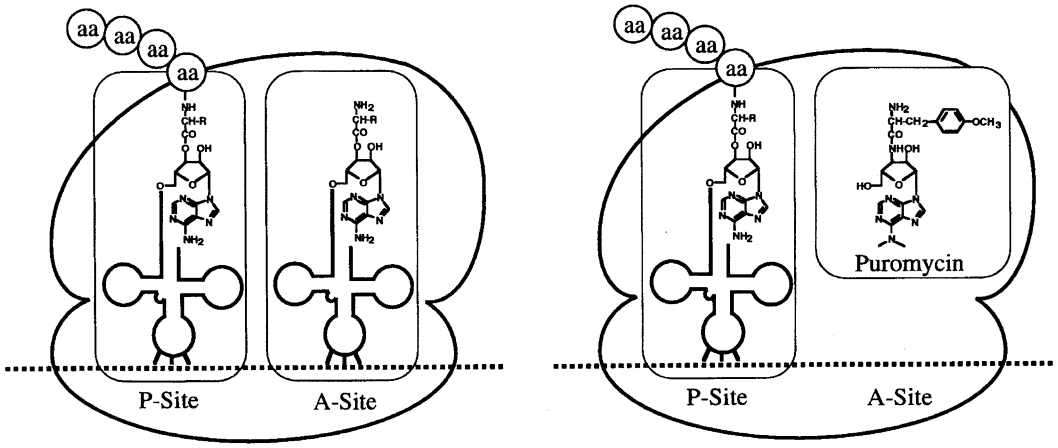


Fig. 7. A rough sketch of the ribosomal system with a P site bound with a peptidyl tRNA and an A site bound with an aminoacyl tRNA. When the A site was bound with puromycin or its analogs, the peptide synthesis will be inhibited.

group contains an X ring. Indeed, the *trans* form binds to the A site and inhibits protein synthesis, but the *cis* form does not bind to the A site. Since the *trans* and *cis* form are interconvertible by photoirradiation, the difference in the inhibitory effects may open a way to the photoregulation of the protein synthesis.

Similar amino acid selectivity was found in the total incorporation efficiency of nonnatural amino acids that are linked to tRNA with a CCCG 4-base anticodon in the *E. coli* in vitro system.[29] The incorporation efficiencies of 19 different nonnatural amino acids are listed in Fig. 10. The incorporation efficiencies of various aminoacyl tRNAs are essentially parallel to the inhibitory efficiencies of various puromycin analogs and follow the same rule as shown in Fig. 9. The parallel relationship indicates that the incorporation efficiency is mainly governed by the accessibility of the aminoacyl tRNAs to the A site of ribosome. The activity of peptidyl transferase to nonnatural amino acids and the accessibility of peptidyl tRNA to the P site, and other factors may also affect the incorporation efficiency, but these factors are not primarily important in determining the amino acid selectivity in the ribosomal system.

Incorporation of other types of amino acids has been attempted. Amino acids with long spacer chains between the α-carbon and 1-pyrenyl group have been synthesized, and their incorporation efficiencies were tested in the *E. coli* in vitro system (Fig. 11).[33] As seen from the figure, moderately efficient incorporation was observed only for the lysine derivative (13%) and only very little incorporation was found for others. The results indicate, however, that large aromatic groups that cannot be incorporated in the form of β-arylalanines, become allowed when they are linked to an amino acid with long spacers.

D-Amino acids have been precluded by the *E. coli* ribosome system.[1] The inaccessibility may be due to the rejection of D-aminoacylated tRNA from binding to the ribosome A site, as suggested by the puromycin analogs.[32] However, since α,α-disubstituted amino acids with small side groups have been successfully incorporated in the *E. coli* system,[34] the behavior of D-amino acids must be re-examined. Glycolic acid and lactic acid are also allowed in the *E. coli* system.[35] The incorporation of an ester linkage in the protein main chain affords an opportunity for site-specific cleavage of the protein main chain. So far, incorporation of β-amino acids has not been reported. However, a –CO–NH–NH–CH-(R)–CO– linkage has been found to form when hydrazinophenylalanine was linked to a tRNA.[36] This suggests that some β-amino acids may also be incorporated into proteins.

The above examples indicate that the *E. coli* ribosomal system has some flexibility and may accept a variety of amino acids as the constituent of proteins, under the constraint

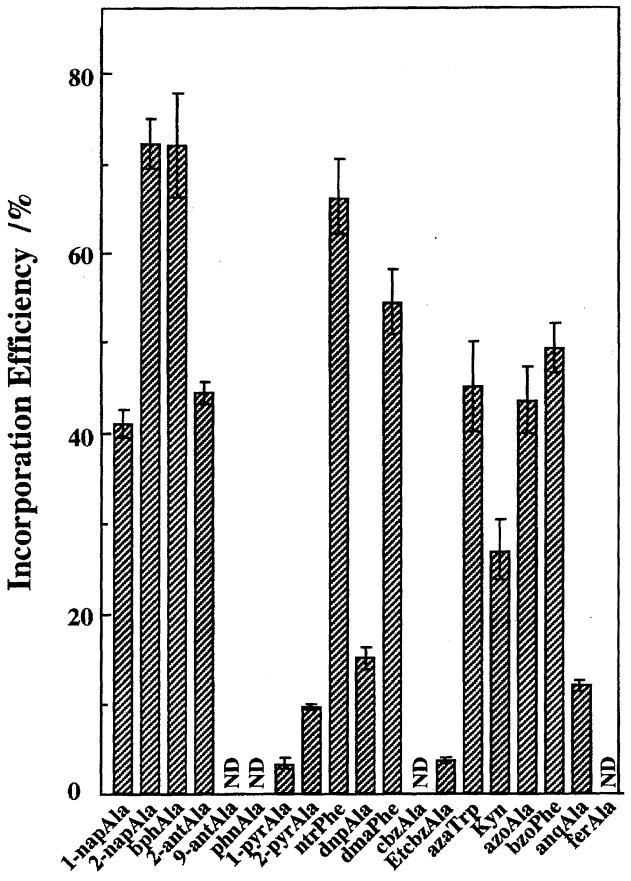


Fig. 10. Incorporation efficiencies of various nonnatural amino acids in the *E. coli* in vitro translation system that contained $tRNA_{CCCG}$'s charged with various nonnatural amino acids and the mRNA of streptavidin with the CGGG 4-base codon at the 83rd site.

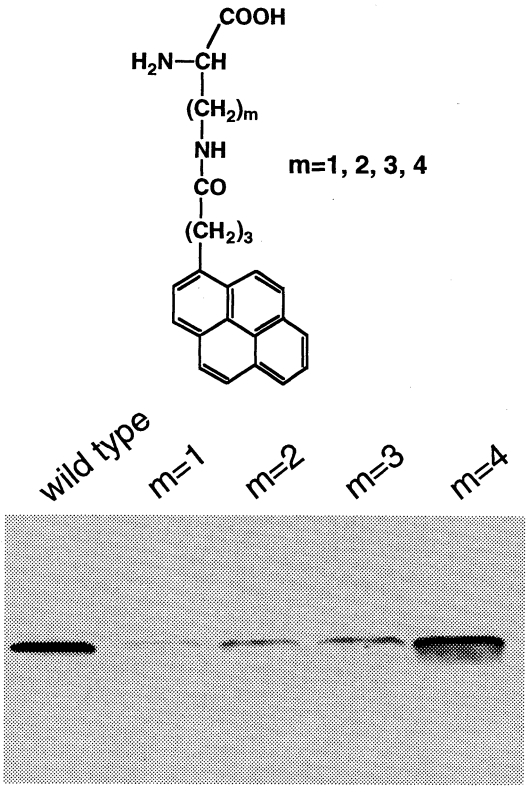


Fig. 11. Incorporation of nonnatural amino acids carrying a pyrenyl group that is linked to α-carbon with a spacer of different lengths.

shown in Fig. 9.

Protein Synthesis in the In Vitro Biosynthesizing System of Rabbit Reticulocyte. Since the amino acids are selected mainly by the accessibility to the ribosomal A site, if the space of the A site were expanded by appropriate mutations, a wider variety of nonnatural amino acids, including β- or γ-amino acids will be accepted. The modification of ribosome structure, however, is not realistic at present, because of the lack of atomic-level structural information.[37,38] Instead of the modification, we compared amino acid selectivities of different ribosomal systems from different biological origins.

