 

*Award Accounts*

**The Chemical Society of Japan Award for Young Chemists for 2002**

# Incorporation of Nonnatural Amino Acids into Proteins through Extension of the Genetic Code

**Takahiro Hohsaka**

School of Materials Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Tatsunokuchi, Ishikawa 923-1292

PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012

Received September 11, 2003; E-mail: hohsaka@jaist.ac.jp

Incorporation of nonnatural amino acids into specific positions of proteins was achieved by using extended codons such as CGGG four-base codon. Aminoacyl-tRNAs having nonnatural amino acids and four-base anticodons were prepared and added to an *E. coli* in vitro translation system together with mutated genes containing four-base codons. Product analysis of the in vitro translation demonstrates that four-base codons are successfully decoded by the corresponding aminoacyl-tRNAs and, as a consequence, nonnatural amino acids are incorporated into desired positions of proteins. By using two independent four-base codons, incorporation of two nonnatural amino acids into single proteins was achieved for the first time. Five-base codons were also found to be available for the nonnatural amino acid incorporation. Extension of the amino acid repertoire was investigated by using four-base codons, and a wide variety of nonnatural amino acids were incorporated into proteins. According to the information on amino acid selectivity of ribosomes, novel fluorescently labeled nonnatural amino acids were designed and utilized for position-specific fluorescence labeling of proteins. The extended codon strategy will allow scientists to construct novel artificial proteins and biomolecular systems.

Development of genetic engineering allows us to produce artificial proteins with modified amino acid sequences. This technique called "Protein Engineering"[1] can replace any amino acid with one or more other ones. By using this technique, a number of proteins with improved catalytic activity, thermostability, or altered substrate specificity have been generated. This replacement is, however, limited within the naturally occurring 20 amino acids. Incorporation of amino acids other than these 20 into proteins will expand a possibility of protein functions. Unfortunately, protein synthesizing systems in usual living organisms do not have any ability to handle nonnatural amino acids.

To realize the incorporation of nonnatural amino acids into proteins, several breakthroughs are needed. In the first place, the nonnatural amino acid must be attached to some transfer RNA (tRNA). In protein biosynthesis, genetic information of DNA is transcribed into messenger RNA (mRNA). The information of mRNA is then translated into proteins in ribosome, the main machinery of the protein synthesizing system. In ribosome, triplet genetic code (i.e., codon) is decoded into one of 20 amino acids according to the genetic code. This decoding process is mediated by tRNAs containing triplet anticodons and aminoacylated with their cognate amino acid catalyzed by their cognate aminoacyl-tRNA synthetase. The aminoacyl-tRNA binds to a ribosomal A-site, if a codon and anticodon pair is correctly formed. Then the amino acid residue of the aminoacyl-tRNA is transferred to an elongating peptide attached to a peptidyl-tRNA on a ribosomal P-site. The principle of the translation reaction inspires us with a general method to incorporate nonnatural amino acids instead of natural ones. Nonnatural amino acid will be acceptable for ribosome as a substrate for protein synthesis when the amino acid is attached to any of the tRNAs. Some amino acid analogs such as azatryptophan and selenomethionine have actually been incorporated into proteins when the amino acid analogs are attached to tRNAs through an incorrect recognition of aminoacyl-tRNA synthetases.[2]

In the second place, the nonnatural amino acid must be encoded by some other codon than the usual triplet codons encoding 20 amino acids. The genetic code is constituted of 64 triplet codons, of which 61 are decoded into one of the 20 amino acids and 3 are decoded as a stop signal of the translation reaction. Since there is no spare codon, the use of the triplet codons competes with the decoding by natural aminoacyl-tRNAs and release factors. In the above cases of azatryptophan and selenomethionine, the incorporation of the amino acid analogs competes with the decoding by the natural aminoacyl-tRNAs, resulting in the production of the heterogeneous proteins containing or not containing the amino acid analogs.

In 1989, two groups, P. G. Schultz's group and A. R. Chamberlin's group, independently reported a methodology to achieve site-specific incorporation of nonnatural amino acids

Published on the web June 9, 2004; DOI 10.1246/bcsj.77.1041

into proteins.[3,4] In this method, an amber stop codon assigning the position of nonnatural amino acids and an amber suppressor tRNA containing a CUA anticodon corresponding to UAG codon were used. The amber suppressor tRNA was aminoacylated with nonnatural amino acids by using a chemical misaminoacylation method that had been developed by Hecht and co-workers.[5] The mRNA containing the amber codon at desired positions and the amber suppressor tRNA chemically aminoacylated with nonnatural amino acids were added into an in vitro translation system, and proteins containing nonnatural amino acids at the positions directed by the amber codon were successfully generated.

Although many researchers have been using the amber suppression method in order to introduce nonnatural amino acids into proteins in which they are interested,[6–15] this method has some serious drawbacks. One is competition with a release factor, which terminates the translation and, as a result, reduces the efficiency of the incorporation of nonnatural amino acids. The other is a limitation of the number of nonnatural amino acids that can be incorporated into a single protein. Because only an amber codon has been practically used for the nonnatural amino acid incorporation,[16] only single nonnatural amino acids could be introduced into single proteins. Chamberlin and coworkers reported that an artificial codon–anticodon pair containing *iso*C and *iso*G could encode nonnatural amino acids without any competition with release factors and endogenous tRNAs.[17] This strategy may overcome the limitation of the genetic code, but the chemical synthesis of the artificial RNAs restricts its wide application.

Our group has been developing an alternative strategy. Instead of triplet, we applied four-base codons to assign the position of nonnatural amino acids. In this article, the details of this novel strategy and some related studies on extended codons are presented.

## Four-Base Codons

Our efforts to introduce nonnatural amino acids into specific positions of proteins have been focused on the development of novel decoding systems that use extended codon and anticodon pairs. First, we have developed a minor codon method.[18] Living organisms do not use 61 triplet codons equally rather, some particular codons are rarely used. In *E. coli*, for example, AGG, AGA, and CGG are rarely used to encode arginine. The concentration of tRNAs carrying anticodons corresponding to these minor codons has been revealed to be quite low.[19,20] This fact suggests that such minor codons could be used to encode nonnatural amino acids rather than natural amino acids, if an excess amount of tRNA, which is aminoacylated with a nonnatural amino acid and carries an anticodon complementary to the minor codon, is added to a translation system. The experiment was done for AGG codon, and we found that the minor codon was decoded as both a natural and an nonnatural amino acid. This result indicates that it is difficult to introduce nonnatural amino acids specifically.

