Ribosomal synthesis of drug-like non-standard peptides

Natural product peptides often contain multiple *N*-methylations on their backbones, *N*-acylated amino acids or cyclic structures closed by a linkage other than a disulfide bond. These structural features contribute to increases in structural rigidity, proteolytic stability, and membrane permeability. With this in mind, we next applied our methodology to the synthesis of natural product-like non-standard peptides.

To achieve this goal, we recently developed a method to reprogram the translation initiation by depleting Met from ϖPURE system and replacing it with an aminoacyl-$tRNA^{fMet}_{CUA}$ prepared by flexizymes (Figure 4a). Surprisingly, we found that virtually any proteinogenic amino acids could be used as an initiator for translation (Goto *et al.*, *ACS Chem Biol*, in press). Particularly, we were pleased to find that a variety of *N*-acyl groups, such as fatty acids and pyroglutamate, could be added to the initiator amino acids designated by the reprogramming. When one of such acyl groups, 2-chloroacyl group (ClAc), attached to an amino acid was used for the initiation and a single cysteine residue (Cys) was included in the elongating chain, spontaneous cyclization occurred through reaction between the ClAc and sulfhydryl group on Cys, forming a physiologically stable thioether bond (Figure 4b). The combination of this cyclization technology with the incorporation of α-*N*-methyl-amino acids into the nascent peptide chain enabled us to synthesize a wide array of macrocyclic peptides bearing *N*-methylated backbones (Kawakami *et al.*, *Chem Biol*, in press), one such example being shown in Figure 4b.

Conclusions and perspectives

Owing to the fact that the technology is still under development, conclusive answers to the questions raised earlier (vide supra) cannot be given yet; however, at the current stage we are able to offer some insights. Regarding the first question, we generally observe that the expression level of non-standard peptides containing up to four non-proteinogenic amino acids is comparable to that of wildtype expression, generally in the range of 5–10 μM (nmol/mL). However, when the number of non-proteinogenic amino acids is increased, particularly in the use of exotic amino acids, the expression level is reduced to sub-μM range (sub-nmol/mL). Regarding the second question, it is possible to reassign 8–10 codons to non-proteinogenic amino acids by our method, and up to 13 codons by the Szostak's method [35••]. Unfortunately, the expression level of non-standard peptides becomes lower with an increase in their lengths. Although we have witnessed significant improvements in these two areas from recent studies, there remain certain limitations, and it is clear that improvements in the expression level by further engineering of the translation system are required for future technical development [51–54]. Regarding the last question, the incorporation efficiency of non-proteinogenic amino acids such as α-*N*-methyl-amino acids tends to be better for the genetic code reprogramming method than the amber suppression method, probably because the competition that causes the reduction in efficiency is repressed to a minimum [11,44]. However, more thorough investigations of exotic amino acids such as β-, γ- and D-amino acids, which have been thus far largely unsuccessful, are awaited [15,44,35••,42••].

It has become clear now that genetic code reprogramming enables us to synthesize non-standard peptides in a template-dependent manner. This technology is currently being extended to the screening of non-standard peptides against therapeutic targets. Particularly, its coupling with suitable *in vitro* display methods, would become a powerful screening strategy to accelerate the discovery of peptidic aptamers [55•,56]. We expect that exciting results of this type should be forthcoming from those laboratories pursuing genetic code reprogramming.

Acknowledgements

We thank Dr PC Reid for discussion and proof reading. This work was supported by grants of National Institute of Health (GM59159), Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (S) (16101007) and a research and development project of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (NEDO) to HS, and AO was supported by Grants-in-Aid for JSPS Fellows (19-1722).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Srinivasan G, James CM, Krzycki JA: **Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA.** *Science* 2002, **296**:1459-1462.
2. Stadtman TC: **Selenocysteine.** *Annu Rev Biochem* 1996, **65**:83-100.
3. Grunewald J, Marahiel MA: **Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides.** *Microbiol Mol Biol Rev* 2006, **70**:121-146.
4. Kohli RM, Walsh CT, Burkart MD: **Biomimetic synthesis and optimization of cyclic peptide antibiotics.** *Nature* 2002, **418**:658-661.
5. Kopp F, Mahlert C, Grunewald J, Marahiel MA: **Peptide macrocyclization: the reductase of the nostocyclopeptide synthetase triggers the self-assembly of a macrocyclic imine.** *J Am Chem Soc* 2006, **128**:16478-16479.
6. Bain JD, Glabe CG, Dix TA, Chamberlin AR: **Biosynthetic site-specific incorporation of a non-natural amino acid *in vitro* a polypeptide.** *J Am Chem Soc* 1989, **111**:8013-8014.
7. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG: **A general method for site-specific incorporation of unnatural amino acids into proteins.** *Science* 1989, **244**:182-188.
8. Bain JD, Diala ES, Glabe CG, Wacker DA, Lyttle MH, Dix TA, Chamberlin AR: **Site-specific incorporation of nonnatural residues during *in vitro* protein biosynthesis with semisynthetic aminoacyl-tRNAs.** *Biochemistry* 1991, **30**:5411-5421.

9. Chung HH, Benson DR, Schultz PG: **Probing the structure and mechanism of Ras protein with an expanded genetic-code.** *Science* 1993, **259**:806-809.

10. Ellman J, Mendel D, Anthony-Cahill S, Noren CJ, Schultz PG: **Biosynthetic method for introducing unnatural amino acids site-specifically into proteins.** *Methods Enzymol* 1991, **202**:301-336.

11. Ellman JA, Mendel D, Schultz PG: **Site-specific incorporation of novel backbone structures into proteins.** *Science* 1992, **255**:197-200.

12. Hendrickson TL, de Crecy-Lagard V, Schimmel P: **Incorporation of nonnatural amino acids into proteins.** *Annu Rev Biochem* 2004, **73**:147-176.

13. Mendel D, Ellman J, Schultz PG: **Protein-biosynthesis with conformationally restricted amino-acids.** *J Am Chem Soc* 1993, **115**:4359-4360.

14. Wang L, Xie J, Schultz PG: **Expanding the genetic code.** *Annu Rev Biophys Biomol Struct* 2006, **35**:225-249.

15. Sando S, Abe K, Sato N, Shibata T, Mizusawa K, Aoyama Y: **Unexpected preference of the *E. coli* translation system for the ester bond during incorporation of backbone-elongated substrates.** *J Am Chem Soc* 2007, **129**:6180-6186.

16. Heckler TG, Chang LH, Zama Y, Naka T, Chorghade MS, Hecht SM: **T4 RNA ligase mediated preparation of novel 'chemically misacylated' tRNAPheS.** *Biochemistry* 1984, **23**:1468-1473.

17. Heckler TG, Roesser JR, Xu C, Chang PI, Hecht SM: **Ribosomal binding and dipeptide formation by misacylated tRNA(Phe)S.** *Biochemistry* 1988, **27**:7254-7262.

18. Link AJ, Tirrell DA: **Reassignment of sense codons *in vivo*.** *Methods* 2005, **36**:291-298.

19. Wang L, Brock A, Herberich B, Schultz PG: **Expanding the genetic code of *Escherichia coli*.** *Science* 2001, **292**:498-500.

20. Xie JM, Schultz PG: **An expanding genetic code.** *Methods* 2005, **36**:227-238.

21. Liu DR, Magliery TJ, Pasternak M, Schultz PG: **Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*.** *Proc Natl Acad Sci U S A* 1997, **94**:10092-10097.

22. Rodriguez EA, Lester HA, Dougherty DA: **Improved amber and opal suppressor tRNAs for incorporation of unnatural amino acids *in vivo*. Part 1: minimizing misacylation.** *Rna—Publ Rna Soc* 2007, **13**:1703-1714.

23. Rodriguez EA, Lester HA, Dougherty DA: **Improved amber and opal suppressor tRNAs for incorporation of unnatural amino acids *in vivo*. Part 2: evaluating suppression efficiency.** *Rna—Publ Rna Soc* 2007, **13**:1715-1722.

24. Hohsaka T, Ashizuka Y, Murakami H, Sisido M: **Incorporation of nonnatural amino acids into streptavidin through *in vitro* frame-shift suppression.** *J Am Chem Soc* 1996, **118**:9778-9779.

25. Ninomiya K, Minohata T, Nishimura M, Sisido M: ***In situ* chemical aminoacylation with amino acid thioesters linked to a peptide nucleic acid.** *J Am Chem Soc* 2004, **126**:15984-15989.

26. Zhang Z, Alfonta L, Tian F, Bursulaya B, Uryu S, King DS, Schultz PG: **Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells.** *Proc Natl Acad Sci U S A* 2004, **101**:8882-8887.

27. Kohrer C, Yoo JH, Bennett M, Schaack J, RajBhandary UL: **A possible approach to site-specific insertion of two different unnatural amino acids into proteins in mammalian cells via nonsense suppression.** *Chem Biol* 2003, **10**:1095-1102.

28. Forster AC, Tan Z, Nalam MN, Lin H, Qu H, Cornish VW, Blacklow SC: **Programming peptidomimetic syntheses by translating genetic codes designed *de novo*.** *Proc Natl Acad Sci U S A* 2003, **100**:6353-6357.