Incorporation efficiencies were examined in rabbit reticulocyte lysate to explore the tolerance of the eukaryotic protein biosynthetic system.[29] Translation was carried out in the lysate of rabbit reticulocyte in the presence of an mRNA that encodes streptavidin with a CGGG 4-base codon at the 83rd position. Yeast $tRNA_{CCCG}$'s linked with a variety of nonnatural amino acids were used as adapters. As in the case of *E. coli* in vitro system, full-length streptavidin was synthesized only in the presence of aminoacyl $tRNA_{CCCG}$. No full-length protein was detected in the absence of the tRNA or in the presence of non-aminoacylated tRNA, indicating that a nonnatural amino acid was successfully incorporated into the protein under direction of the 4-base codon. Roughly speaking, the amino acid selectivity in the rabbit reticulocyte system was not much different from those of the *E. coli* system. It must be noted that 9-anthrylalanine, 9-phenanthrylalanine, 9-carbazolylalanine, and ferrocenylalanine could be incorporated neither in the rabbit system nor in the *E. coli* system and the simple selection rule shown in Fig. 9 holds also in the rabbit reticulocyte system. The parallel amino acid selectivity in the two systems suggests that the ribosome of rabbit reticulocyte recognizes the nonnatural amino acids in the same manner as that of *E. coli*. Presumably, the structures of the A-sites of the two types of ribosomes are well preserved.

A detailed comparison of the incorporation efficiencies in the two systems, however, discloses a subtle but important difference. The ratios of the incorporation efficiencies in the rabbit system to those in the *E. coli* are shown in Fig. 12. It is seen that the amino acids with large side groups such as pyrenyl, dinitrophenyl, and anthraquinonyl group are more efficiently incorporated in the rabbit system than in the *E. coli* system. This suggests that the molecular recognition in the rabbit ribosomal A site is a little looser than that of *E. coli* and the former system is more suitable for preparing proteins incorporated with nonnatural amino acids carrying large side groups.

Applications of Nonnatural Mutagenesis for Probing Protein Structures and Stabilities. As seen from the list of nonnatural amino acids in Fig. 2, a variety of artificial functions are ready to be incorporated into proteins. So far, however, the applications have been limited to relatively small fields, mainly due to the selectivity of the nonnatural amino acids by the ribosomes and the difficulty in keeping active protein conformations when a bulky nonnatural amino acid was built in.

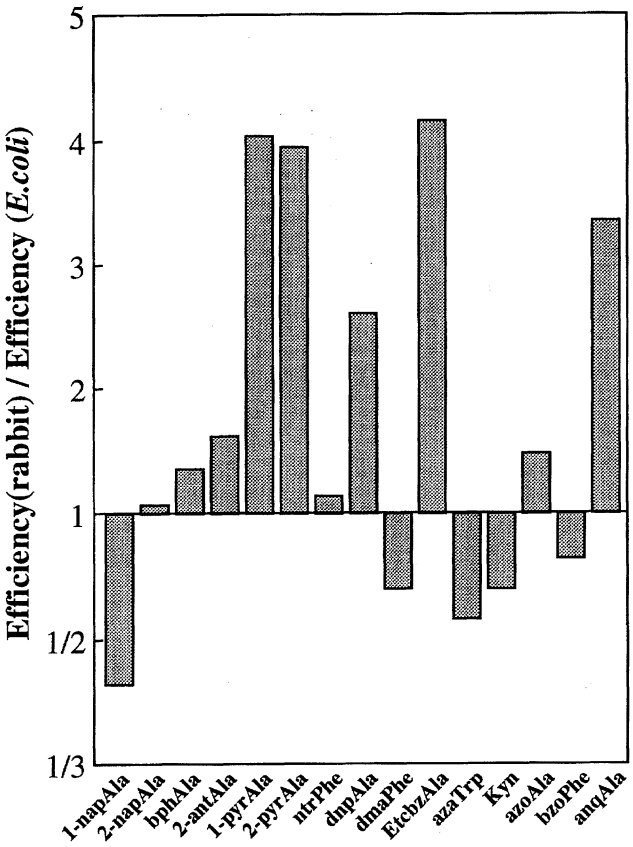


Fig. 12. Ratios of the incorporation efficiencies of nonnatural amino acids in rabbit reticulocyte system to those in the *E. coli* system.

The effects of introducing nonnatural amino acids that have unusual side-chain structures,[39—49] or additional hydrogen donors or acceptors,[50—52] on the protein stability have been studied by Schultz and co-workers. These studies have been reviewed by Schultz[53] and will not be repeated in this account.

Caged Proteins. β-*o*-Nitrobenzyl aspartate Asp(*O*-ntrBzl), has been known as the "caged" Asp that can be photodecomposed to regenerate an intact Asp unit. The caged Asp unit has been first incorporated into T4 lysozyme in place of Asp.[54] The re-activation of the caged protein by photoirradiation was actually observed when the caged Asp was incorporated into ras protein in place of Asp38.[55] Although the activity of ras protein in the absence of GTPase-activating protein (GAP) was affected neither by the introduction of the caged Asp unit, nor by the photoirradiation of the mutant, the GAP-mediated activity of the ras protein almost completely disappeared by the introduction of the caged Asp. The latter activity recovered up to about 50% of the wild-type activity after photoirradiation. However, since the photocleavage of *o*-nitrobenzyl ester must regenerate an intact Asp unit, a 100% recovery is expected in this experiment. The incomplete recovery has been explained in terms of side reactions associated with the photoirradiation.

A 100% recovery has been observed in a preliminary experiment of camel anti-lysozyme antibody incorporated with

Asp(*O*-ntrBzl) at the 99th position in place of Asp unit.[56] The mutant antibody lost its antigen-binding activity almost completely. After UV irradiation, the caged antibody fully restored its binding activity, as illustrated in Fig. 13. In contrast to this, the mutant antibody with 121Asp(*O*-ntrBzl) retained original antigen-binding activity and the activity was not affected by the photoirradiation. The positions of the caged amino acid play a crucial role in suppressing and restoring the original activity. The technique of caged proteins will find wide biochemical applications in triggering protein functions by photoirradiation.

Spin, Photoaffinity, and Fluorescence Labeling. A spin probe has been introduced into T4 lysozyme.[57] Taking the advantage of the recent developments of highly-sensitive ESR instruments, the application of spin probes in biochemical research is expanding rapidly. Detailed structural and dynamic information will become available by introducing a spin probe at a specific position.

p-Benzoylphenylalanine has been introduced into two types of cytochrome b_2-dihydrofolate reductase fusion proteins at various positions.[58] The mutant fusion proteins enter into the intermembrane space or the matrix of mitochondria and photocrosslink to the inner or outer membrane of mitochondria by UV irradiation. The analysis of the photocrosslinked product revealed the location and conformation of the protein in the membrane.

Fluorescence labeling is a more versatile method to explore location, conformation, and mobility of proteins. However, only a few types of nonnatural amino acids bearing highly fluorescent side group have been incorporated in the in vitro system. $^{\varepsilon}N$-Dansyl-L-lysine has been incorporated into

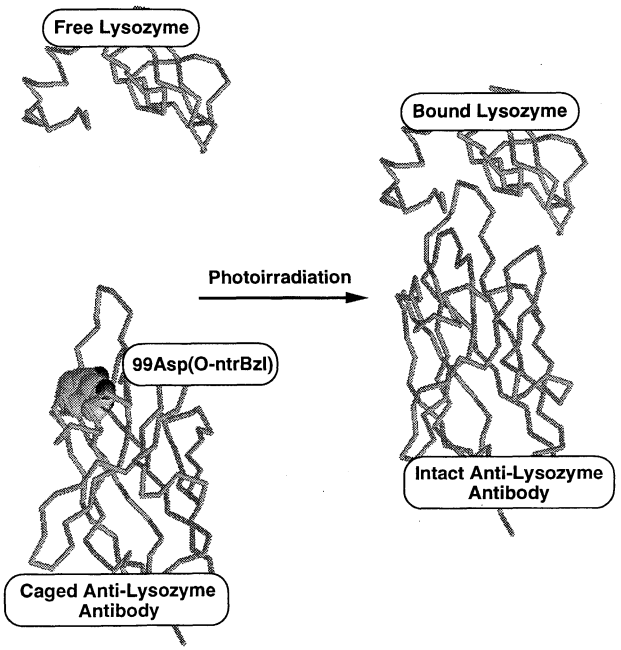


Fig. 13. Computer-predicted conformation of 99Asp(*O*-ntrBzl)-*anti*-lysozyme camel antibody (left). The caged antibody showed no binding activity, but restored full activity after photodecomposition of the caged Asp unit (right).