To overcome the disadvantage of the minor codon method, we then developed a novel extended codon–anticodon pair. It is known that some kinds of frameshift suppressor strains have irregular tRNAs that contain extended four-base anticodons and decode four-base sequences as single codons.[21,22] This fact suggests that ribosomes accept the unusual tRNAs as substrates for protein synthesis. Hardesty and co-workers have reported that a $tRNA_{ACCU}$ (which is an abbreviation for tRNA containing ACCU anticodon) aminoacylated enzymatically with alanine decodes an AGGU codon.[23] The use of four-base codons instead of minor codons could eliminate the inappropriate incorporation of natural amino acid into a synthesized full-length protein through a triplet decoding by naturally occurring tRNAs.

The details of the four-base codon strategy are illustrated in Fig. 1. A mutated gene containing a four-base codon at a position for nonnatural amino acid incorporation is supplied to a translation system in the presence of an aminoacylated tRNA containing the corresponding four-base anticodon. When the four-base codon is decoded by the aminoacyl-tRNA with the corresponding four-base anticodon, the nonnatural amino acid attached to the tRNA is incorporated into the elongating peptide and, as a consequence, the correct reading frame of the translated gene is maintained. On the other hand, when the three letters of the four-base codon is decoded by the corresponding naturally occurring aminoacyl-tRNA, the cognate natural amino acid is incorporated and the reading frame is shifted. The resulting reading frame has a downstream stop codon, and the elongating protein that does not contain the nonnatural amino acid is truncated. In other words, the full-length protein must have the desired nonnatural amino acids at the position directed by the four-base codon.

As a four-base codon that assigns nonnatural amino acids, we have first chosen AGGU codon.[24] Similar to the minor codon strategy, the decoding of the four-base codon competes with naturally occurring aminoacyl-tRNAs. Although the triplet decoding of the four-base codon does not give a full-length protein, it will inhibit the four-base decoding and the incorporation of nonnatural amino acids. To avoid the triplet decoding, four-base codons derived from minor codons are preferable. As described above, AGG is seldom used in *E. coli*. Thus AGGU codon is expected to be preferentially decoded as a four-base codon rather than a triplet.

Experiments were carried out as shown in Fig. 2. Nitrophenylalanine as an example of nonnatural amino acids was derived to *N*-Boc (*t*-butoxycarbonyl)-nitrophenyalalnine cyanomethyl ester. The protected amino acid was attached to the 2′ or 3′ hydroxy group of a dinulceotide pdCpA,[25] which corresponds to the 3′ terminal of tRNAs. The tRNA has a nucleotide sequence of yeast tRNA for phenylalanine,[26] but has been mutated to contain an ACCU anticodon. The tRNA gene including a T7 promoter sequence was chemically synthesized. Then the tRNA(-CA) was prepared by using T7 RNA polymerase and the PCR-amplified tRNA gene lacking 3′ terminal dinucleotide. The aminoacylated pdCpA and the tRNA(-CA) was combined by using T4 RNA ligase to obtain the desired full-length aminoacylated tRNA. Mutated genes containing four-base codons were designed to include T7-tag and His-tag at N- and C-terminal, respectively. T7-tag is an 11 amino acid sequence and was utilized to detect proteins on Western blotting. His-tag is a histidine hexamer and was utilized to purify full-length proteins and to exclude truncated products that do not contain nonnatural amino acids. Streptavidin was chosen as a protein to examine the incorporation of nonnatural amino acids. The chemically

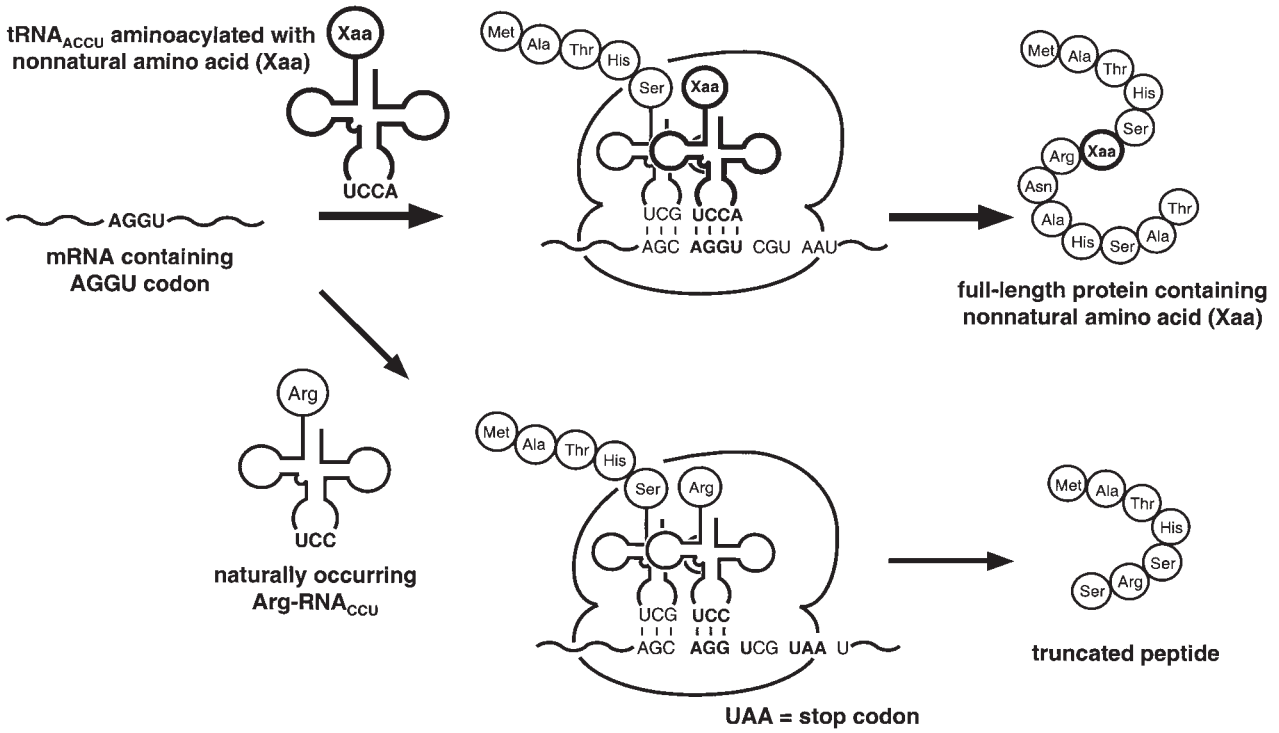


Fig. 1. Schematic illustration of four-base decoding. (Upper panel) Four-base decoding and incorporation of nonnatural amino acid. (Lower panel) Triplet decoding by a naturally occurring aminoacyl-tRNA, followed by termination of the peptide elongation.