29. Tan Z, Blacklow SC, Cornish VW, Forster AC: ***De novo* genetic codes and pure translation display.** *Methods* 2005, **36**:279-290.

30. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T: **Cell-free translation reconstituted with purified components.** *Nat Biotechnol* 2001, **19**:751-755.

31. Shimizu Y, Kanamori T, Ueda T: **Protein synthesis by pure translation systems.** *Methods* 2005, **36**:299-304.

32. Hartman MC, Josephson K, Szostak JW: **Enzymatic aminoacylation of tRNA with unnatural amino acids.** *Proc Natl Acad Sci U S A* 2006, **103**:4356-4361.

33. • Josephson K, Hartman MC, Szostak JW: **Ribosomal synthesis of unnatural peptides.** *J Am Chem Soc* 2005, **127**:11727-11735.

Using the PURE system supplemented with non-proteinogenic amino acids, the authors demonstrated mRNA-directed non-standard polypeptide synthesis containing 10 different non-proteinogenic amino acids.

34. Seebeck FP, Szostak JW: **Ribosomal synthesis of dehydroalanine-containing peptides.** *J Am Chem Soc* 2006, **128**:7150-7151.

35. •• Hartman MC, Josephson K, Lin CW, Szostak JW: **An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides.** *PLoS ONE* 2007, **2**:e972.

Ninety-nine non-proteinogenic amino acids were examined for their ability to be added onto tRNAs by aaRS in the *w*PURE system and incorporated into a specific site. It was shown that non-proteinogenic amino acids with various non-natural sidechains were incorporated, while backbone-modifying ones showed modest incorporation efficiencies.

36. Saito H, Kourouklis D, Suga H: **An *in vitro* evolved precursor tRNA with aminoacylation activity.** *EMBO J* 2001, **20**:1797-1806.

37. Hodgson DR, Suga H: **Mechanistic studies on acyl-transferase ribozymes and beyond.** *Biopolymers* 2004, **73**:130-150.

38. Kourouklis D, Murakami H, Suga H: **Programmable ribozymes for mischarging tRNA with nonnatural amino acids and their applications to translation.** *Methods* 2005, **36**:239-244.

39. Murakami H, Kourouklis D, Suga H: **Using a solid-phase ribozyme aminoacylation system to reprogram the genetic code.** *Chem Biol* 2003, **10**:1077-1084.

40. Murakami H, Saito H, Suga H: **A versatile tRNA aminoacylation catalyst based on RNA.** *Chem Biol* 2003, **10**:655-662.

41. Ramaswamy K, Saito H, Murakami H, Shiba K, Suga H: **Designer ribozymes: programming the tRNA specificity into flexizyme.** *J Am Chem Soc* 2004, **126**:11454-11455.

42. •• Murakami H, Ohta A, Ashigai H, Suga H: **A highly flexible tRNA acylation method for non-natural polypeptide synthesis.** *Nat Methods* 2006, **3**:357-359.

This article reports a highly flexible *de novo* tRNA acylation system consisting of ribozymes (flexizymes, eFx and dFx). This system enables the charging of virtually any non-proteinogenic amino acids and hydroxy acids. This system was also applied to mRNA-programmed non-standard polypeptide synthesis.

43. Ohuchi M, Murakami H, Suga H: **The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus.** *Curr Opin Chem Biol* 2007.

44. Tan Z, Forster AC, Blacklow SC, Cornish VW: **Amino acid backbone specificity of the *Escherichia coli* translation machinery.** *J Am Chem Soc* 2004, **126**:12752-12753.

45. Zhang B, Tan Z, Dickson LG, Nalam MN, Cornish VW, Forster AC: **Specificity of translation for *N*-alkyl amino acids.** *J Am Chem Soc* 2007, **129**:11316-11317.

46. Merryman C, Green R: **Transformation of aminoacyl tRNAs for the *in vitro* selection of 'drug-like' molecules.** *Chem Biol* 2004, **11**:575-582.

47. Frankel A, Millward SW, Roberts RW: **Encodamers: unnatural peptide oligomers encoded in RNA.** *Chem Biol* 2003, **10**:1043-1050.

48. England PM, Zhang Y, Dougherty DA, Lester HA: **Backbone mutations in transmembrane domains of a ligand-gated ion channel: implications for the mechanism of gating.** *Cell* 1999, **96**:89-98.

49. Fahnestock S, Rich A: **Ribosome-catalyzed polyester formation.** *Science* 1971, **173**:340-343.

50. Ohta A, Murakami H, Higashimura E, Suga H: **Synthesis of polyester by means of genetic code reprogramming.** *Chem Biol* 2007, **14**:1315-1322.

51. Doi Y, Ohtsuki T, Shimizu Y, Ueda T, Sisido M: **Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system.** *J Am Chem Soc* 2007, **129**:14458-14462.

52. Dedkova LM, Fahmi NE, Golovine SY, Hecht SM: **Enhanced D-amino acid incorporation into protein by modified ribosomes.** *J Am Chem Soc* 2003, **125**:6616-6617.

53. Dedkova LM, Fahmi NE, Golovine SY, Hecht SM: **Construction of modified ribosomes for incorporation of D-amino acids into proteins.** *Biochemistry* 2006, **45**:15541-15551.

54. Wang K, Neumann H, Peak-Chew SY, Chin JW: **Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion.** *Nat Biotechnol* 2007, **25**:770-777.

55. • Millward SW, Fiacco S, Austin RJ, Roberts RW: **Design of cyclic peptides that bind protein surfaces with antibody-like affinity.** *ACS Chem Biol* 2007, **2**:625-634.
Using a bis-NHS cross-linking reagent to join the N-terminus to an internal lysine residue of peptide, a trillion-member library of cyclic peptides was prepared in the mRNA display format. After *in vitro* selection against signaling protein Gαi1, a high affinity cyclic peptide ligand with K_d = 2.1 nM was identified.

56. Ja WW, West AP Jr, Delker SL, Bjorkman PJ, Benzer S, Roberts RW: **Extension of Drosophila melanogaster life span with a GPCR peptide inhibitor.** *Nat Chem Biol* 2007, **3**: 415-419.

MINIREVIEW / MINISYNTHÈSE

Ribosomal synthesis of nonstandard peptides[1]

Taek Jin Kang and Hiroaki Suga

Abstract: It is well known that standard peptides, which comprise proteinogenic amino acids, can act as specific chemical probes to target proteins with high affinity. Despite this fact, a number of peptide drug leads have been abandoned because of their poor cell permeability and protease instability. On the other hand, nonstandard peptides isolated as natural products often exhibit remarkable pharmaco-behavior and stability in vivo. Although it is likely that numerous nonstandard therapeutic peptides capable of recognizing various targets could have been synthesized, enzymes for nonribosomal peptide syntheses are complex; therefore, it is difficult to engineer such modular enzymes to build nonstandard peptide libraries. Here we describe an emerging technology for the synthesis of nonstandard peptides that employs an integrated system of reconstituted cell-free translation and flexizymes. We summarize the historical background of this technology and discuss its current and future applications to the synthesis of nonstandard peptides and drug discovery.

Key words: nonstandard peptide, misaminoacylation, therapeutic peptide, translation.

Résumé : Il est bien connu que les peptides standards, qui comprennent les acides aminés protéinogènes, peuvent agir comme sondes chimiques spécifiques pour cibler des protéines avec une haute affinité. Malgré cela, plusieurs prototypes de drogues peptidiques ont été abandonnés à cause de leur faible perméabilité cellulaire et à leur sensibilité vis-à-vis les protéases. Cependant, des peptides non standards isolés comme produits naturels font souvent preuve de qualités pharmacologiques uniques et de stabilité *in vivo*. Quoiqu'il soit probable que plusieurs peptides thérapeutiques non standards capables de reconnaître différentes cibles puissent être synthétisés, les enzymes de synthèse de peptides non ribosomiques sont complexes; il est alors difficile de concevoir de telles enzymes modulaires pour construire des banques de peptides non standards. Dans cet article, nous décrivons une technologie en émergence qui permet de synthétiser des peptides non standards à l'aide d'un système intégré de traduction acellulaire reconstituée et de flexizymes. Nous résumons ici la base historique de cette technologie et ses applications actuelles et futures pour la synthèse de peptides non standards et pour la découverte de médicaments.

Mots-clés : peptide non standard, mis-aminoacylation, peptide thérapeutique, traduction.