β-galactosidase.[59] However, the incorporation efficiency was very low (about 3% of the wild-type protein). Higher incorporation efficiency has been found for 5-hydroxytryptophan and 7-azatryptophan, but the fluorescence wavelengths of these amino acids are largely overlapped with the intrinsic fluorescence of tryptophan groups in proteins. Fluorescein group has been introduced by a keto-selective reaction of fluorescein hydrazide to mutant T4 lysozyme incorporated with *O*-2-ketopropyl-L-tyrosine.[60] Although the attachment of the fluorescent label was specific to the keto group, the reaction yield was not perfect (about 50%).

Direct incorporation of a fluorescent amino acid that emits stronger fluorescence at much longer wavelengths than the intrinsic fluorescence of proteins has been achieved by Chollet and co-workers.[61] They employed 3-*N*-(7-nitro-2,1,3-benzooxadiazol-4-yl)-2,3-diaminopropionic acid (NBD-Dap) that fluoresces at 550 nm with an approximate quantum yield 0.2. They introduced the fluorescent amino acids through in vivo synthesis in *Xenopus oocytes* that contained chemically misaminoacylated tRNA and the mRNA encoding NK2 receptor. The in vivo synthesis will be described in the last section of this account.

A simple fluorescent amino acid, L-2-anthrylalanine, was incorporated through *E. coli* in vitro system with a reasonable efficiency.[62] The anthryl group emits strong fluorescence at 390—450 nm; this may be detected without severe interference from the intrinsic fluorescence of tryptophan groups. The very strong absorption band around 255 nm with the absorption coefficient of the order of 10^5 is another advantage of the anthryl group, that makes it possible to detect fluorescence in less than *n*M concentrations.

As mentioned above, fluorescent probes are finding wide biochemical applications to observe location, conformation, and mobility of proteins under steady-state and pulsed irradiation. Taking advantage of recent progress in photon counting techniques, the behavior of even a single fluorescence molecule can be monitored. Moreover, the use of a confocal fluorescence microscope expands the fluorescence probe method into the field of μm-scale resolution. The site-specific incorporation of fluorescence probes, especially under constrained side-chain orientations as in the case of 2-anthrylalanine and 2-pyrenylalanine, will find wide applications in these fields.

Alteration of Colors of Biochemical Luminescence and Fluorescence. Firefly luciferase catalyzes oxidation of luciferin in the presence of ATP, Mg^{2+}, and molecular oxygen and emits luminescence around 600 nm. The biochemical luminescence has been used to measure the amount of ATP, because of its easy and sensitive detection. The luciferase gene has been used as a reporter gene. Hecht and coworkers[63,64] introduced natural and nonnatural amino acids into firefly luciferase and examined the luminescent intensities and wavelengths. The most striking color change was observed when Ser286 was replaced by Leu. The substitution induced the change in the luminescence color from 581 nm (yellow-green) to 622 nm (red). They also introduced *O*-glucosylated serine, serine phosphonate, tyrosine phosphate,

and tyrosine methylenephosphonate [4-(phosphanomethyl)-phenylalanine] at the 286 position. Unfortunately, introduction of these non-proteinogenic amino acids caused no striking changes in the luminescence behavior of luciferase. However, since glucosylation and phosphorylation of serine and tyrosine are common modification that are believed to occur after the translation, the ability to introduce such modifications at specific sites will find wide applications in biochemistry.

Green fluorescent protein (GFP) is a highly fluorescent protein found in a jelly fish, *Aequorea victoria*. The fluorophore consists of three amino acid units (65Ser–66ΔTyr–67Gly, ΔTyr = α,β-dehydrotyrosine) that form a cyclic structure as shown in Fig. 14. The conjugated π-system when incorporated into the protein framework emits strong fluorescence at 508 nm with fluorescence quantum yield of about 0.77.[65,66] The cyclization and dehydrogenation to give the fluorescent chromophore containing a ΔTyr unit take place spontaneously as a post-translation modification, as recently demonstrated by the fluorescence from the chemically synthesized GFP.[67] We have attempted in vitro synthesis of the wild-type and mutant GFP that contained a nonnatural amino acid at the 66th position in place of Tyr.[68] Although a wide variety of nonnatural amino acids have been introduced successfully, only two were fluorescent, i.e., 66*O*-methyltyrosine, and 66*p*-aminophenylalanine mutants. The former emitted around 450 nm and the latter at 494 nm with smaller quantum yields than the wild-type. The fluorescence from the *O*-methyltyrosine mutant indicates that the ionization of the phenol group of tyrosine is not crucial for emitting fluorescence. The alteration of fluorescence wavelength has been also achieved by the mutations at remote positions from the chromophore with naturally-occurring amino acids.[65]

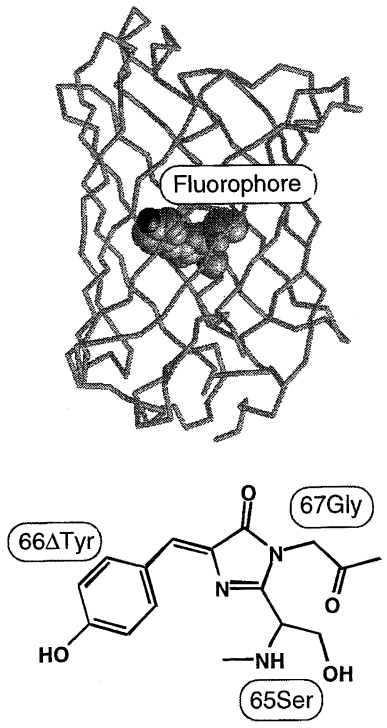


Fig. 14. X- Ray crystallographic structure of GFP and the fluorescent chromophore group consisting 65Ser–66ΔTyr–67Gly units.

Building up Paths for Electron Transfer on Proteins. One of attractive targets of nonnatural mutagenesis is to build paths for electron transfer (ET) on proteins. Since the ET process is very sensitive to the distance between donor and acceptor over the range of 5—20 Å, the site-specific incorporation of these groups into the 3-dimensional protein framework will be a powerful approach to achieve artificial photoenergy conversions and to design other photoelectronic devices. We have incorporated single L-*p*-nitrophenylalanine (ntrPhe) into streptavidin at 22 different positions (22nd, 25th, 43rd, 44th, 49th, 51st, 52nd, 54th, 65th, 79th, 80th, 83rd, 84th, 85th, 87th, 92nd, 101st, 108th, 114th, 117th, 120th, and 124th), respectively.[69] Streptavidin is known to bind biotin and biotin derivatives with an association constant of the order of 10^{15} M^{-1}. We have linked a pyrenyl group to biotin and the latter was bound to the mutant streptavidin. In this way, we have incorporated a pyrenyl group (P)–*p*-nitrophenyl group (T) pair into a streptavidin with different P–T distances and orientations. In the P–T pair, P works as a photosensitizer and T as an electron acceptor and the fluorescence of P group is quenched by the ET from excited P group (P^*) to T group. From the fluorescence quenching efficiency or the fluorescence lifetime of P group, the rate constants k_{ET}'s, of the photoinduced ET were determined as a function of the P–T edge-to-edge distances on the protein framework.