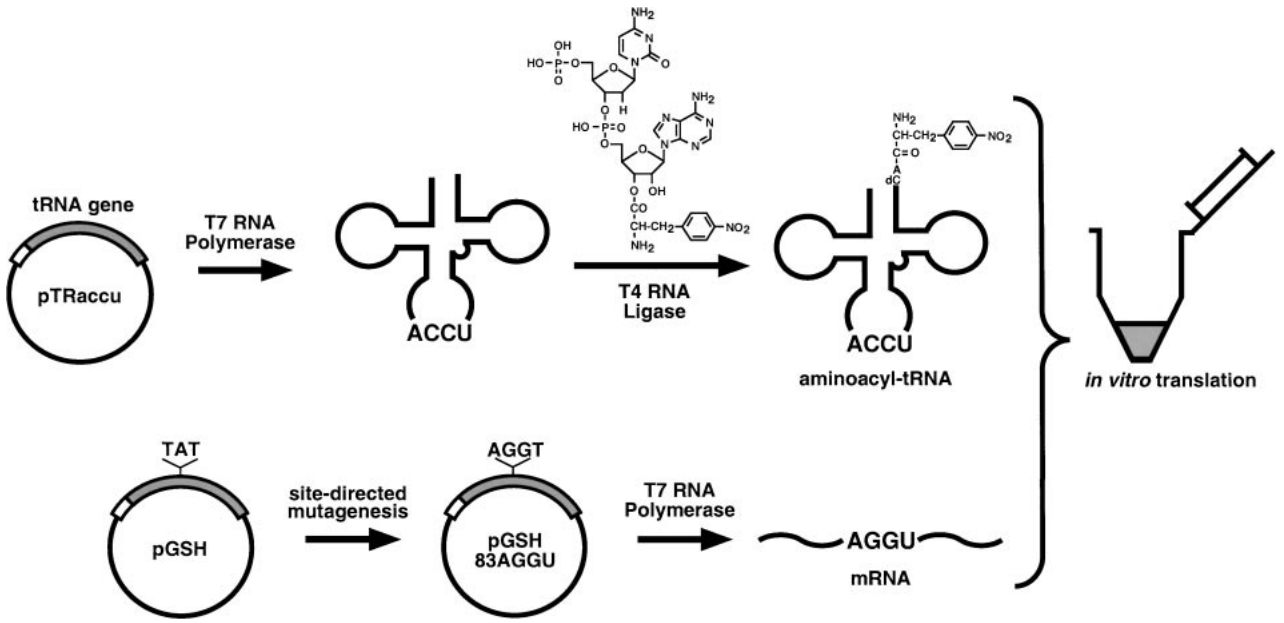


Fig. 2. Experimental scheme for synthesizing proteins containing nonnatural amino acids by using four-base codons.

synthesized streptavidin gene[27] was obtained commercially, and site-directed mutagenesis was carried out to introduce a four-base codon AGGU at its tyrosine 83 position. The wild-type and four-base codon mutant genes were cloned into an expression vector carrying T7 promoter. The mRNAs were transcribed from the constructed gene by using T7 RNA polymerase. Then the mRNA and aminoacyl-tRNA containing AGGU codon and ACCU anticodon, respectively, were added to an *E. coli* in vitro translation system consisting of an *E. coli* extract and supplements such as amino acids and ATP. The products were applied to an SDS-PAGE, followed by Western blot analysis using anti-T7-tag antibody.

As shown in Fig. 3, the wild-type full-length streptavidin was obtained as a band of 20 kDa. In the absence of tRNA$_{ACCU}$ and in the presence of non-aminoacylated tRNA$_{ACCU}$, the full-length protein was not observed, but a truncated peptide was obtained around 8 kDa. On the other hand, addition of tRNA$_{ACCU}$ aminoacylated with nitrophenylalanine to the in vi-

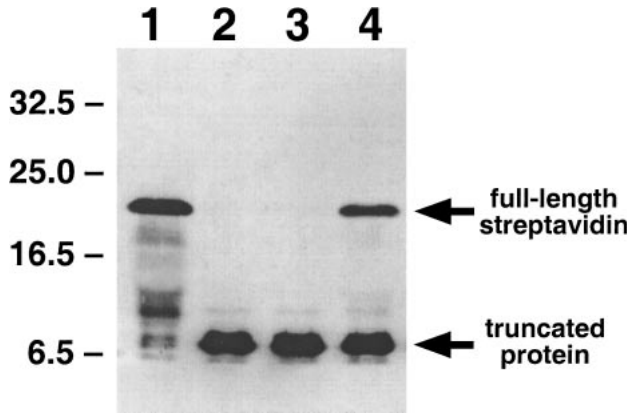


Fig. 3. Western blot analysis of an in vitro translation of streptavidin mRNA containing AGGU codon at its Tyr83 position. Lane 1, wild-type streptavidin mRNA. Lanes 2–4, Tyr83CGGG mRNA in the absence of $tRNA_{ACCU}$ (lane 2); in the presence of non-aminoacylated tRNA (lane 3); in the presence of 3-(*p*-nitrophenyl)alanyl-tRNA (lane 4).