[Traduit par la Rédaction]

Introduction: standard peptides versus nonstandard peptides

The recent successful development of therapeutic proteins has made a significant impact on the pharmaceutical industry (Leader et al. 2008), yet most drug-development initiatives focus on small organic molecules, because of their target flexibility (intra- and extra-cellular target proteins), stability, and oral bioavailability. Despite such advantages, the binding of small organic molecules to proteins mainly relies on the existence of the narrow and deep binding pockets of their target proteins, where they can fit in. Unfortunately, not all therapeutically relevant target proteins have such a characteristic site. Instead, it is known that protein–protein interactions often occur on a wide and shallow protein interface, of which the standard size spans about 1600 Å^2 (Lo Conte et al. 1999). Additionally, small, deep cavities that can serve as binding sites for small organic molecules are rarely found at the interface of protein–protein interaction pairs (Arkin and Wells 2004). Therefore, the development of small organic molecules capable of disrupting protein–protein interactions may be intrinsically difficult. In this sense, therapeutic antibodies and antibody-like proteins are ideal molecules to disrupt these interactions (Leader et al. 2008); however, they are costly, and their poor membrane permeability restricts their use to extracellular targets (Arkin and Wells 2004). Thus, it is desirable to develop therapeutic agents that are big enough to cover the interface of protein targets but still able to penetrate into cells so that the agents can function against both extracellular and intracellular targets.

Received 22 October 2007. Revision received 25 January 2008. Accepted 28 January 2008. Published on the NRC Research Press Web site at bcb.nrc.ca on 20 March 2008.

T.J. Kang and H. Suga.[2] Research Center for Advanced Science and Technology, University of Tokyo, 153-8904 Tokyo; and Department of Chemistry and Biotechnology, Graduate School of Engineering, University of Tokyo, 113-8656 Tokyo, Japan.

[1]This paper is one of a selection of papers published in this Special Issue, entitled CSBMCB — Systems and Chemical Biology, and has undergone the Journal's usual peer review process.

[2]Corresponding author (e-mail: hsuga@rcast.u-tokyo.ac.jp).

 doi:10.1139/O08-009

In the last few decades, a number of short, naturally occurring peptides possessing a variety of biological activities have been discovered. Potencies of these peptides are remarkably high, often exhibiting activities at low concentrations. For instance, human urotensin-II (hU-II), consisting of 11 amino acids, is a potent vasoconstrictor that strongly binds to one of the G-protein-coupled receptors and induces calcium mobilization at subnanomolar concentrations (Ames et al. 1999; Coulouarn et al. 1998). Moreover, peptide fragments artificially designed from substrates of target proteins or their receptors have been shown to retain their biological activities. For instance, a short peptide derived from the erythropoietin receptor (EPO-R) was able to bind to EPO-R and activate the signaling pathway (Naranda et al. 1999). Furthermore, phage display technology has enabled us to screen artificial peptides from random peptide libraries (Parmley and Smith 1988). Along the same lines, phage display selection against EPO-R has given rise to a peptide sequence that has no sequence homology to EPO yet exhibits activity both in vitro and in vivo (Johnson and Jolliffe 2000; Wrighton et al. 1996). Even though the phage display method has been successfully used to generate a variety of peptide sequences that bind to target proteins, such peptides rarely exhibit high therapeutic potencies in vivo. This is because peptides consisting of the 20 proteinogenic (standard) amino acids are generally susceptible to proteases and are often digested before exhibiting their expected biological activities. Thus, to devise protease-resistant peptides based on the phage-selected peptides, each of the analogs must be chemically synthesized and nonproteinogenic amino acids incorporated, follwed by tedious rescreening against the target to optimize such sequences.

Nature overcomes this limitation by synthesizing nonstandard peptides containing unusual monomers, e.g., N^{α}-methylated amino acids, hydroxy acids, and amino acids with D-configuration or non-proteinogenic side-chains. Whereas standard peptides are synthesized by the mRNA-directed polymerization of amino acids by the translation machinery, nonstandard peptides are generally synthesized by the template-independent synthesis machinery consisting of clusters of modular protein enzymes, called nonribosomal peptide sysnthetases (NRPSs) (Fischbach and Walsh 2006). Remarkably, these nonstandard peptides, which had been isolated as natural products, exhibit a wide range of biological activities not only against microorganisms but also in human cells (Schwarzer et al. 2003). Engineering appropriate modules in NRPSs is expected to generate new machineries capable of synthesizing novel kinds of nonstandard peptides. In a preliminary study, a small library of nonstandard peptides was prepared by NRPSs containing engineered donor and acceptor communication domains (Hahn and Stachelhaus 2006). However, current knowledge about the generality and portability of communication domains is yet insufficient to generate randomly shuffled enzyme domains. Thus, it is still a demanding task to further engineer the clusters that produce a variety of nonstandard peptide libraries for the discovery of novel therapeutic molecules.

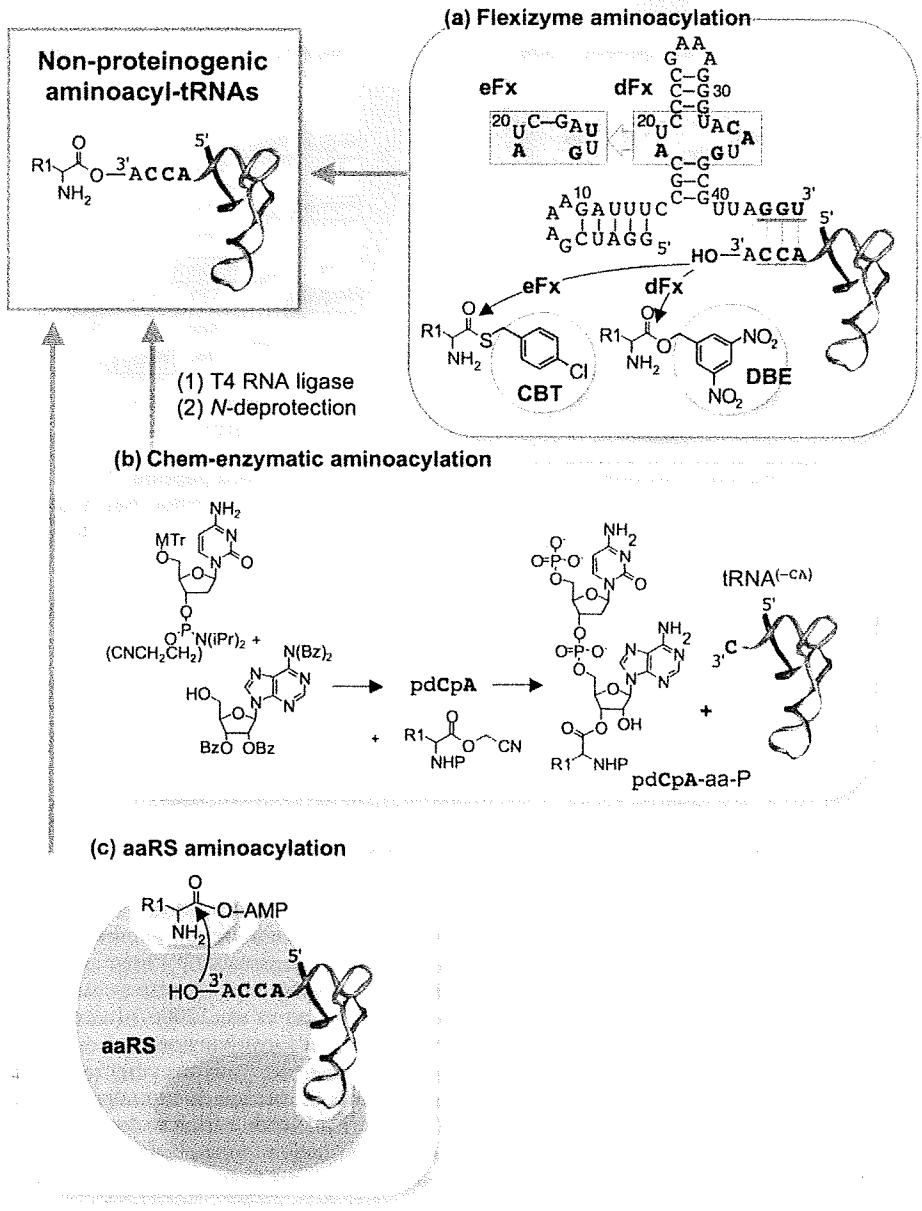
Chemical synthesis has been the only alternative method to generate nonstandard peptide libraries, but it is not necessarily ideal for handling diverse libraries against various protein targets. When the library size reaches one million, it is impractical to test all the individual peptides for their binding properties (Weiss and Chamberlin 2003); thus, diverse libraries are usually subject to various selection strategies to reduce the number of selectants (Turk and Cantley 2003). Additionally, selectants are not usually amplifiable in the chemically synthesized libraries; thus, the absolute amount necessary to identify positive hits restricts practical diversity. Instead, a new technology, termed genetic code reprogramming, has recently been devised for the synthesis of nonstandard peptides in a template-directed format using a reconstituted cell-free translation system. Here, we discuss the development of this emerging technology, as well as its strategies for the synthesis of nonstandard peptides and their libraries and future applications.

Nonsense and 4-base codon suppressions versus genetic code reprogramming

Although the ordinary translation system strictly incorporates 20 proteinogenic amino acids into the nascent peptide chain, some organisms use the same codon for different purposes. For example, UGA and UAG codons occasionally encode selelnocysteine and pyrrolysine, respectively, in a variety of organisms, whereas these codons denote the translation stop signal in general (Zhang and Gladyshev 2007). It is also known that an appropriate manipulation of the translation system enables us to incorporate nonproteinogenic amino acids into peptides. A classical example is that chemically generated misaminoacylated tRNA (alanyl-$tRNA^{Cys}$) supported translation similar to normal enzymatically prepared aminoacyl-tRNAs in vitro (Chapeville et al. 1962; Ehrenstein et al. 1963), indicating that the tRNA aminoacylation event by aminoacyl-tRNA synthetases (aaRSs), and not the ribosome, is the major player controlling translation fidelity. This created the possibility of using the ribosome for making peptides containing nonproteinogenic amino acid(s), and later efforts were devoted to devising methods for assigning nonproteinogenic amino acids to a single codon by using appropriate misacylation technologies.