The positions and orientations of the ntrPhe units on mutant streptavidins bound with a *N*-biotinyl-L-1-pyrenylalanine at the biotin binding site were predicted from molecular mechanics calculations. The starting conformation was taken from the X-ray crystallographic structure of native streptavidin with a biotin in its binding site.[70] A computer software (PROCON) for molecular mechanics calculations of proteins that contain nonnatural amino acids and a modified prosthetic group has been programmed by one of the authors (MS). The computer-predicted orientations of the ntrPhe units at the 14 different positions are illustrated in Fig. 15, together with a biotinyl-L-1-pyrenylalanine bound to the biotin-binding site.

The quenching efficiency was very sensitive to the positions of the ntrPhe unit. For example, efficiency for the 83ntrPhe mutant was virtually zero, whereas that of the 84ntrPhe mutant was 89%. The marked position sensitivity indicates that the ET rate depends very sharply on distance and that the fluctuations of the main chain and the side chains are not so significant as to average out the difference of the positions of 83ntrPhe and 84ntrPhe unit.

The difference of ET efficiencies in different mutants may be interpreted simply in terms of the difference of the edge-to-edge distances between the P and T group. For example, the very different efficiencies of the 83ntrPhe and 84ntrPhe mutant may be explained in terms of the orientations of the ntrPhe unit at the two sites. As seen in Fig. 15, the 83ntrPhe unit is oriented to the opposite direction from the P group

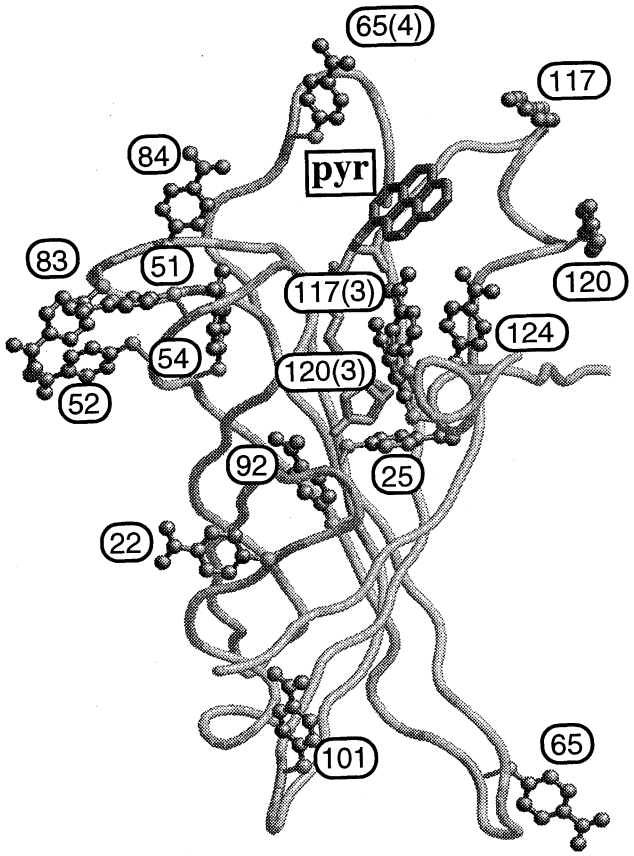


Fig. 15. Computer-predicted orientations of ntrPhe units in mutant streptavidins. The numbers indicate the position on the protein. The numbers with parentheses indicate that the ntrPhe unit belongs to different subunit in the tetramer. The orientation of pyrenyl group in the bound biotinyl-L-1-pyrenylalanine is also shown. The ntrPhe units in different mutants are shown together in this figure.

(r_{ee} = 16.8 Å), due mainly to the constraints of the main-chain conformation and a steric repulsion with 51Glu. On the other hand, the 84ntrPhe unit is oriented toward the P group (r_{ee} = 9.1 Å) due to the main-chain constraints. The ET to 65ntrPhe was moderately efficient (63%), but the latter is far separated from the P group (r_{ee} = 38.1 Å) on the same subunit. The question may be solved by taking the tetramer structure into consideration. The 65ntrPhe unit in the forth subunit was found to be close to the P group in the first unit (r_{ee} = 9.3 Å), shown as 65(4)ntrPhe in Fig. 15. Judging from these facts, we concluded that the ntrPhe units are correctly incorporated into the specific positions in the protein and take the orientations predicted from the molecular mechanics calculations. The quenching efficiencies and the ET rate constants are primarily determined simply by the distances in the tetrameric proteins.

The ET rate constants measured from the fluorescence quenching efficiencies are plotted against the closest edge-to-edge distances between P and T group in Fig. 16 by solid circles. The rate constants from the decay kinetics are also plotted in Fig. 16 as open circles. The distance dependence of ET rate constant may be roughly expressed by an exponential

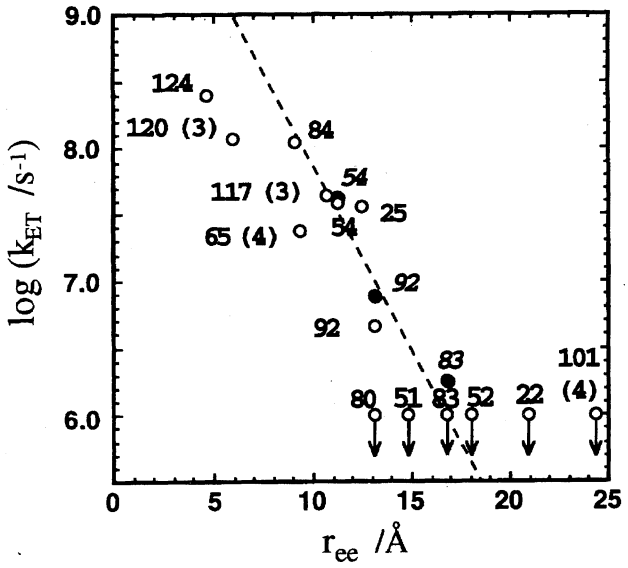


Fig. 16. ET rate constants from the excited pyrenyl group to nitrophenyl group on mutant streptavidins. The abscissa is the edge-to-edge distance.

function $k_{ET0}\exp[-\beta(r_{ee}-3)]$ with the rate constant for a contact P–T pair $k_{ET0} = 6.5\times10^9$ (s^{-1}) at r_{ee} = 3 Å, and the distant decay constant β = 0.63 ($Å^{-1}$). The linear relation in Fig. 16 again indicates that the ntrPhe units are correctly incorporated at the positions predicted in Fig. 15.

Site-Directed Incorporation of 2-Anthrylalanine into Streptavidin and Cytochrome b_5. As described before, L-2-anthrylalanine (antAla) is a simple and tractable nonnatural amino acid that emits strong fluorescence and can be incorporated into proteins with reasonable efficiency. We have incorporated single antAla into various positions of streptavidin.[71] Due to the bulkiness of the anthryl group, however, only 6 of the 19 mutants retained strong biotin-binding activity. The computer-predicted orientations of the 6 antAla units together with those of tryptophan units are shown in Fig. 17.

Fluorescence spectra of 120antAla-streptavidin, for example, excited at 290 nm showed strong fluorescence from anthryl group. At this excitation wavelength, however, the absorption of single anthryl group is much smaller than that of five tryptophan groups of streptavidin. Indeed, fluorescence spectrum of an equimolar mixture of wild-type streptavidin and *N*-acetylanthrylalanine showed negligible contribution of anthryl fluorescence. Therefore, the strong anthryl fluorescence indicates efficient intramolecular energy transfer from tryptophan groups to the 120th anthryl group. The efficiencies of energy transfer were evaluated from the fluorescence excitation spectra of all mutants and are plotted in Fig. 18. The abscissa of Fig. 18 represents the sum of the theoretical energy transfer efficiencies calculated from the center-to-center distances between the anthryl group to the tryptophan groups in the same subunit. The qualitative agreement between the observed and calculated efficiencies suggests that the antAla units are correctly incorporated at the specified positions and the protein conformations are not