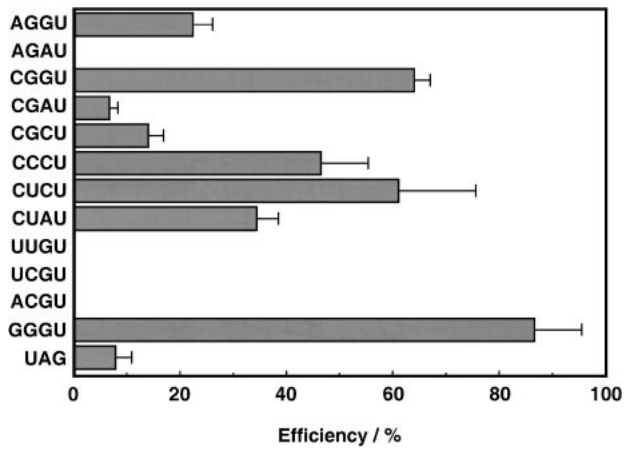


Fig. 4. Decoding of various four-base codons. Relative yields of the full-length streptavidin synthesized through a four-base decoding are represented as efficiencies of the four-base decoding.

tro translation reaction gave the full-length streptavidin. These results indicate that the aminoacylated $tRNA_{ACCU}$ successfully decodes the AGGU four-base codon, and introduces nitrophenylalanine to the desired position of streptavidin. The ineffectiveness of the non-aminoacylated tRNA supports the conclusion that the tRNA is not aminoacylated by any endogenous aminoacyl-tRNA synthetases.

Next, other four-codons were designed and examined.[28] Taking into consideration the codon usage and the copy number of tRNA gene in *E. coli* genome,[20] we investigated other 11 four-base codons. Arginine is coded by six codons: i.e., CGN (N indicates one of nucleotides A, G, C or U) and AGR (R indicates A or G), but CGU and CGC are primarily used. The gene of the corresponding $tRNA_{ICG}$ has four copies in the *E. coli* genome. On the other hand, CGG and AGA as well as AGG are rarely used, and their tRNA gene is a single. The above consideration suggests that the four-base codons derived from these minor codons could reduce serious competition with naturally occurring arginyl-tRNAs. At first, U was added to the codons, and the resulting four-base codons were examined for introducing nitrophenylalanine into Tyr83 position of streptavidin in the *E. coli* in vitro translation system. For comparison, CGAU and CGCU decoded by major $tRNA_{ICG}$ were also examined. It should be noted that the streptavidin has one CGC codon, which has been replaced by CGU codon when the CGCU codon is examined.

Four-base codon extension from triplet codons encoding leucine, serine, proline, threonine, and glycine were also considered. Leucine is coded by six codons as is the case for arginine. CUG codon decoded by $tRNA_{CAG}$ (four copies in the genome) is the most frequently used. On the other hand, the use of CUA, CUC, and UUG codons is infrequent; these are decoded by tRNAs that are coded by single copy genes. Thus CUA, CUC, and UUG were chosen for the four-base codon extension. Serine is also coded by six codons that are used with about the same frequency. For comparison with the case of arginine and leucine, UCG was picked in this study. Proline and glycine are coded by four codons, CCN and GGN, respectively. The most rarely used CCC and GGG codons were chosen. Threonine is also coded by four codons, ACN. Two tRNAs corresponding to ACA and ACG are minor, but ACA is used twice in the streptavidin gene. Therefore, ACG decoded by $tRNA_{CGU}$ was chosen. Valine and alanine are also encoded by four codons, but they were not taken up because all of them are frequently used.

The 12 four-base codons including AGGU were applied to the streptavidin expression assay. The four-base codons were introduced into the Tyr83 position of the streptavidin mRNA, and the nitrophenylalanyl-tRNAs containing the anticodons complementary to the four-base codons were prepared in a similar manner. The translation was carried out in the absence of tRNA, in the presence of non-aminoacylated tRNA, and in the presence of aminoacylated tRNA. The quantitative result of Western blot analysis is summarized in Fig. 4. Yields of the full-length streptavidin generated in the presence of the nitrophenylalanyl-tRNA were quantified by comparing the band intensity of the full-length streptavidin to those of serially diluted wild-type streptavidin, and these yields were represented as relative efficiencies of the four-base codon decoding.

Figure 4 indicates that a variety of four-base codons are available for highly efficient incorporation of nonnatural amino acids into proteins. Especially, the efficiencies of the GGGU (86%) and CGGU (64%) codons were extremly high compared to that of the amber codon (8%) in the same *E. coli* translation system. Efficiencies of the CCCU, CUCU, and CUAU codons were 46, 61, and 34%, respectively, which seem to be high enough for practical applications. Contrary to these codons, AGAU, UUGU, UCGU, and ACGU were ineffective.

The efficiency of the four-base decoding may be influenced by two major factors. First, the four-base decoding would be affected by the concentration of naturally occurring tRNAs. For example, the fact that the efficiency of the CGGU codon is higher than those of the CGCU and CGAU codons would reflect the concentration of $tRNA_{CCG}$ being lower than that of $tRNA_{ICG}$. Inability of the UUGU, UCGU, and ACGU codons suggests that relatively large amounts of naturally occurring tRNAs exist in the *E. coli* extract. The second factor may be interactions between four-base codons and anticodons. The four-

base codons with higher GC content are more efficiently decoded than those of lower GC. In the cases of the CGAU and CGCU codons, the efficiency of the CGCU codon is higher than that of the CGAU codon although they are decoded by the same $tRNA_{ICG}$. This difference may be simply due to a different strength of the interaction at the third position.

The fourth letter U was replaced by other nucleotides in the case of the four-base codons AGGU, CGGU, CUCU, and GGGU. As a result, all of them were accepted as four-base codons, but the efficiency depended on the fourth letter. In the case of AGGN, the AGGG codon was slightly more efficient. Similarly, the CGGG codon was slightly better in the CGGN series. In the case of GGGN, however, the GGGU codon was the best one but the GGGG codon was much less effective, possibly due to a higher-order structure of the GGGG sequence.