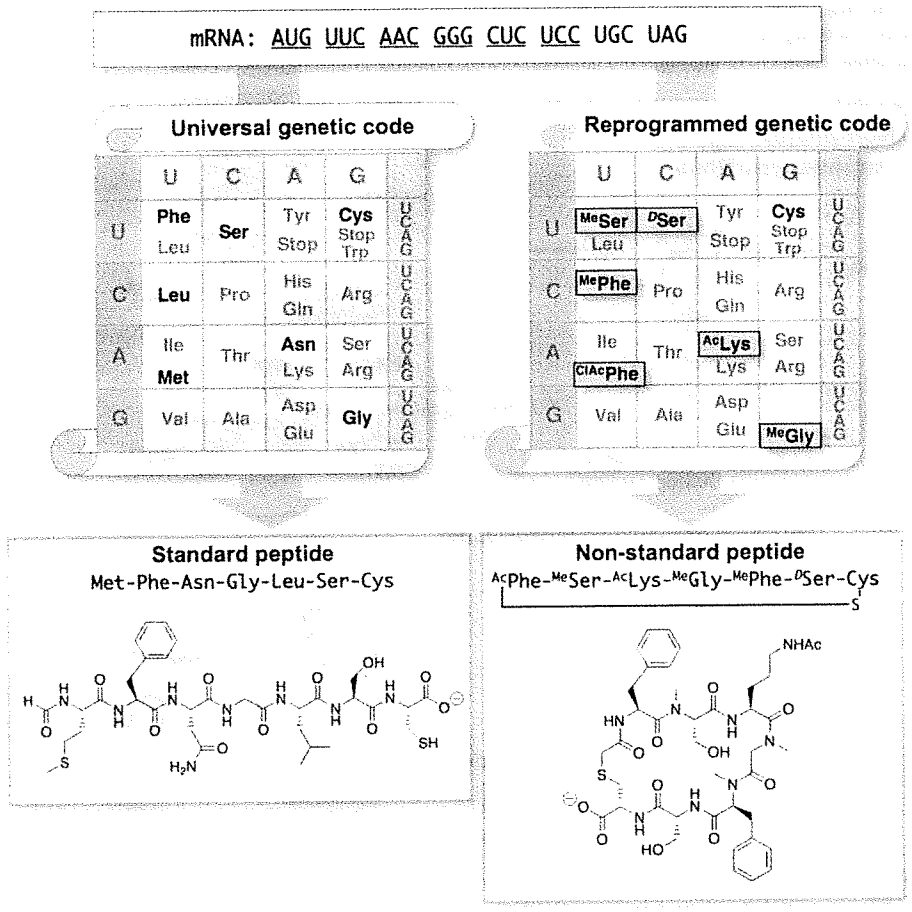
There are 64 possible combinations of triplet nucleotides (codons), all of which are used to code for the 20 amino acids, with the exception of 3 codons, which denote the translation stop signal. These stop codons (or nonsense codons) can be re-assigned to nonproteinogenic amino acid(s) (Noren et al. 1989). However, this method, generally called nonsense suppression, suffers from inherent competition with release factor(s), occasionally yielding low incorporation efficiency, depending upon downstream or upstream codon sets. In addition, this method is suitable only for the incorporation of a single type of nonproteinogenic amino acid into a peptide chain at the specific site assigned by one of the stop codons (usually a UAG stop codon), which is not diverse enough for the construction of a nonstandard peptide library. Complementary to this nonsense suppression, nucleotide quadruplet codons (4-base codons) have been used to assign non-proteinogenic amino acids (Hohsaka et al. 1996), where the programmed frameshift occurs at the 4-base codon as a correct reading frame. This method enabled the incorporation of two, and occasionally three, nonproteinogenic amino acids charged onto

Fig. 1. Preparation of misaminoacyl-tRNAs. (*a*) Flexizyme method (left). Two ribozymes, dFx and eFx, show different substrate specificities, as shown. Flexizyme recognizes only the 3 nucleotides at the 3′-end of tRNA (N_{73}-C_{75}) and the active ester of amino acid substrates, thus this method is compatible with a wide range of acids and tRNAs. (*b*) Chem-enzymatic method (top). The terminal dinucleotide (pdCpA) is chemically synthesized using the standard phosphoramidite method (1); the substrate amino acid (aa) protected by an appropriate protective group is chemically linked to the 3′ end of pdCpA (2); the resulting pdCpA-aa is subsequently ligated to the truncated tRNA (tRNA–3′CA) lacking the 3′-terminal CA by the action of T4 RNA ligase (3); deprotection of the protective group to generate aa-tRNA (4). MTr, methoxytrityl; Bz, benzoyl. (*c*) Enzymatic method using engineered aminoacyl-tRNA synthetase (aaRS, bottom). Both the rational design and the molecular evolution were reported to render the extremely high specificity of some aaRSs. Unlike the previous two methods, this approach enables the large scale syntheses of nonstandard amino acids by expressing engineered aaRS in the host cell. However, because of high specificity, each engineered aaRS must be prepared for each nonstandard amino acid and for each anticodon. So far, the choice of nonstandard amino acids is limited and the anticodon is also limited to the anti-UAG codon.



tRNAs bearing 4-base anticodons into the peptide chain (Ohtsuki et al. 2005). However, because 4-base codons must be designed based on rarely used codons in *Escherichia coli*, the number of usable codons is still restricted, particularly upon using exotic nonproteinogenic acids that are often difficult to incorporate into a peptide chain, and this method suffers from incomplete synthesis of the peptide as a result of undesired reading of the 4-base codon(s) by com-

Fig. 2. Genetic reprogramming approach for the preparation of nonstandard peptides. When the mRNA sequence shown is translated to a peptide, the translation system faithfully follows the universal code book (shown on the left) in the usual in vitro translation system (PURE system). The resulting heptapeptide (fMet-Phe-Asn-Gly-Leu-Ser-Cys) is shown below. In this example, by removing Met, Phe, Asn, Gly, Leu, and Ser (and (or) corresponding aaRSs) from the translation system (wPURE system), codons for those amino acids can be re-assigned to nonstandard amino acids. For this genetic reprogramming, the key component is misaminoacyl-tRNA (^{ClAc}Phe-tRNA$_{ini}$, ^{Me}Ser-tRNA$_{GAA}$, ^{Ac}Lys-tRNA$_{GUU}$, ^{Me}Gly-tRNA$_{CCC}$, ^{Me}Phe-tRNA$_{GAG}$, and ^{D}Ser-tRNA$_{GGA}$, in this example; ^{ClAc}Phe, N^{α}-(2-chloroacetyl)phenylalanine; ^{Ac}Lys, N^{ε}-acetyl-lysine; ^{Me}Gly and ^{Me}Phe, N^{α}-methyl-glycine and N^{α}-methyl-phenylalanine, repectively). The resulting nonstandard peptide is shown below.

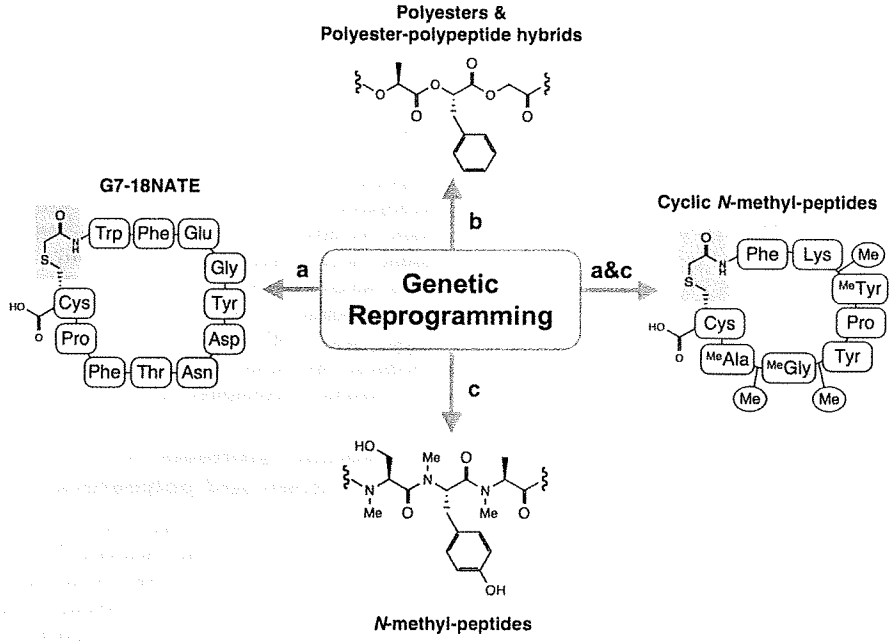


peting aminoacyl-tRNA(s) bearing the normal 3-base anticodon(s).