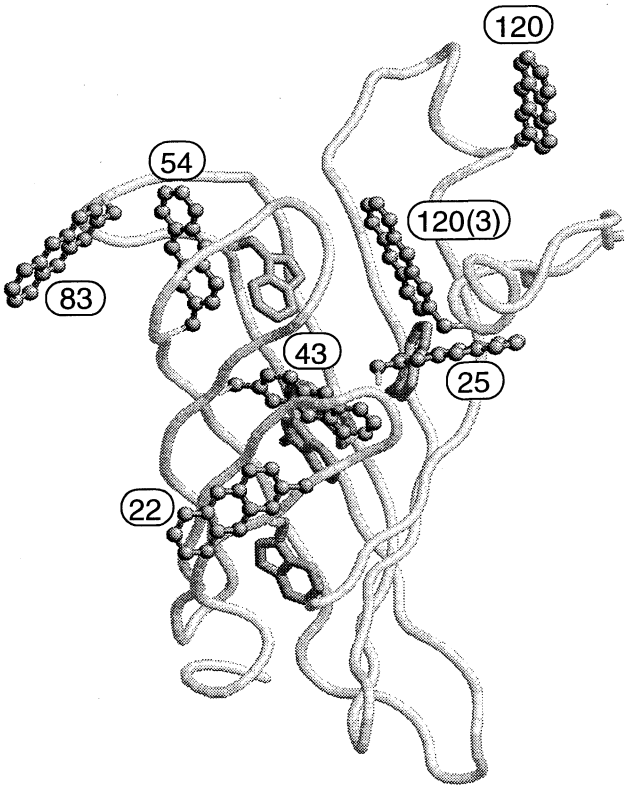


Fig. 17. Computer-predicted orientations of 2-anthrylalanines and tryptophans on mutant streptavidins. Anthrylalanines in six different mutants are shown together in this figure.

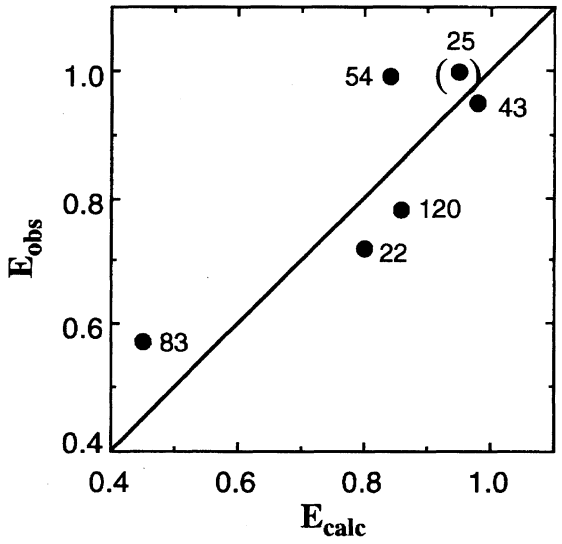


Fig. 18. Energy transfer efficiencies from tryptophan groups to single anthryl group in various antAla-streptavidins plotted against the theoretical efficiencies.

denatured by the incorporation of single anthryl group.

2-Anthrylalanine has been also introduced into cytochrome b_5 (Cyt b_5) at the 7th, 22nd, 45th, 58th, and 61st positions.[72] The computer-predicted orientations of anthryl groups on the mutant Cyt b_5 bound with heme are shown in Fig. 19. When hemin was added to the mutant, the anthryl fluorescence was largely quenched. The fluorescence quenching, in this case, can be attributed to both energy transfer and electron transfer. The quenching efficiencies decreased with the distance predicted from the molecular mechanics calculation, indicating again the correct locations and orientations of the anthryl groups as shown in Fig. 19.

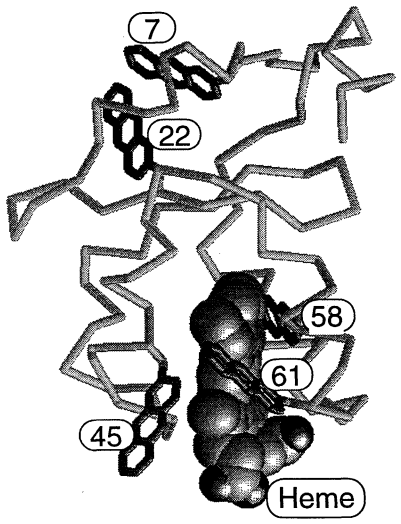


Fig. 19. Computer-predicted orientations of 2-anthrylalanines in mutant cytochrome b_5's with a heme group in the binding site.

In vivo Mutagenesis with Nonnatural Amino Acids—Co-Injection of Aminoacyl-tRNA and mRNA to *Xenopus Oocytes*. The expression of nonnatural mutants in living cells is an interesting issue. The in vivo system is advantageous, because of the possibility of direct control of cell functions by the introduction of nonnatural components. Novak et al.[73] reported the first attempt for the expression of a nonnatural mutant in an intact cell. Their procedure is illustrated in Fig. 20. They prepared a chemically misaminoacylated yeast $tRNA_{CUA}$ and an mRNA that encodes α-chain of nicotinic acetylcholine receptor (nAChR) and contains a UAG nonsense codon at the position of interest. Then the aminoacyl tRNA and the mRNA were co-injected into *Xenopus oocyte*. The yeast $tRNA_{CUA}$ was expected to behave independently from other tRNAs in the *oocyte*. Further, several mutations were made on the yeast tRNA to suppress re-aminoacylation by the endogenous ARSs in the *oocyte*. They successfully introduced some phenylalanine derivatives to the α-chain of nAChR that was expressed on the membrane surface of the *oocyte*. The incorporation of nonnatural amino acids produced detectable changes in the function of the ion channel. Later, they found that an engineered tetrahymena $tRNA^{Gln}$ was 100-fold less aminoacylated by the endogenous tRNA synthetases in *Xenopus oocyte* and 4—10 fold more efficient in incorporating nonnatural amino acids.[74]

The in vivo system has been finding applications as a powerful tool in biochemical studies.[75—80] A biotinylated amino acid $^{\varepsilon}N$-biotinyl-L-lysine = biocytin, has been incorporated into several positions in the α-chain of nAChR in *Xenopus oocytes*.[75] In this case, the highly specific binding of strep-

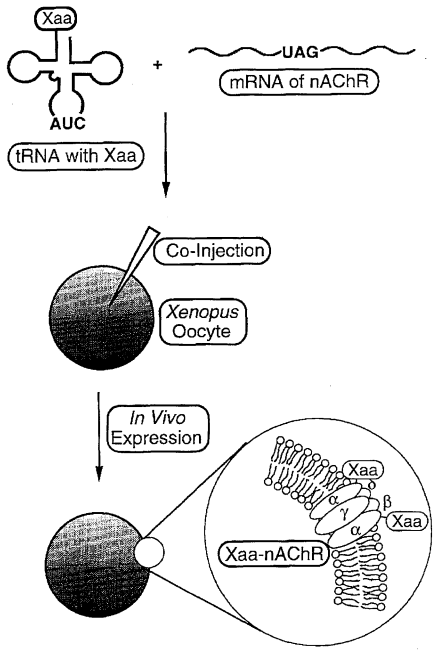


Fig. 20. Nonnatural mutagenesis expressed in the living *Xenopus oocyte*. The aminoacyl tRNA carrying a nonnatural amino acid and the target mRNA that includes a UAG nonsense codon are co-injected into the cell. The nonnatural mutant of *n*AChR is expressed on the membrane of the cell.

tavidin to the biotinylated subunit of nAChR was used to determine the surface exposure of individual amino acids of nAChR.

A photoreactive nonnatural amino acid, *o*-nitrophenylglycine (Npg), has been introduced into two ion channels on the membrane surface of *Xenopus oocyte*: the *Drosophila* Shaker BK^+ channel and the nAChR.[76] *o*-Nitrophenylglycine, when incorporated into a polypeptide chain, was expected to induce a backbone cleavage by photoirradiation. The photochemical proteolysis was actually suggested by the disappearance of the regulative activity of the channel current in the mutant channel. The photochemical deactivation of protein functions is an opposite of the caged protein technique and will add a new version of phototriggering of biological reactions.

As described before, Chollet and co-workers[61,81] reported a fluorescence labeling of NK2 receptor through in vivo synthesis in *Xenopus oocyte*. They used a fluorescent amino acid (NBD-Dap) that shows a sharp emission peak at 550 nm, far from tryptophan fluorescence. They bound a tetramethylrhodamine (TMR)-labeled peptide antagonist to the receptor and observed energy transfer from the excited NBD group to the TMR group. From the energy-transfer efficiencies, they successfully evaluated distances between the ligand and the receptor.