### Incorporation of Two Nonnatural Amino Acids into Single Proteins

The four-base codon strategy is more advantageous than the amber codon strategy thanks to its efficiency as described above, and thanks to its ability to incorporate multiple nonnatural amino acids into single proteins. These results indicate that proteins containing two nonnatural amino acids could be synthesized by using two four-base codons. To test this possibility, the streptavidin mRNA was designed to contain two four-base codons at two different positions.[29] As shown in Fig. 5a, the CGGG and AGGU codons were introduced into Tyr54 and Thr57 sites of streptavidin, respectively. If the CGGG codon is decoded by $Xaa^1$-$tRNA_{CCCG}$ and the AGGU codon is decoded by $Xaa^2$-$tRNA_{ACCU}$ ($Xaa^1$ and $Xaa^2$ indicate two different nonnatural amino acids), the full-length streptavidin containing the two nonnatural amino acids ($Xaa^1$ and $Xaa^2$) will be produced. If the CGGG codon is undesirably decoded as CGG triplet by endogenous arginyl-$tRNA_{CCG}$, the reading frame is unshifted, resulting in an encounter of UAG stop codon three codons downstream. Similarly, if the AGGU codon is decoded as AGG triplet by arginyl-$tRNA_{CCU}$, the protein synthesis will be stopped at UGA stop codon four codons downstream. In consequence, the full-length streptavidin must contain two nonnatural amino acids at desired positions.

Figure 5b shows the result of the double incorporation. The Western blot clearly indicates that the full-length streptavidin successfully synthesized only in the presence of both the aminoacyl-$tRNA_{CCCG}$ and aminoacyl-$tRNA_{ACCU}$ (lane 5). The efficiency of the double incorporation was quite low in this case (8% relative to the wild-type streptavidin), because the total efficiency was a product of the efficiency of the CGGG codon decoding and that of the AGGU codon decoding. As expected, however, the use of the GGGU codons, the most efficient four-base codon, in place of AGGU, greatly increased the double incorporation efficiency.[28]

The double incorporation system will be utilized for the construction of chemical systems on protein frameworks that require at least two chemical groups. Actually, the double incorporation technique was applied to study electron transfers on a single protein.[30] Fluorescence resonance energy transfer (FRET) is one of the most significant applications of the double incorporation. FRET-mediated structural and functional analyses of proteins are now in progress in our laboratory.

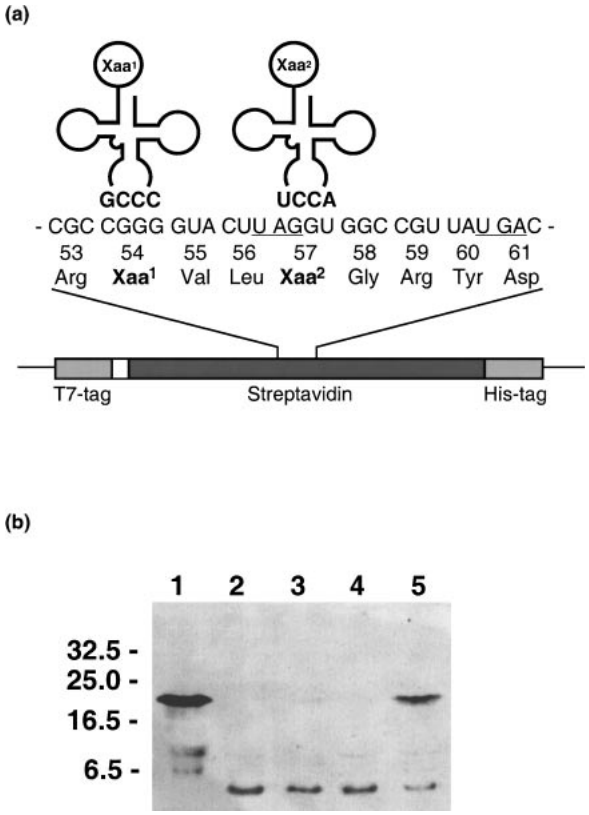


Fig. 5. Double incorporation of two nonnatural amino acids into single proteins. (a) Sequence of streptavidin mRNA containing CGGG and AGGU codons at its Tyr54 and Thr57 positions, respectively. (b) Western blot analysis of an in vitro translation. Lane 1, wild-type streptavidin mRNA. Lanes 2–4, Tyr54CGGG and Thr57AGGU mRNA in the absence of tRNA (lane 2); in the presence of NBD-lysyl-$tRNA_{CCCG}$ (lane 3); in the presence of 3-(2-naphthyl)alanyl-$tRNA_{ACCU}$ (lane 4); in the presence of both aminoacyl-tRNAs (lane 5).

### Five-Base Codons

The ability of protein synthesizing systems to decode four-base codons raises the possibility that even five-base codons could be decoded by aminoacyl-tRNAs containing five-base anticodons. The use of five-base codons has never been considered by any researchers for any living organisms. The examination will not only serve to extend the genetic code further, but also to study mechanisms of ribosomal frameshifting.

In order to evaluate five-base decoding, we uased the streptavidin expression system as in the case of the four-base codon. At first, a combination of a CGGUA codon and a UACCG anticodon was investigated.[31] Since CGG is a minor codon in *E. coli*, the use of CGG-derived five-base codons will avoid serious competition with endogenous arginyl-$tRNA_{CCG}$. The CGGUA codon was introduced into the Tyr54 position of the streptavidin gene and the UACCG anticodon was introduced into the anticodon of the tRNA (Fig. 6a). As is true for the four-base codons, when the CGGUA codon is decoded by the aminoacyl-$tRNA_{UACCG}$, the five-base codon will be translated into the nonnatural amino acid and a correct reading frame will

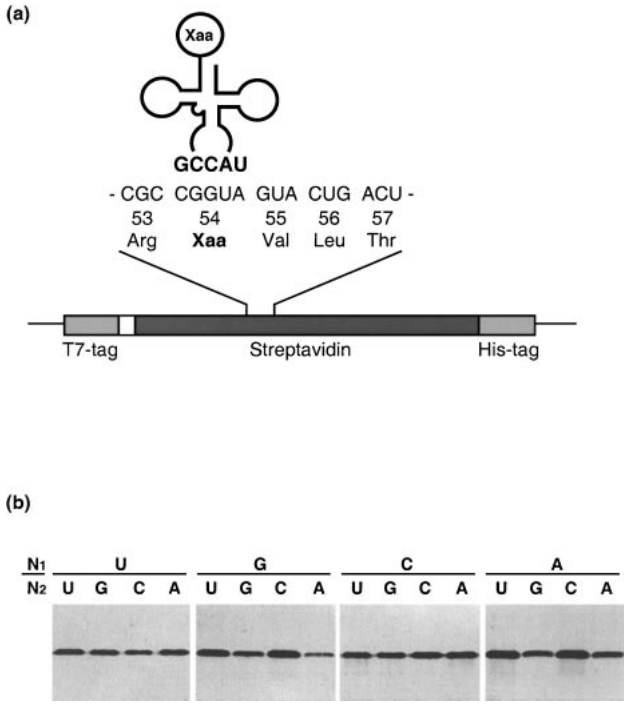


Fig. 6. Decoding of five-base codons. (a) Sequence of streptavidin mRNA containing a five-base codon at its Tyr54 position. (b) Western blot analysis of an in vitro translation supplied with mRNAs containing $CGGN_1N_2$ codons and 3-(*p*-nitrophenyl)alanyl-tRNAs containinig the corresponding five-base anticodons.