In addition to the above limitations, the step of tRNA aminoacylation with nonproteinogenic amino acids relies on laborious and technically demanding chem-enzymatic processes (Hecht et al. 1978; Heckler et al. 1984; Noren et al. 1989; Robertson et al. 1989). This technical barrier made this method so specialized that only a subset of researchers was able to use it (Fig. 1). More recently, Schultz and other groups have succeeded in developing mutant aaRSs that charge a certain family of nonproteinogenic amino acids (mostly aromatic amino acid analogs) onto orthogonal tRNA (Kiga et al. 2002; Wang and Schultz 2005; Yoo and Tirrell 2007; Zhang et al. 2004). However, this method is limited to the incorporation of a single nonproteinogenic amino acid, and also the choice of the amino acids remains limited (Fig. 1).

More recently, independent efforts were made to overcome such limitations in conventional nonproteinogenic amino acid mutagenesis by Foster et al., followed by those of several other groups (Forster et al. 2003; Josephson et al. 2005; Murakami et al. 2006). In contrast to the aforementioned methods involving nonsense or 4-base codon suppression, this method involves the reassignment of arbitral codons from proteinogenic to nonproteinogenic amino acids. However, nucleotide triplet codons in the mRNA pass genetic information to the peptide through the tightly regulated aminoacylation of cognate tRNAs, making it very difficult to reassign the genetic code in the usual translation system. A key technology that makes it possible to break the tight regulation of the correspondence is a reconstituted *E. coli* cell-free translation system, often referred to as PURE (protein synthesis using recombinant elements) system (Shimizu et al. 2001). The most important feature of this translation

Fig. 3. Genetic code reprogramming and its applications. Genetic code reprogramming made it possible to incorporate the key amino acids into a peptide for various purposes. (*a*) N^{α}-(2-chloroacetyl)tyrosine was incorporated into the sequence of G7-18NATE via initiation codon reprogramming. Subsequent cyclization through thioether yielded a cyclized G7-18NATE analog. (*b*) α-Hydroxy acids, lactic acid, phenyllactic acid, and glycolic acid, were incorporated into a peptide structure consecutively, yielding an ester-backbone-containing peptide. (*c*) N^{α}-Methyl-serine, N^{α}-methyl-phenylalanine, and N^{α}-methyl-alanine are incorporated into a peptide. The resulting peptide has a nonstandard backbone that can be found in many therapeutically important natural products. This approach can be combined with the initiation codon reprogramming to ribosomally synthesize cyclized N^{α}-methylated peptides, yielding a structure that has great potential in peptide therapeutics.



system is that certain amino acids and aaRSs can be withdrawn from the translation elements. For example, when the translation system is deprived of leucine and its corresponding aaRS, leucinyl-tRNA cannot be synthesized in the system, and thus is practically withdrawn, even though corresponding tRNA is present in the system. Using such a withdrawn PURE system, termed wPURE, vacant codons can be created at a researcher's will (Fig. 2). The significant differences between this method and the nonsense or 4-base codon suppression are (*i*) because natural aminoacyl-tRNAs are removed from the translation system, competition between natural and nonstandard aminoacyl-tRNAs would not occur, and (*ii*) through appropriate selection of codons for the nonstandard amino acid, the maximal sense suppression efficiency can be obtained. A major barrier that remains for this method is how nonproteinogenic amino acids can be assigned to these vacant codons, i.e., how can they be attached onto the desired tRNAs capable of reading the vacant codons. Clearly, a nonlaborious and reliable method is necessary to perform this process.

Methods for tRNA aminoacylation, the key technology

The Szostak group showed a pioneering example of codon table redesigning (Josephson et al. 2005). Using nonproteinogenic amino acids compatible with aaRSs, they reassigned the universal genetic code to one that contains 12 nonproteinogenic amino acids, and also showed that this translation system was compatible with the newly assigned genetic code. The translation system resulted in peptides containing nonproteinogenic amino acids, as designated by the mRNA sequence. This technology, together with their proprietary technology, mRNA display, had opened a possibility for in vitro screening of nonstandard peptide aptamers (peptides capable of binding to targets). However, since this system utilizes aaRSs for mischarging tRNAs with nonproteinogenic amino acids, only those that structurally resemble proteinogenic amino acids (and thus, are compatible with aaRSs) can be used, thereby limiting the variety of peptides that can be used. A more serious problem of this approach is contamination of proteinogenic amino acids carried over into the wPURE system with purified ribosome or recombinant factors. When exotic nonproteinogenic amino acids, which are usually inefficiently incorporated into the nascent peptide chain are used for the suppressions, a small amount of such unavoidable proteinogenic contaminants readily outcompete the nonproteinogenic amino acids, leading to peptides composed primarily of these standard amino acids.

Obviously, a general method for the preparation of nonstandard aminoacyl-tRNAs is a prerequisite for genetic reprogramming. Although the chem-enzymatic method is suitable to this application, and indeed, was used in the earlier work by Foster et al. (Forster et al. 2003), its technical difficulties and laboriousness hinder its widespread use in creating peptide libraries with diverse kinds of non-proteino-

Table 1. Yields of acyl-tRNAs.

Acid substrate	Flexizyme	Yield (%)
Standard amino acids		
Ala-DBE	dFx	36[a]
Asn-DBE	dFx	22[b]
Asp-DBE	dFx	52[b]
Cys-DBE	dFx	46[b]
Gln-DBE	dFx	46[b]
Glu-DBE	dFx	17[a]
Gly-DBE	dFx	39[a]
His-DBE	dFx	29[a]
Leu-DBE	dFx	37[a]
Lys-DBE	dFx	36[a]
Met-DBE	dFx	35[a]
Phe-CME	eFx	47[a]
Pro-DBE	dFx	37[b]
Ser-DBE	dFx	38[a]
Trp-CME	eFx	36[a]
Tyr-CME	eFx	34[a]
Val-DBE	dFx	13[a]
Amino acids with nonstandard side-chains		
Aly-DBE	dFx	33[a]
Bly-DBE	dFx	30[a]
Cit-DBE	dFx	35[a]
α-Hydroxy acids		
Hbi-DBE	dFx	25[a]
Hle-DBE	dFx	51[b]
Hph-CME	eFx	82[b]
β-Amino acid and N^{α}-methyl amino acid		
Bal-DBE	dFx	17[a]
Mle-DBE	dFx	55[b]

Note: Yields were calculated using either a streptavidin-dependent gel-shift assay or acid PAGE, depending on the substrate (Murakami et al. 2006). Amino acids are represented in standard three-letter abbreviations except as follows: Aly, ε-*N*-acetyl-lysine; Bly, ε-*N*-biotinyl-lysine; Cit, L-citrulline; Hbi, δ-*N*-biotinyl-(*S*)-hydroxybutanoic acid; Hle, (*S*)-3-isoproyllactic acid; Hph, (*S*)-3-phenyllactic acid; Bal, β-alanine; Mle, α-*N*-methyl-leucine. Substrates were activated by either 3,5-dinitrobenzyl (DBE) or cyanomethyl ester (CME).

[a]Yield calculated using a streptavidin-dependent gel-shift assay.

[b]Yield calculated using acid PAGE.

genic amino acids. Instead, an artificial RNA enzyme (ribozyme) capable of catalyzing aminoacylation of tRNAs was generated from a random sequence pool of RNA by our group (Saito et al. 2001), and later, this ribozyme turned into a highly flexible tool for tRNA aminoacylation and was named the flexizyme system (Murakami et al. 2006). The system consists of two kinds of flexizymes, called dFx and eFx, which can be used depending upon the choice of a leaving group on the substrates; dFx for 3,5-dinitobenzyl ester (DBE), and eFx for 4-chlorobenzyl thioester (CBT) or cyanomethyl ester (CME) (Fig. 1). Because the flexizymes recognize their cognate leaving group, and not the side-chain or the free amino group of a given substrate, they are able to charge a wide variety of α-amino acids with nonproteinogenic side-chains. However, the most remarkable feature is that the combination of these two flexizymes gives virtually no restriction of substrates such as N^{α}-methylated amino acids, D-α-amino acids, β-amino acids, and even α-hydroxy acids (Table 1). Moreover, flexizymes recognize N_{73}–C_{75} of tRNA by base pairings with G43–U45, where C_{74} and C_{75} are common to all tRNAs and N_{73} can be any of A, G, and U (even C can be accepted if the incubation time is prolonged); thus virtually any tRNA can be a substrate for flexizymes (Fig. 1).

Thus, we combined the flexizyme system with a wPURE system to demonstrate the genetic code reprogramming for the synthesis of nonstandard peptides. In the first demonstration of the combined systems, three codons (AGU, AAC, and CAG for serine, asparagines, and glutamine, respectively) were reassigned to three nonstandard amino acids (ε-*N*-acetyl-lysine, citrulline, and *p*-iodo-phenylalnine, respectively) so that a 17-mer peptide possessing 6 nonproteinogenic amino acids could be synthesized by the ribosome as efficiently as the original sequence with standard amino acids (Murakami et al. 2006). Typical yields range from several to 20 μmol/L in the translation mixture. With this foundation, our efforts have been made to prepare a wide variety of nonstandard peptides with highly altered chemical structures. We will describe some of our most recent results to exemplify the versatility of this new method.