One of the problems in using fluorescence labels in the in vivo system is that the fluorescence from free amino acids or from the aminoacyl tRNA that are remaining unreacted in the cell cannot be eliminated. Therefore, the latter contribution must be subtracted from the observed fluorescence

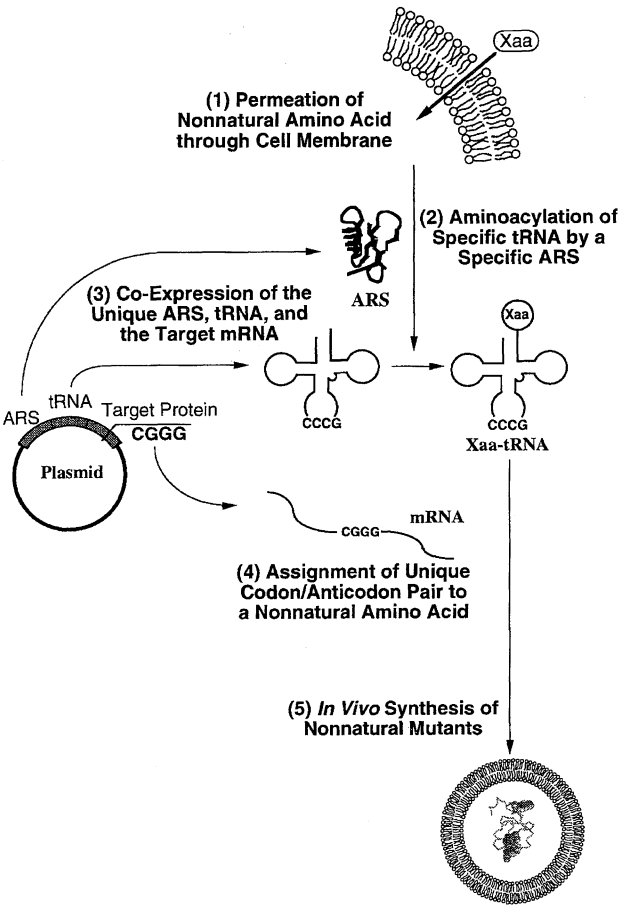


Fig. 21. Key steps required to achieve in vivo biosynthesis of nonnatural mutants.

to obtain fluorescence from the labeled protein. The use of fluorescence labels that show different emission peaks and intensities in the free and bound state, respectively will partially solve the above problem.

Attempts to Prepare an Orthogonal ARS–tRNA pair that is Independent from the Endogenous ARS–tRNA Pairs. In nature, a non-standard amino acid, selenocystein, has been incorporated into some redox enzymes during translation.[82] The UGA stop codon is assigned to selenocysteine and a $tRNA^{Sec}_{UCA}$ is the adapter. The $tRNA^{Sec}_{UCA}$ is first aminoacylated with serine and the latter was converted to selenocysteinyl-$tRNA^{Sec}_{UCA}$ by selenocysteine tRNA synthetase. The selenocysteine has been found, for example, in *E. coli* formate dehydrogenase and mammalian tetraiodothyronine deiodinase type I. The latter enzyme links the selenium status of an organism with the thyroid hormone level.

In this context, the nonnatural mutants can not only be synthesized in the in vitro cell-free systems or in the in vivo system that contained an mRNA and a misaminoacylated tRNA with a unique codon/anticodon pair, but also may be expressed in in vivo systems from free amino acids. For the in vivo expression of nonnatural mutants from free amino acids, however, four problems must be solved (Fig. 21): (1) A nonnatural amino acid must enter into the cell through the

cell membrane. (2) An artificial aminoacyl tRNA synthetase (ARS) has to be designed and expressed in the cell. The artificial ARS should bind a specific nonnatural amino acid and link it to a special tRNA. (3) The special tRNA has to be produced also in the cell. The tRNA should not be aminoacylated with the ARSs that are inherently present in the cell, but may be linked to a specific nonnatural amino acid by the aid of the artificial ARS. (4) As in the in vitro system, a unique codon/anticodon pair must be assigned to the nonnatural amino acid. (5) Finally, the mutant proteins must be neither lethal to the cell nor quickly decomposed in the cell.

In the above experiments of co-injection of mRNA and misaminoacylated tRNA into *Xenopus oocyte*, the difficulty in designing an ARS–tRNA pair that is orthogonal to the endogenous ARS–tRNA pairs has been bypassed. However, to create a living organism that accepts or requires a nonnatural amino acid for its living and growth, the unique ARS–tRNA pair is essential. The design of the orthogonal ARS–tRNA pair has been elaborated by some workers. Liu et al.[83,84] engineered a tRNA–ARS pair to achieve an orthogonal incorporation of a nonnatural amino acid in the *E. coli* system. They introduced eight mutations into *E. coli* $tRNA^{Gln}$ to suppress aminoacylation by the endogenous GlnRS and other ARSs in *E. coli*. They also synthesized mutant ARS that may efficiently aminoacylate the mutant tRNA with glutamine in vitro. The engineered tRNA–ARS pair may constitute a functional pair to introduce nonnatural amino acids.

Furter[85] also reported an orthogonal tRNA–ARS pair that works in *E. coli* in vivo system. They coexpressed a yeast $tRNA^{Phe}_{CUA}$–PheRS pair in an *E. coli* strain that is resistant to *p*-fluorophenylalanine. Since yeast $tRNA^{Phe}_{CUA}$ was not aminoacylated by *E. coli* ARSs and the yeast PheRS did not aminoacylate *E. coli* tRNAs, the coexpressed tRNA–ARS pair will work independently from the *E. coli* endogenous tRNA–ARS pairs and may introduce a nonnatural amino acid at the amber position. They actually observed site-directed incorporation of *p*-fluorophenylalanine into dihydrofolate reductase (DHFR) in the *E. coli* in vivo system. In their case, however, the charge of phenylalanine onto yeast $tRNA_{CUA}$ cannot be avoided, resulting in a protein mixture with *p*-fluorophenylalanine and phenylalanine at the amber position.

A similar approach using yeast $tRNA^{Tyr}_{CUA}$–TyrRS pair has been reported by Nishikawa and co-workers.[86] They confirmed the effectiveness and uniqueness of the above pair in the synthesis of β-galactosidase, although no nonnatural amino acid has been charged to the yeast tRNA yet.

The biggest problem for the in vivo incorporation of nonnatural amino acids is the design of a unique ARS that accepts a specific nonnatural amino acid exclusively and charges it onto the unique tRNA. The approaches toward the artificial ARS have been described before. Since a wide diversity of nonnatural amino acids are available, as exemplified in Fig. 2, preparation of a unique ARS for each nonnatural amino acid seems to be unrealistic. However, for some nonnatural amino acids that have very unique side groups, the production of a unique ARS through combinatorial method, for example, will be promising.

Conclusion

Nonnatural mutagenesis is a powerful and versatile technique not only as a new tool for biochemical research but also as a new way to produce functional molecules or materials that are used in medicinal and other fields. In the past decade, however, efforts have been focused mostly on the development and extension of the technique itself from in vitro synthesis to in vivo. In the next decade, attention will be focused on the application of the technique to the development of new protein drugs, including those for diagnostic uses, photoenergy-chemical energy conversion systems, and biosensors. These applications will be possible by the incorporation of a wide variety of nonnatural amino acids into various proteins.

The authors are grateful to professor Shigeyuki Yokoyama, University of Tokyo for introducing us the technique of in vitro biosynthesis. Recent experiments included in this account have been conducted with a Grant-in-Aid for Scientific Research No. 09555287 from the Ministry of Education, Science, Sports and Culture.

References

1 C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, and P. G. Schultz, *Science*, **244**, 182 (1989).

2 S. J. Anthony-Cahill, M. C. Griffith, C. J. Noren, D. J. Suich, and P. G. Schultz, *Trends Biochem. Sci.*, **14**, 400 (1989).