be maintained. However, when the first three bases of the CGGUA codon are decoded as a triplet codon by endogenous arginyl-$tRNA_{CCG}$, the translation will be terminated at a subsequent stop codon. As a result, the five-base decoding will give the full-length streptavidin containing the desired nonnatural amino acid at the desired position, whereas failure of the five-base decoding will give the truncated protein.

The mRNA containing the CGGUA codon at the Tyr54 position of the streptavidin gene and the $tRNA_{UACCG}$ chemically aminoacylated with nitrophenylalanine were added to the *E. coli* in vitro translation system. Then, the reaction mixture was analyzed by Western bloting. As a result, in the presence of nitrophenylalanyl-$tRNA_{UACCG}$, the full-length streptavidin was synthesized. The yield of the full-length streptavidin increased with increasing the concentration of the aminoacyl-$tRNA_{UACCG}$. However, in the absence of $tRNA_{UACCG}$ and in the presence of non-aminoacylated $tRNA_{UACCG}$, only a truncated product was observed instead of full-length product. These results indicate that the CGGUA is successfully decoded as a five-base codon by the aminoacyl-$tRNA_{UACCG}$.

The five-base decoding system may produce up to 16 extended codons from a single three-base codon. To test this possibility, streptavidin mRNAs containing five-base codons $CGGN_1N_2$, where $N_1$ and $N_2$ indicate one of four bases, at the Tyr54 position were added to the translation mixture together with nitrophenylalanyl-tRNAs containing the complementary five-base anticodons. Western blot analysis (Fig. 6b) showed that all combinations of the complementary five-base codon–anticodon pairs gave the full-length streptavidin. On the other hand, negligible amounts of the full-length streptavidin were given in the presence of non-aminoacylated tRNAs. These results indicate that all of the 16 five-base codons can be successfully decoded by the aminoacylated tRNAs containing the complementary five-base anticodons. The efficiency of five-base decoding was estimated from the band intensity on the Western blot. The highest yield of full-length streptavidin was observed in the case of the CGGAC codon (42% relative to wild-type streptavidin). This efficiency is lower than that of the four-base codons. We also found that the five-base anticodon could decode the corresponding four-base codon without using the first letter of the anticodon. Taken together, a five-base decoding model is proposed, in which ribosome moves not through a five-base translocation but through a triplet translocation followed by duplicated +1 frameshifting.[31]

These results demonstrate for the first time that the five-base codons can be decoded by using aminoacyl-tRNAs containing five-base anticodons. This novel frameshift strategy will find wide applications, especially multiple incorporation of nonnatural amino acids into single proteins. This study also shows the feasibility and utility of the application of chemically aminoacylated tRNAs. In previous studies of frameshift suppression, suppressor tRNAs must be aminoacylated by any one of the aminoacyl-tRNA synthetases. Most aminoacyl-tRNA synthetases, however, recognize anticodons as identity determinants, limiting the mutagenesis of the anticodon of frameshift suppressor tRNAs. Since the chemical aminoacylation method is free from enzymatic recognition, it will serve to probe structure-function relationships of unusual tRNAs such as frameshift suppressor tRNAs, mitchondorial tRNAs, and tmRNAs.[32]

### Extension of Amino Acid Repertoire

In addition to the extension of the genetic code, we have explored the extension of the amino acid repertoire that can be accepted as substrates for translation systems. Results show that the incorporation efficiency depends sharply on the side groups of nonnatural amino acids. For instance, 3-(*p*-benzoylphenyl)-alanine has been successfully incorporated into T4 lysozyme in good yield, whereas the incorporation of $N^{\varepsilon}$-(dansyl)lysine was unsuccessful.[33] We evaluated the incorporation of nonnatural amino acids with various aromatic side groups as listed in Fig. 7.[34] The incorporation of these nonnatural amino acids into the Tyr83 position of the streptavidin was examined by using a CGGG–CCCG codon–anticodon pair. The translation products were analyzed by Western blot analysis and the incorporation efficiency was determined from the band intensity.

The result summarized in Fig. 8 clearly shows that the incorporation efficiency markedly depends on the structure of amino acids. For example, some aromatic nonnatural amino acids such as 3-(2-naphthyl)alanine **2** and 3-(4-biphenylyl)alanine **3** are efficiently incorporated. However, 3-(9-phenanthryl)alanine **6** and 3-(1-pyrenyl)alanine **7** are rejected. In comparison to 3-(2-naphthyl)alanine **2** (72%), the incorporation of 3-(1-naphthyl)alanine **1** was less efficient (30%). Similarly, 3-(2-anthryl)alanine **4** (45%) and 3-(2-pyrenyl)alanine **8** (10%) were more efficiently incorporated than 3-(9-anthryl)alanine **5** (less than 2%) and 3-(1-pyrenyl)alanine **7** (3%), respectively. These results suggest that the protein biosynthesizing system discriminates nonnatural amino acids not by their size but by their