Ribosomal synthesis of nonstandard polypeptides and polyesters

The SH2 domain of Grb7 plays a role in signaling by binding to an intracellular phosphorylated tyrosine of several receptor tyrosine kinases. Pero et al. performed a selection of peptides against the SH2 domain using phage display and isolated novel high affinity peptides (Pero et al. 2002). G7-18, one of the most potent peptides selected, has the characteristics of a nonphosphorylated and cyclic form, closed by a disulfide bond between two internal cysteine residues. Unfortunately, this disulfide bond in G7-18 was reducible so that its linear form not only failed to exhibit activity but also was susceptible to proteases in vivo. The same team ingeniously substituted the disulfide bond with a thioether bond by chemical synthesis; the N terminus of the cysteine residue of G7-18 was substituted with a 2-chloroacetyl group (ClAc), resulting in the intramolecular attack of the C-terminal cysteine side-chain to the 2-position of the acetyl group. Despite the fact that this new peptide, called G7-18NATE, lost approximately 10-fold of its affinity to Grb7 compared with that of G7-18, it still exhibited antitumor activity in an animal study. We recently succeeded in the ribosomal synthesis of G7-18NATE using initiation codon reprogramming, in which ^{ClAc}Trp was used to prime the translation in a methionine-depleted wPURE system (Goto et al. 2008). Initiator tRNA was charged with ^{ClAc}Trp by the action of flexizyme, and the resulting aminoacyl-tRNA was used for the translation initiation instead of initiator tRNA charged with α-*N*-formyl-methione. Remarkably, the peptide was spontaneously and only intramolecularly cyclized through the thioether bond upon the completion of translation, and therefore no extra treatment was necessary. We envision that our method will allow us to re-investigate the sequence of G7-18NATE by coupling initiation codon reprogramming with an appropriate in vitro display system.

Since this genetic code reprogramming system lacks competitors for the incorporations of nonproteinogenic amino acids, it is possible to perform multiple incorporations or the polymerization of monomers that would alter the backbone from the ordinary peptide bond to other types of bonds. For instance, we have recently performed mRNA-directed polyester synthesis using seven varieties of α-hydroxy acids that were each individually assigned to one of seven codons (Ohta et al. 2007). This work represents the first demonstration of up to 12 consecutive additions of α-hydroxy acids with various different compositions designated by codons on mRNA. We also recently performed the mRNA-directed synthesis of linear and cyclic *N*-methyl-peptides (Kawakami et al. 2008). As a proof-of-concept experiment, up to 10 successive incorporations of N^{α}-methyated amino acids with six different kinds of side-chains were performed. In this work, we also combined this method with the aforementioned thioether cyclization method using the ClAc group, showing that cyclic *N*-methyl-peptides with three or four *N*-methylated backbone residues could be synthesized. This work has opened the possibility of the mRNA-encoded synthesis of *N*-methyl-peptide libraries for the screening of biologically active molecules.

Conclusion and perspectives

Since the advent of the combinatorial synthesis of peptides, peptides have drawn significant attention with the anticipation of finding novel drugs. However, only a few therapeutic peptides have reached the market because of problems associated with poor stability and cell-membrane permeability. A lesson from peptide-like natural products has given a possible direction; the peptide should have non-standard structures with not only side-chains but a cyclic and N^{α}-methylated backbone, as well. We now have a new tool for synthesizing such nonstandard peptides in a cell-free translation system so that their sequences can be encoded by oligonucleotides (mRNA or cDNA) that are readily amplifiable and sequence readable by conventional molecular biological techniques.

A controllable translation system and a flexible tool for the preparation of a wide variety of nonstandard aminoacyl-tRNAs are essential to genetic code reprogramming. As described here, the combined use of wPURE system and the flexizyme technology is a reliable methodology for the ribosomal syntheses of nonstandard peptides containing the desired modifications (Fig. 3). By means of genetic code reprogramming, we can expect a whole new era of peptide chemistry that is compatible not only with cell-based high-throughput screening, but with in vitro selection methods, such as mRNA or ribosome display against isolated targets, as well. We are on the verge of witnessing the discovery of active nonstandard peptides against various therapeutic targets.

Acknowledgements

This work was supported by grants from the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (S) (16101007) and the US National Institutes of Health (GM59159).

References

Ames, R.S., Sarau, H.M., Chambers, J.K., Willette, R.N., Alyar, N.V., Romanic, A.M., et al. 1999. Human urotensin-II is a potent vasoconstrictor and agonist for the orphan receptor GPR14. Nature, **401**: 282–286. doi:10.1038/45809. PMID:10499587.

Arkin, M.R., and Wells, J.A. 2004. Small-molecule inhibitors of protein–protein interactions: Progressing towards the dream. Nat. Rev. Drug Discov. **3**: 301–317. doi:10.1038/nrd1343. PMID:15060526.

Chapeville, F., Ehrenstein, G.V., Benzer, S., Weisblum, B., Ray, W.J., and Lipmann, F. 1962. On role of soluble ribonucleic acid in coding for amino acids. Proc. Natl. Acad. Sci. U.S.A. **48**: 1086–1092. doi:10.1073/pnas.48.6.1086. PMID:13878159.

Coulouarn, Y., Lihrmann, I., Jegou, S., Anouar, Y., Tostivint, H., Beauvillain, J.C., et al. 1998. Cloning of the cDNA encoding the urotensin II precursor in frog and human reveals intense expression of the urotensin II gene in motoneurons of the spinal cord. Proc. Natl. Acad. Sci. U.S.A. **95**: 15803–15808. doi:10.1073/pnas.95.26.15803. PMID:9861051.

Ehrenstein, G., Weisblum, B., and Benzen, S. 1963. Function of sRNA as amino acid adaptor in synthesis of hemoglobin. Proc. Natl. Acad. Sci. U.S.A. **49**: 669–675. doi:10.1073/pnas.49.5.669. PMID:16591086.

Fischbach, M.A., and Walsh, C.T. 2006. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. Chem. Rev. **106**: 3468–3496. doi:10.1021/cr0503097. PMID:16895337.

Forster, A.C., Tan, Z.P., Nalam, M.N.L., Lin, H.N., Qu, H., Cornish, V.W., and Blacklow, S.C. 2003. Programming peptidomimetic syntheses by translating genetic codes designed de novo. Proc. Natl. Acad. Sci. U.S.A. **100**: 6353–6357. doi:10.1073/pnas.1132122100. PMID:12754376.

Goto, Y., Ohta, A., Sako, Y., Yamagishi, Y., Murakami, H., and Suga, H. 2008. Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. ACS Chem. Biol. **3**: 120–129. doi:10.1021/cb700233t. PMID:18215017.

Hahn, M., and Stachelhaus, T. 2006. Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. Proc. Natl. Acad. Sci. U.S.A. **103**: 275–280. doi:10.1073/pnas.0508409103. PMID:16407157.

Hecht, S.M., Alford, B.L., Kuroda, Y., and Kitano, S. 1978. "Chemical aminoacylation" of tRNA's. J. Biol. Chem. **253**: 4517–4520. PMID:248056.

Heckler, T.G., Chang, L.H., Zama, Y., Naka, T., Chorghade, M.S., and Hecht, S.M. 1984. T4 RNA ligase mediated preparation of novel "chemically misacylated" tRNAPhes. Biochemistry, **23**: 1468–1473. doi:10.1021/bi00302a020. PMID:6372858.

Hohsaka, T., Ashizuka, Y., Murakami, H., and Sisido, M. 1996. Incorporation of nonnatural amino acids into streptavidin through in vitro frame-shift suppression. J. Am. Chem. Soc. **118**: 9778–9779. doi:10.1021/ja9614225.

Johnson, D.L., and Jolliffe, L.K. 2000. Erythropoietin mimetic peptides and the future. Nephrol. Dial. Transplant. **15**: 1274–1277. doi:10.1093/ndt/15.9.1274. PMID:10978375.

Josephson, K., Hartman, M.C.T., and Szostak, J.W. 2005. Ribosomal synthesis of unnatural peptides. J. Am. Chem. Soc. **127**: 11727–11735. doi:10.1021/ja0515809. PMID:16104750.

Kawakami, T., Murakami, H., and Suga, H. 2008. Messenger RNA-programmed incorporation of multiple *N*-methyl-amino acids into linear and cyclic peptides. Chem. Biol. **15**: 32–42. doi:10.1016/j.chembiol.2007.12.008. PMID:18215771.

Kiga, D., Sakamoto, K., Kodama, K., Kigawa, T., Matsuda, T., Yabuki, T., et al. 2002. An engineered *Escherichia coli* tyrosyl-

tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. Proc. Natl. Acad. Sci. U.S.A. **99**: 9715–9720. doi:10.1073/pnas.142220099. PMID:12097643.