3 J. D. Bain, C. G. Glabe, T. A. Dix, and A. R. Chamberlin, *J. Am. Chem. Soc.*, **111**, 8013 (1989).

4 C. J. Noren, S. J. Anthony-Cahill, D. J. Suich, K. A. Noren, M. C. Griffith, and P. G. Schultz, *Nucleic Acids Res.*, **18**, 83 (1990).

5 J. Ellman, D. Mendel, S. Anthony-Cahill, C. J. Noren, and P. G. Schultz, *Methods Enzymol.*, **202**, 301 (1991).

6 J. D. Bain, E. S. Diala, C. G. Glabe, D. A. Wacker, M. H. Lyttle, T. A. Dix, and A. R. Chamberlin, *Biochemistry*, **30**, 5411 (1991).

7 F. Chapeville, F. Lipmann, G. von Ehrenstein, B. Weisblum, W. J. Ray, and S. Benzer, *Proc. Natl. Acad. Sci. U.S.A.*, **48**, 1086 (1962).

8 J. S. Thorson, V. W. Cornish, J. E. Barrett, S. T. Cload, T. Yano, and P. G. Schultz, *Methods Mol. Biol.*, **77**, 43 (1998).

9 L. E. Steward and A. R. Chamberlin, *Methods Mol. Biol.*, **77**, 325 (1998).

10 T. Kigawa, T. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, T. Shibata, and S. Yokoyama, *FEBS Lett.*, **442**, 15 (1999).

11 L. Jermutus, L. A. Ryabova, and A. Pluckthun, *Curr. Opin. Biotechnol.*, **9**, 534 (1998).

12 T. Lescrinier, C. Hendrix, L. Kerremans, J. Rozenski, A. Link, B. Samyn, A. Van Aerschot, E. Lescrinier, R. Eritja, J. Van Beeumen, and P. Herdewijn, *Chem. Eur. J.*, **4**, 425 (1998).

13 Y. Konishi, "Peptide Synthesis Database Homepage": http://aminoacid.bri.nrc.ca:1125/

14 a) L-4-biphenylylalanine: M. Kuragaki and M. Sisido, *J. Phys. Chem.*, **100**, 16019 (1996). b) L-9-anthrylalanine: S. Egusa, M. Sisido, and Y. Imanishi, *Bull. Chem. Soc. Jpn.*, **59**, 3175 (1986);

M. Sisido, *Macromolecules*, **22**, 4367 (1989). c) DL-9-phenanthrylalanine: H. Sasaki, M. Sisido, and Y. Imanishi, *Langmuir*, **7**, 1949 (1991). d) L-1-pyrenylalanine: S. Egusa, M. Sisido, and Y. Imanishi, *Chem. Lett.*, **1983**, 1307; S. Egusa, M. Sisido, and Y. Imanishi, *Macromolecules*, **18**, 88 (1985). e) L-3-(*N*-ethylcarbazolyl)lalanine: K. Taku, H. Sasaki, S. Kimura, and Y. Imanishi, *Amino Acids*, **7**, 311 (1994). f) L-2-anthraquinonylalanine: T. Matsubara, H. Shinohara, and M. Sisido, *Macromolecules*, **30**, 2651 (1997). g) L-ferrocenylalanine: M. Kira and M. Sisido, *Chem. Lett.*, **1997**, 89. h) L-9-Carbazolylalanine was synthesized accroding to the private communication from Professor N. Nishino of Kyushu Institute of Technology. i) L-*p*-dimethylaminophenylalanine: M. Sisido, R. Tanaka, Y. Inai, and Y. Imanishi, *J. Am. Chem. Soc.*, **111**, 6790 (1989). j) L-*p*-phenylazophenylalanine: M. Goodman and A. Kossoy, *J. Am. Chem. Soc.*, **88**, 5010 (1966).

15 T. G. Heckler, Y. Zama, T. Naka, and S. M. Hecht, *J. Biol. Chem.*, **258**, 4492 (1983).

16 T. G. Heckler, L. H. Chang, Y. Zama, T. Naka, and S. M. Hecht, *Tetrahedron*, **40**, 87 (1984).

17 S. A. Robertson, J. A. Ellman, and P. G. Schultz, *J. Am. Chem. Soc.*, **113**, 2722 (1991).

18 M. Lodder, S. Golovine, A. L. Laikhter, V. A. Karginov, and S. M. Hecht, *J. Org. Chem.*, **63**, 794 (1998).

19 M. J. Dougherty, S. Kothakota, T. L. Mason, D. A. Tirrell, and M. J. Fournier, *Macromolecules*, **26**, 1779 (1993).

20 E. Yoshikawa, M. J. Fournier, T. L. Mason, and D. A. Tirrell, *Macromolecules*, **27**, 5461 (1994).

21 S. Kothakota, T. L. Mason, D. A. Tirrell, and M. J. Fournier, *J. Am. Chem. Soc.*, **117**, 536 (1995).

22 M. H. Richmond, *Bacteriol. Rev.*, **26**, 398 (1962).

23 T. Ueda, M. Ueda, A. Tanaka, M. Sisido, and Y. Imanishi, *Bull. Chem. Soc. Jpn.*, **64**, 1576 (1991).

24 P. Kast and H. Hennecke, *J. Mol. Biol.*, **222**, 99 (1991).

25 M. Ibba, P. Kast, and H. Hennecke, *Biochemistry*, **33**, 7107 (1994).

26 M. Ibba and H. Hennecke, *FEBS Lett.*, **364**, 272 (1995).

27 J. R. Jacobsen, J. R. Prudent, L. Kochersperger, S. Yonkovich, and P. G. Schultz, *Science*, **256**, 365 (1992).

28 a) T. Hohsaka, Y. Ashizuka, H. Murakami, and M. Sisido, *J. Am. Chem. Soc.*, **118**, 9778 (1996). b) T. Hohsaka, Y. Ashizuka, H. Sasaki, H. Murakami, and M. Sisido, in preparation.

29 T. Hohsaka, D. Kajihara, Y. Ashizuka, H. Murakami, and M. Sisido, *J. Am. Chem. Soc.*, **121**, 34 (1999).

30 J. D. Bain, C. Switzer, A. R. Chamberlin, and S. A. Benner, *Nature*, **356**, 537 (1992).

31 T. Kanda, K. Takai, S. Yokoyama, and H. Takaku, *FEBS Lett.*, **440**, 273 (1998).

32 T. Hohsaka, K. Sato, M. Sisido, K. Takai, and S. Yokoyama, *FEBS Lett.*, **335**, 47 (1993).

33 S. Tokunaga, T. Hohsaka, and M. Sisido, unpublished work.

34 D. Mendel, J. Ellman, and P. G. Schultz, *J. Am. Chem. Soc.*, **115**, 4359 (1993).

35 H. H. Chung, D. R. Benson, and P. G. Schultz, *Science*, **259**, 806 (1993).

36 J. A. Killian, M. S. Van Cleve, Y. F. Shayo, and S. M. Hecht, *J. Am. Chem. Soc.*, **120**, 3032 (1998).

37 J. Frank, J. Zhu, P. Penczek, Y. Li, S. Srivastava, A. Verschoor, M. Radermacher, R. Grassucci, R. K. Lata, and R. K. Agrawal, *Nature*, **376**, 441 (1995).

38 R. K. Agrawal, P. Penczek, R. A. Grassucci, Y. Li, ArD. Leith, K. H. Nierhaus, and J. Frank, *Science*, **271**, 1000 (1996).

39 J. A. Ellman, D. Mendel, and P. G. Schultz, *Science*, **255**, 197 (1992).

40 D. Mendel, J. A. Ellman, Z. Chang, D. L. Veenstra, P. A. Kollman, and P. G. Schultz, *Science*, **256**, 1798 (1992).

41 J. K. Judice, T. R. Gamble, E. C. Murphy, A. M. de Vos, and P. G. Schultz, *Science*, **261**, 1578 (1993).

42 H. H. Chung, D. R. Benson, V. W. Cornish, and P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 10145 (1993).

43 V. W. Cornish, M. I. Kaplan, D. L. Veenstra, P. A. Kollman, and P. G. Schultz, *Biochemistry*, **33**, 12022 (1994).

44 Y. Kimata, H. Shimada, T. Hirose, and Y. Ishimura, *Biochem. Biophys. Res. Commun.*, **208**, 96 (1995).

45 Z. Zhao, X. Liu, Z. Shi, L. Danley, B. Huang, R.-T. Jiang, and M.-D. Tsai, *J. Am. Chem. Soc.*, **118**, 3535 (1996).

46 Y. Park, J. Luo, P. G. Schultz, and J. F. Kirsch, *Biochemistry*, **36**, 10517 (1997).

47 J. T. Koh, V. W. Cornish, and P. G. Schultz, *Biochemistry*, **36**, 11314 (1997).

48 V. A. Karginov, S. V. Mamaev, H. An, M. D. Van Cleve, S. M. Hecht, G. A. Komatsoulis, and J. N. Abelson, *J. Am. Chem. Soc.*, **119**, 8166 (1997).