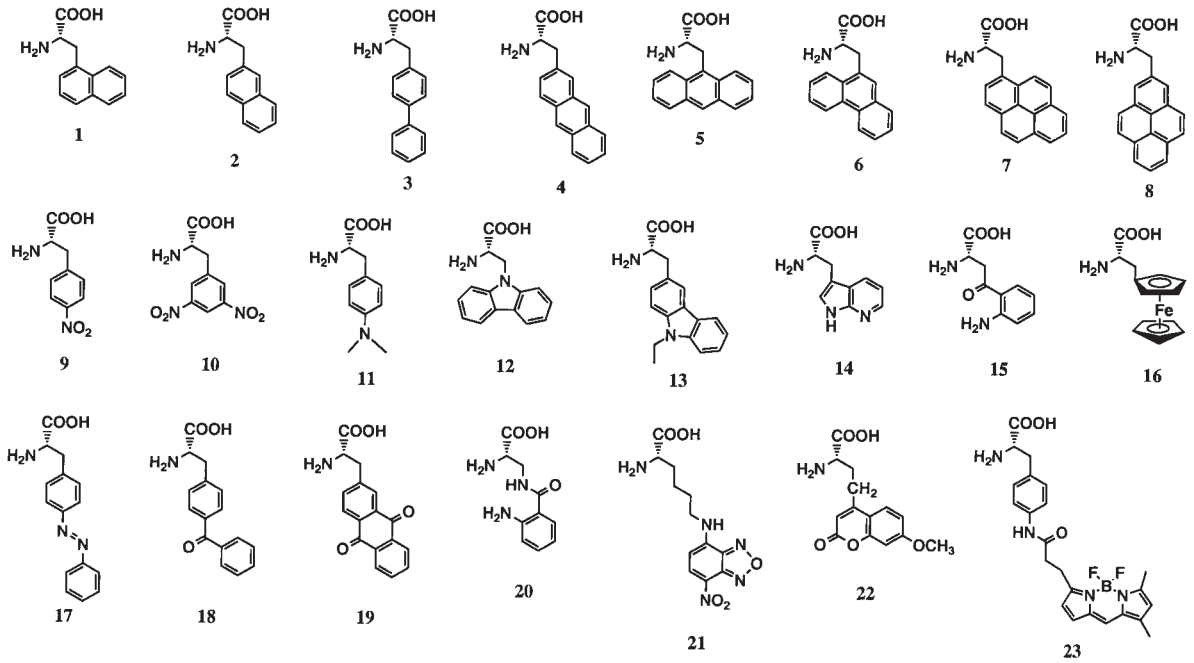


Fig. 7. Structure of nonnatural amino acids examined for incorporation into proteins by using four-base codons.

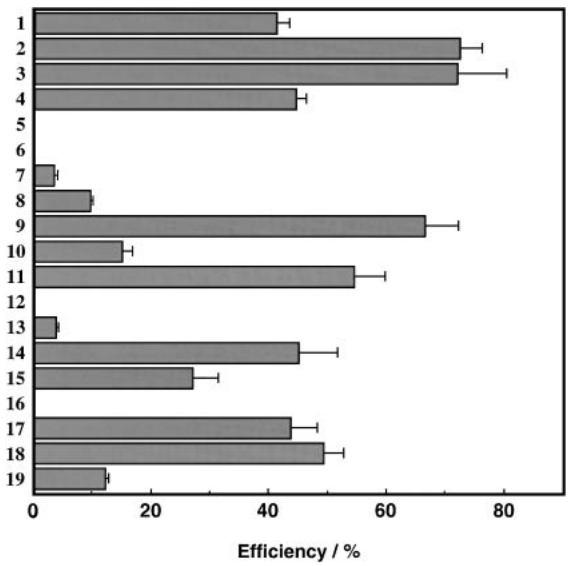


Fig. 8. Incorporation of various aromatic nonnatural amino acids into proteins by using CGGG four-base codon in an *E. coli* in vitro translation system. Relative yields of the full-length streptavidin synthesized in the presence of the $tRNA_{CCCG}$ aminoacylated with amino acids **1**–**19** are represented as efficiencies of the incorporation of these amino acids.

shape. It appears that amino acids with linearly extended aromatic groups such as 3-(2-naphthyl)alanine **2**, 3-(4-biphenylyl)alanine **3**, 3-(2-anthryl)alanine **4**, 3-(*p*-phenylazophenyl)-alanine **17**, and 3-(*p*-benzoylphenyl)alanine **18** are favored, while those with rather widely expanded or bent aromatic groups such as 3-(9-anthryl)alanine **5**, 3-(9-phenanthryl)alanine **6**, and 3-(9-carbazolyl)alanine **12** are rejected.

The amino acid selectivity will serve to design novel nonnatural amino acids that can be incorporated in the translation system. Relatively large but linearly extended nonnatural amino acids with fluorescent probes **20**–**22** were designed and incorporated into proteins.[35,36] Recently, we have developed fluorescently labeled nonnatural amino acids that can be excited with visible light. BODIPY FL-labeled 3-(*p*-aminophenyl)alanine **23** attached to the $tRNA_{CCCG}$ was successfully incorporated into specific positions of proteins, whereas the incorporation of fluorescein-labeled aminophenylalanine was not observed.[37] Compared with chemical modification method that has been used to prepare fluorescently labeled proteins, our novel labeling method enables scientists to prepare proteins labeled site-specifically and quantitatively. This labeling method will be serve as a powerful tool for structural and functional analysis of proteins.

## Concluding Remarks

A novel method using extended codons has been developed to realize efficient and multiple incorporation of nonnatural amino acids into desired positions of proteins. By using four-base codon strategy, we can obtain proteins containing nonnatural amino acids with higher yield than with the amber suppression strategy. Moreover, double incorporation of two different nonnatural amino acids into single proteins is first achieved by the four-base codon strategy. Various nonnatural amino acids so far have been incorporated into proteins including fluorescently labeled nonnatural amino acids. Combination of the double incorporation technique and the fluorescently labeled nonnatural amino acids will give a valuable tool for protein analysis by fluorescence resonance energy transfer (FRET), which is now in progress in our laboratory.

Recently, mutated aminoacyl-tRNA synthetases recognizing some nonnatural amino acids were developed to incorporate nonnatural amino acids into proteins in living cells.[38] Although this in vivo translation system gave a larger amount of nonna-

tural proteins than in vitro systems, the in vitro systems have the advantages of high through-put and wide amino acid repertoire. Methodological improvement of aminoacylation of tRNAs with nonnatural amino acids will serve to increase the yield of nonnatural proteins in vitro.

The extended codon strategy also implies that there are possibilities of extension of natural biomolecules and biomolecular systems. Extension of the genetic code is an example of these possibilities. Extension of protein architectures through the nonnatural amino acid incorporation can be utilized to investigate structure and function of proteins and to design and synthesize artificial proteins. By adapting functionalized proteins to living cells, extension of cell functions will be also expected. Extended biomolecules and biomolecular systems will make great contributions to future bioscience and biotechnology.