Leader, B., Baca, Q.J., and Golan, D.E. 2008. Protein therapeutics: a summary and pharmacological classification. Nat. Rev. Drug Discov. **7**: 21–39. doi:10.1038/nrd2399. PMID:18097458.

Lo Conte, L., Chothia, C., and Janin, J. 1999. The atomic structure of protein–protein recognition sites. J. Mol. Biol. **285**: 2177–2198. doi:10.1006/jmbi.1998.2439. PMID:9925793.

Murakami, H., Ohta, A., Ashigai, H., and Suga, H. 2006. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. Nat. Methods, **3**: 357–359. doi:10.1038/nmeth877. PMID:16628205.

Naranda, T., Wong, K., Kaufman, R.I., Goldstein, A., and Olsson, L. 1999. Activation of erythropoietin receptor in the absence of hormone by a peptide that binds to a domain different from the hormone binding site. Proc. Natl. Acad. Sci. U.S.A. **96**: 7569–7574. doi:10.1073/pnas.96.13.7569. PMID:10377456.

Noren, C.J., Anthonycahill, S.J., Griffith, M.C., and Schultz, P.G. 1989. A general-method for site-specific incorporation of unnatural amino-acids into proteins. Science, **244**: 182–188. doi:10.1126/science.2649980. PMID:2649980.

Ohta, A., Murakami, H., Higashimura, E., and Suga, H. 2007. Synthesis of polyester by means of genetic code reprogramming. Chem. Biol. **14**: 1315–1322. PMID:18096500.

Ohtsuki, T., Manabe, T., and Sisido, M. 2005. Multiple incorporation of non-natural amino acids into a single protein using tRNAs with non-standard structures. FEBS Lett. **579**: 6769–6774. doi:10.1016/j.febslet.2005.11.010. PMID:16310775.

Parmley, S.F., and Smith, G.P. 1988. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. Gene, **73**: 305–318. doi:10.1016/0378-1119(88)90495-7. PMID:3149606.

Pero, S.C., Oligino, L., Daly, R.J., Soden, A.L., Liu, C., Roller, P.P., et al. 2002. Identification of novel non-phosphorylated ligands, which bind selectively to the SH2 domain of Grb7. J. Biol. Chem. **277**: 11918–11926. doi:10.1074/jbc.M111816200. PMID:11809769.

Robertson, S.A., Noren, C.J., Anthonycahill, S.J., Griffith, M.C., and Schultz, P.G. 1989. The use of 5′-phospho-2 deoxyribocytidylylriboadenosine as a facile route to chemical aminoacylation of transfer-RNA. Nucleic Acids Res. **17**: 9649–9660. doi:10.1093/nar/17.23.9649. PMID:2602139.

Saito, H., Kourouklis, D., and Suga, H. 2001. An in vitro evolved precursor tRNA with aminoacylation activity. EMBO J. **20**: 1797–1806. doi:10.1093/emboj/20.7.1797. PMID:11285242.

Schwarzer, D., Finking, R., and Marahiel, M.A. 2003. Nonribosomal peptides: from genes to products. Nat. Prod. Res. **20**: 275–287. doi:10.1039/b111145k.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. 2001. Cell-free translation reconstituted with purified components. Nat. Biotechnol. **19**: 751–755. doi:10.1038/90802. PMID:11479568.

Turk, B.E., and Cantley, L.C. 2003. Peptide libraries: at the crossroads of proteomics and bioinformatics. Curr. Opin. Chem. Biol. **7**: 84–90. doi:10.1016/S1367-5931(02)00004-2. PMID:12547431.

Wang, L., and Schultz, P.G. 2005. Expanding the genetic code. Angew. Chem. Int. Ed. **44**: 34–66. doi:10.1002/anie.200460627.

Weiss, G.A., and Chamberlin, R. 2003. Bridging the synthetic and biopolymer worlds with peptide-drug conjugates. Chem. Biol. **10**: 201–202. doi:10.1016/S1074-5521(03)00056-5. PMID:12670531.

Wrighton, N.C., Farrell, F.X., Chang, R., Kashyap, A.K., Barbone, F.P., Mulcahy, L.S., et al. 1996. Small peptides as potent mimetics of the protein hormone erythropoietin. Science, **273**: 458–463. doi:10.1126/science.273.5274.458. PMID:8662529.

Yoo, T.H., and Tirrell, D.A. 2007. High-throughput screening for methionyl-tRNA synthetases that enable residue-specific incorporation of noncanonical amino acids into recombinant proteins in bacterial cells. Angew. Chem. Int. Ed. **46**: 5340–5343. doi:10.1002/anie.200700779.

Zhang, Y., and Gladyshev, V.N. 2007. High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm Olavius algarvensis. Nucleic Acids Res. **35**: 4952–4963. doi:10.1093/nar/gkm514. PMID:17626042.

Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D.S., and Schultz, P.G. 2004. Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **101**: 8882–8887. doi:10.1073/pnas.0307029101. PMID:15187228.

J|A|C|S
COMMUNICATIONS

Published on Web 05/14/2008

Ribosomal Synthesis of Bicyclic Peptides via Two Orthogonal Inter-Side-Chain Reactions

Yusuke Sako,[†,‡] Jumpei Morimoto,[†,‡] Hiroshi Murakami,[†] and Hiroaki Suga*[,†,‡,§]

Research Center for Advanced Science and Technology, The University of Tokyo, Komaba, Meguro, Tokyo 153-8904, Japan, Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo, Tokyo 113-8656, Japan, and Department of Advanced Interdisciplinary Studies, Graduate School of Engineering, The University of Tokyo, 4-6-1, Komaba, Meguro, Tokyo 153-8904, Japan

Received February 6, 2008; E-mail: hsuga@rcast.u-tokyo.ac.jp

Many naturally occurring peptides possess two pairs of disulfide bonds, representing a bicyclic structural feature.[1] Such a structure rigidifies the peptide conformation into a specific active form and consequently plays a crucial role in the biological function of the peptide. To form the two disulfide bonds, two particular pairs of cysteines (Cys) out of four Cys residues must selectively react with each other, avoiding two other possible combinations[2] (Figure 1A, a−d = Cys). Specific formation of such disulfide bonds in naturally occurring peptides is generally dictated by the proximity of the Cys residues in the tertiary structure[3] or in cells it is occasionally driven by a disulfide isomerase that rearranges kinetically trapped incorrect disulfide pairs to the correct pairs.[4] Although it is of great interest to generate libraries of such bicyclic peptides using the translation apparatus and subsequently screen their biological functions, upon randomization of the sequence, the disulfide bond formation would very likely be scrambled (Figure 1A); thus the 4-Cys-containing peptide scaffold cannot be utilized as a reliable bicyclic platform. We here report a new peptide scaffold containing a single Cys and three nonproteinogenic amino acids (Figure 1B), selectively forming the desirable cross-links. Significantly, this peptide scaffold can be expressed by using a reconstituted *Escherichia coli* cell-free translation system (called PURE[5]) under the reprogrammed genetic code. Therefore, it is readily applicable to the preparation of libraries of bicyclic peptides with a uniform skeleton.

We designed and synthesized two pairs of amino acids containing orthogonal functional groups (Figure 1B). 4-(2-Chloroacetyl)aminobutyric acid (Cab) bears the thiol-reactive functional group in its side chain. We have previously reported that, upon the site-specific incorporation of Cab paired with a Cys into the peptide chain, these two side chains spontaneously react with each other to give a thioether linkage.[6] Because this chemistry requires no additional reagent, the Cab−Cys pair is a convenient and reliable amino acid pair to generate the desired monocyclic peptides closed by a nonreducible bond. The functional groups of the other pair, azidohomoalanine (Aha) and propargylglycine (Pgl), are known to react with each other via Cu(I)-catalyzing azide−alkyne cycloaddition, and this type of chemistry has been extensively used in bioconjugate applications.[7] Despite the fact that the individual molecules were shown to be compatible with translation,[8] their site-specific double incorporations have not yet been demonstrated. Therefore, we first attempted to establish the assignment of Aha and Pgl to certain codons by genetic code reprogramming and demonstrate their double incorporations into the nascent chain of a model peptide, and then perform the intramolecular cyclization.

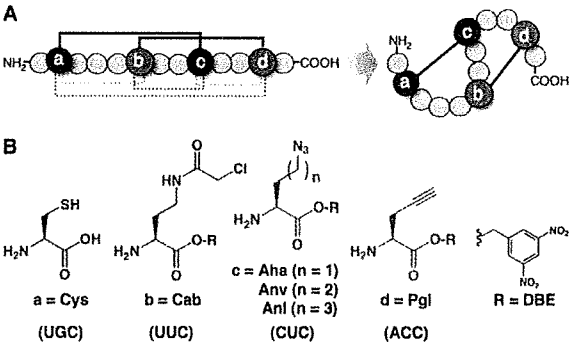


Figure 1. Ribosomal synthesis of bicyclic peptides. (A) Schematic representation of a bicyclic peptide closed by two orthogonal inter-side-chain reactions. (B) Structure and assignment of amino acids to the reprogrammed genetic code. Cab, 4-(2-Chloroacetyl)aminobutyric acid; Aha, azidohomoalanine; Anv, azidonorvaline; Anl, azidonorleucine; Pgl, propargylglycine; DBE, 3,5-dinitrobenzyl ester. Codon assignment of each amino acid is shown in parentheses.