49 J. S. Thorson, I. Shin, E. Chapman, G. Stenberg, B. Mannervik, and P. G. Schultz, *J. Am. Chem. Soc.*, **120**, 451 (1998).

50 J. S. Thorson, C. Eli, and P. G. Schultz, *J. Am. Chem. Soc.*, **117**, 1157 (1995).

51 I. Shin, A. Y. Ting, and P. G. Schultz, *J. Am. Chem. Soc.*, **119**, 12667 (1997).

52 E. Chapman, J. S. Thorson, and P. G. Schultz, *J. Am. Chem. Soc.*, **119**, 7151 (1997).

53 V. W. Cornish, D. Mendel, and P. G. Schultz, *Angew. Chem., Int. Ed. Engl.*, **34**, 621 (1995).

54 D. Mendel, J. A. Ellman, and P. G. Schultz, *J. Am. Chem. Soc.*, **113**, 2758 (1991).

55 S. K. Pollitt and P. G. Schultz, *Angew. Chem., Int. Ed. Engl.*, **37**, 2104 (1998).

56 S. Nakamura, T. Hohsaka, and M. Sisido, in preparation.

57 V. W. Cornish, D. R. Benson, C. A. Altenbach, K. Hideg, W. L. Hubbell, and P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 2910 (1994).

58 T. Kanamori, S. Nishikawa, I. Shin, P. G. Schultz, and T. Endo, *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 485 (1995).

59 L. E. Steward, C. S. Collins, M. A. Gilmore, J. E. Carlson, J. B. A. Ross, and A. R. Chamberlin, *J. Am. Chem. Soc.*, **119**, 6 (1997).

60 V. W. Cornish, K. M. Hahn, and P. G. Schultz, *J. Am. Chem. Soc.*, **118**, 8150 (1996).

61 G. Turcatti, K. Nemeth, M. D. Edgerton, J. Knowles, H. Vogel, and A. Chollet, *Receptors Channels*, **5**, 201 (1997).

62 T. Hohsaka, K. Sato, M. Sisido, K. Takai, and S. Yokoyama, *FEBS Lett.*, **344**, 171 (1994).

63 S. V. Mamaev, A. L. Laikhter, T. Arslan, and S. M. Hecht, *J. Am. Chem. Soc.*, **118**, 7243 (1996).

64 T. Arslan, S. V. Mamaev, N. V. Mamaeva, and S. M. Hecht, *J. Am. Chem. Soc.*, **119**, 10877 (1997).

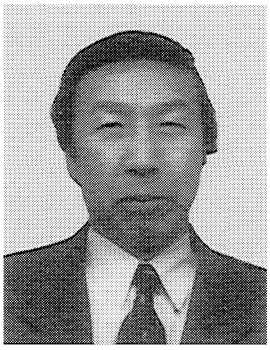
65 H. Niwa, T. Inouye, T. Hirano, T. Matsuno, S. Kojima, M. Kubota, M. Ohashi, and F. I. Tsuji, *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 13617 (1996).

66 R. Heim and R. Y. Tsien, *Current Biol.*, **6**, 178 (1996).

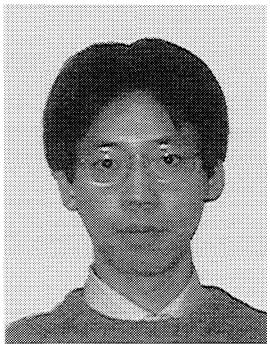
67 Y. Nishiuchi, T. Inui, H. Nishio, J. Bódi, T. Kimura, F. I. Tsuji, and S. Sakakibara, *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 13549 (1998).

68 D. Kajihara, T. Hohsaka, H. Murakami, and M. Sisido, in preparation.

69 H. Murakami, T. Hohsaka, Y. Ashizuka, and M. Sisido, *J. Am. Chem. Soc.*, **120**, 7520 (1998).
70 P. C. Weber, D. H. Ohlendorf, J. J. Wendoloski, and F. R. Salemme, *Science*, **243**, 85 (1989).
71 H. Murakami, T. Hohsaka, Y. Ashizuka, and M. Sisido, in preparation.
72 K. Yamanaka, T. Yasukouchi, T. Hohsaka, and M. Sisido, in preparation.
73 M. W. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty, and H. A. Lester, *Science*, **268**, 439 (1995).
74 M. E. Saks, J. R. Sampson, M. W. Nowak, P. C. Kearney, F. Du, J. N. Abelson, H. A. Lester, and D. A. Dougherty, *J. Biol. Chem.*, **271**, 23169 (1996).
75 J. P. Gallivan, H. A. Lester, and D. A. Dougherty, *Chem. Biol.*, **4**, 739 (1997).
76 P. M. England, H. A. Lester, N. Davidson, and D. A. Dougherty, *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 11025 (1997).
77 P. C. Kearney, H. Zhang, W. Zhong, D. A. Dougherty, H. A. Lester, and M.-D. Tsai, *Neuron*, **17**, 1221 (1996).
78 P. C. Kearney, M. W. Nowak, W. Zhong, S. K. Silverman, H. A. Lester, and D. A. Dougherty, *Mol. Pharmacol.*, **50**, 1401 (1996).
79 D. A. Dougherty, *Pure Appl. Chem.*, **69**, 1969 (1997).
80 M. W. Nowak, J. P. Gallivan, S. K. Silverman, C. G. Labarca, D. A. Dougherty, and H. A. Lester, *Methods Enzymol.*, **293**, 504 (1998).
81 G. Turcatti, K. Nemeth, M. D. Edgerton, U. Meseth, F. Talabot, M. Peitsch, J. Knowles, H. Vogel, and A. Chollet, *J. Biol. Chem.*, **271**, 19991 (1996).
82 A. Böck, K. Forchhammer, J. Heider, and C. Baron, *TIBS*, **16**, 463 (1991).
83 D. R. Liu, T. J. Magliery, and P. G. Schultz, *Chem. Biol.*, **4**, 685 (1997).
84 D. R. Liu, T. J. Magliery, M. Pastrank, and P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 10092 (1997).
85 R. Furter, *Protein Sci.*, **7**, 419 (1998).
86 S. Ohno, T. Yokogawa, I. Fujii, H. Asahara, H. Inokuchi, and K. Nishikawa, *J. Biochem. (Tokyo)*, **124**, 1065 (1998).



Masahiko Sisido was born in Kyoto in 1944. He graduated from the Department of Polymer Chemistry, Kyoto University, in 1967 and received his D. Engineering from the same department in 1973. He was appointed as a research associated at the Department of Polymer Chemistry in 1972, then as an associated professor in Research Center for Medical Polymers and Biomaterials, Kyoto University in 1981. He moved to Research Laboratory of Resources Utilization, Tokyo Institute of Technology in 1987 and was promoted to be a full professor at Okayama University in 1993. During 1977–1979, he stayed at the Department of Chemistry and Biochemistry, Northwestern University, as a visiting assistant professor. His current research interest includes synthetic peptides and engineered proteins that contain nonnatural amino acids with specialty side groups. He received the scientific award from the Society of Polymer Science, Japan in 1990.



Takahiro Hohsaka was born in Nagano in 1970 and graduated from Tokyo Institute of Technology with B. Eng. (1992) and M. Eng. (1994) degrees. In April 1994, he was appointed as a research associated in Okayama University. His current interest is directed to extended genetic and protein engineering.