I would like to express my sincere gratitude to Professor Masahiko Sisido for his guidance and discussion, and to all collaborators in Sisido laboratory at Okayama University for their efforts in contributing to this study.

**References**

1 K. M. Ulmer, *Science*, **219**, 666 (1983).
2 M. J. Wilson and D. L. Hatfield, *Biochim. Biophys. Acta*, **781**, 205 (1984).
3 C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, and P. G. Schultz, *Science*, **244**, 182 (1989).
4 J. D. Bain, C. G. Glabe, T. A. Dix, A. R. Chamberlin, and E. S. Diala, *J. Am. Chem. Soc.*, **111**, 8013 (1989).
5 T. G. Heckler, Y. Zama, T. Naka, and S. M. Hecht, *J. Biol. Chem.*, **258**, 4492 (1983).
6 J. Ellman, D. Mendel, S. J. Anthony-Cahill, C. J. Noren, and P. G. Schultz, *Methods Enzymol.*, **202**, 301 (1991).
7 V. W. Cornish, D. Mendel, and P. G. Schultz, *Angew. Chem., Int. Ed. Engl.*, **34**, 621 (1995).
8 M. A. Gillmore, L. E. Steward, and A. R. Chamberlin, *Top. Curr. Chem.*, **202**, 77 (1999).
9 T. Arslan, S. V. Mamaev, N. V. Mamaeva, and S. M. Hecht, *J. Am. Chem. Soc.*, **119**, 10877 (1997).
10 B. Wang, K. C. Brown, M. Lodder, C. S. Craik, and S. M. Hecht, *Biochemistry*, **41**, 2805 (2002).
11 D. A. Dougherty, *Curr. Opin. Chem. Biol.*, **4**, 645 (2000).
12 D. L. Beene, D. A. Dougherty, and H. A. Lester, *Curr. Opin. Neurobiol.*, **13**, 264 (2003).
13 Y. Kimata, H. Shimada, T. Hirose, and Y. Ishimura, *Biochem. Biophys. Res. Commun.*, **208**, 96 (1995).
14 T. Kanamori, S. Nishikawa, M. Nakai, I. Shin, P. G. Schultz, and T. Endo, *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 3634 (1999).
15 T. Hohsaka and M. Sisido, *Curr. Opin. Chem. Biol.*, **6**, 809 (2002).
16 S. T. Cload, D. R. Liu, W. A. Froland, and P. G. Schultz, *Chem. Biol.*, **3**, 1033 (1996).
17 J. D. Bain, C. Switzer, R. A. Chamberlin, and S. A. Benner, *Nature*, **356**, 537 (1992).
18 T. Hohsaka, K. Sato, M. Sisido, K. Takai, and S. Yokoyama, *FEBS Lett.*, **344**, 171 (1994).
19 T. Ikemura, *J. Mol. Biol.*, **151**, 389 (1981).
20 Y. Komine, T. Adachi, H. Inokuchi, and H. Ozeki, *J. Mol. Biol.*, **212**, 579 (1990).
21 D. Riddle and J. Carbon, *Nat. New Biol.*, **242**, 230 (1973).
22 R. F. Gaber and M. R. Culbertson, *Gene*, **19**, 163 (1982).
23 C. Ma, W. Kudlicki, O. W. Odom, G. Kramer, and B. Hardesty, *Biochemistry*, **32**, 7939 (1993).
24 T. Hohsaka, Y. Ashizuka, H. Murakami, and M. Sisido, *J. Am. Chem. Soc.*, **118**, 9778 (1996).
25 S. A. Robertson, J. A. Ellman, and P. G. Schultz, *J. Am. Chem. Soc.*, **113**, 2722 (1991).
26 J. R. Sampson and O. C. Uhlenbeck, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 1033 (1988).
27 L. D. Thompson and P. C. Weber, *Gene*, **136**, 243 (1993).
28 T. Hohsaka, Y. Ashizuka, H. Taira, H. Murakami, and M. Sisido, *Biochemistry*, **40**, 11060 (2001).
29 T. Hohsaka, Y. Ashizuka, H. Sasaki, H. Murakami, and M. Sisido, *J. Am. Chem. Soc.*, **121**, 12194 (1991).
30 M. Taki, T. Hohsaka, H. Murakami, K. Taira, and M. Sisido, *J. Am. Chem. Soc.*, **124**, 14586 (2002).
31 T. Hohsaka, Y. Ashizuka, H. Murakami, and M. Sisido, *Nucleic Acids Res.*, **29**, 3646 (2001).
32 A. Muto, C. Ushida, and H. Himeno, *Trends Biochem. Sci.*, **23**, 25 (1998).
33 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).
34 T. Hohsaka, D. Kajihara, Y. Ashizuka, H. Murakami, and M. Sisido, *J. Am. Chem. Soc.*, **121**, 34 (1999).
35 H. Murakami, T. Hohsaka, Y. Ashizuka, K. Hashimoto, and M. Sisido, *Biomacromolecules*, **1**, 118 (2000).
36 M. Taki, T. Hohsaka, H. Murakami, K. Taira, and M. Sisido, *FEBS Lett.*, **507**, 35 (2001).
37 T. Hohsaka, R. Abe, and M. Sisido, in preparation.
38 L. Wang and P. G. Schultz, *Chem. Commun.*, **2002**, 1.

Takahiro Hohsaka was born in Nagano in 1970. He graduated from the Department of Biomolecular Engineering, Faculty of Engineering, Tokyo Institute of Technology in 1992. He received the M.Eng. (1994), and the Ph.D. (2002) from Tokyo Institute of Technology. He was appointed as a research associate at Faculty of Engineering, Okayama University in 1994, then as an associate professor of School of Materials Science, Japan Advanced Institute of Science and Technology in 2003. Since 2002, he has been a PRESTO researcher of Japan Science and Technology Agency (JST). He received the CSJ Award for Young Chemists for 2002 from the Chemical Society of Japan and the Gold Medal from Tokyo Techno-Forum 21 in 2003. His current researches have focused on the development of novel biotechnologies by using engineered biomolecules and biomolecular systems.