In the present study, we chose the Leu codon (CUC) for the assignment of Aha and the Thr codon (ACC) for Pgl. Each amino acid activated with a 3,5-dinitrobenzyl ester (DBE) (Figure 1B) was charged onto orthogonal $tRNA^{Asn\text{-}E1}$ using flexizyme,[9] a tRNA acylation ribozyme (Supporting Information Figure S1). Then, these aminoacyl-tRNAs, Aha-$tRNA^{Asn\text{-}E1}_{GAG}$ and Pgl-$tRNA^{Asn\text{-}E1}_{GGU}$, were added in a PURE system in which both Thr and Leu were withdrawn (referred to as *w*PURE system). We observed double incorporations of ^{5}Aha and ^{11}Pgl into the nascent chain of a model peptide **1** (Figure 2) in 36% yield relative to the yield of wild-type expressed under the normal genetic code (Supporting Information Figure S2A,B); this value translated to the quantity of **1** with an approximately 0.64 pmol/μL compared with 1.75 pmol/μL of the wild-type.[10] MALDI-TOF analysis of the peptide showed a single major peak corresponding to **1** accompanied by a minor peak corresponding to an amine molecule **3** (see more details in the legend of Figure 2).[11] We also demonstrated that the replacement of Aha with its derivatives bearing a longer arm (Figure 1B, Anv, n = 2, and Anl, n = 3) yielded the desired peptides with nearly equivalent quantities and qualities to **1** (Supporting Information Figure S2B,C). Thus, the genetic code reprogramming approach enables us to prepare Aha/Pgl-containing peptides.

To test if the inter-side-chain closure of ^{5}Aha/^{11}Pgl-containing peptide **1** could be achieved by the cycloaddition, **1** was treated with $CuSO_4$ and ascorbate (Figure 2, **1**→**2**). Because we expected that both the unreacted linear form and the cyclic form generated by the cycloaddition would show no change in molecular weight, the occurrence of cycloaddition was confirmed by reducing the azide

[†] Research Center for Advanced Science and Technology.
[‡] Department of Chemistry and Biotechnology.
[§] Department of Advanced Interdisciplinary Studies.

10.1021/ja800953c CCC: $40.75

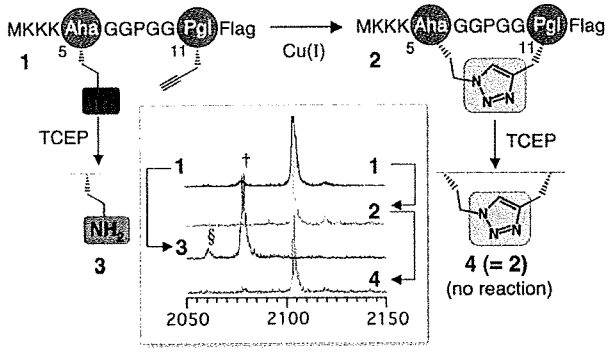


Figure 2. Cyclization via the Cu(I)-catalyzing azide–alkyne cycloaddition. The observed molecular mass (m_{obs}) of **1** is 2103.52 Da, while the calculated molecular mass (m_{calc}) is 2102.96 Da. Likewise, **2** is m_{obs} = 2103.47 and m_{calc} = 2102.96, while **3** is m_{obs} = 2076.54 Da and m_{calc} = 2076.97 Da. Upon treatment of **2** with TCEP, no change in m_{obs} was detected (m_{obs} = 2102.56). The minor peak shown by † in the spectrum of **1** is consistent with m_{obs} of **3**, possibly generated by the reduction of the azide group by DTT and/or mercaptoethanol contained in the translation buffer;[11] § corresponds to a peak of an imido ester likely produced by a similar mechanism reported by Back et al.[12]

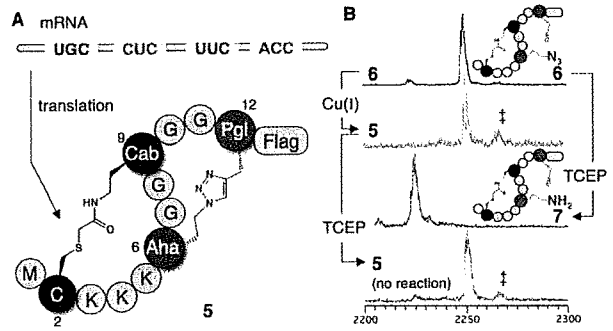


Figure 3. Ribosomal synthesis of bicyclic peptides. (A) Schematic structure of the bicyclic peptide **5** containing ^{9}Cab, ^{6}Aha, and ^{12}Pgl. (B) MALDI-TOF analysis of the nascent peptide **6** (m_{obs} = 2248.29 Da, m_{calc} = 2248.97 Da), which was spontaneously cyclized by the thioether bond between Cab and Cys side chains. TCEP reduction of peptide **6** resulted in the formation of the reduced peptide **7** (m_{obs} = 2222.94 Da, m_{calc} = 2222.98 Da), while bicyclic peptide **5** produced by the copper-catalyzed cycloaddition of **6** was not reduced (m_{obs} = 2248.36); ‡ denotes a peak of peptide **5** in which Met was oxidized during the MALDI-TOF analysis.

group by incubating with tris(carboxyethyl)phosphine (TCEP). Prior to the cycloaddition, the azide group of **1** was readily reduced by TCEP to yield the corresponding amine **3**, consistent with its MALDI-TOF analysis (Figure 2, **1**→**3**). In contrast, the TCEP treatment of the product generated by incubation of **1** with Cu(I) showed no change in the molecular weight (Figure 2, **2**→**4**). This result confirmed that **1** was converted to **2** by the cycloaddition. Likewise, ^{5}Anv/^{11}Pgl- or ^{5}Anl/^{11}Pgl-containing peptide, in which the ring structure was assisted by a longer arm, also afforded the corresponding cyclic peptide upon the Cu(I)-catalyzing reaction (Supporting Information Figure S4).

With the above cyclization methodology in our hands, we next applied it to the synthesis of bicyclic peptides in combination with the intramolecular thioether formation between Cab and Cys. We prepared an mRNA template for the expression of peptide **5** containing ^{2}Cys, ^{6}Aha, ^{9}Cab, and ^{12}Pgl in the peptide sequence (Figure 3A). ^{9}Cab was assigned to a Phe codon (UUC) by charging onto $tRNA^{Asn\text{-}E1}_{GAA}$ using flexizyme. We then performed the triple sense suppressions of ^{6}Aha, ^{12}Pgl, and ^{9}Cab using a *w*PURE system lacking Leu, Thr, and Phe. MALDI-TOF analysis of the peptide showed a major peak corresponding to the monocyclic **6** closed by a thioether bond between the side chains of ^{2}Cys and ^{9}Cab (Figure 3B, **6**). These results indicated that the "first" eight-membered ring cyclization occurred in the translation mixture in situ cleanly. **6** was then subjected to the "second" seven-membered ring cyclization step in the presence of Cu(I) to yield **5**, which was further treated with TCEP. As expected, the molecular weight of the product (**5**) did not show any changes before or after TCEP treatment (Figure 3B, **6**→**5**→**5**), as opposed to the observation that the Cu(I)-untreated **6** was reduced to the corresponding amine **7** by TCEP (**6**→**7**). This result indicates that peptide **5** has the expected bicyclic structure with an eight/seven-membered ring fusion.

Using the same DNA template that expressed **5**, Aha was substituted with Anv or Anl by simply altering the use of the corresponding Anv- or Anl-$tRNA^{Asn\text{-}E1}_{GAG}$ in the translation. The respective precursor peptide containing ^{2}Cys, ^{9}Cab, and ^{12}Pgl along with ^{6}Anv or ^{6}Anl was expressed under the reprogrammed genetic code and treated with Cu(I). Again, to ensure the cycloaddition, these peptides as well as the precursor monocyclic peptides were treated with TCEP, and the resulting peptides were analyzed by MALDI-TOF (Supporting Information Figure S5). The data were indeed consistent with the idea that the desired bicylic peptides were produced upon the Cu(I) treatment.

Encouraged by the above observations, we next attempted to apply this methodology to the expression of peptides composed of different sequences with other combinations of ring sizes. We designed three DNA templates (see Tables 1 and 2 in Supporting Information) that would produce precursor peptides **8**, **11**, and **14** (Figure 4A–C) under the assignments of Aha, Cab, and Pgl to the Leu, Phe, and Thr codons. It was expected that these precursor peptides were spontaneously closed by the first sulfhydryl–chloroacetyl cyclization upon the translation and then converted to bicyclic peptides **10**, **13**, and **16**, respectively, via the second azide–alkyne cycloaddition catalyzed by Cu(I) (Figure 4A–C, left panels). Consequently, peptide **10** had the same bicyclic structure as **5** with an eight/seven-membered ring fusion, but its sequence composition of amino acids in each ring differed from that of **5**. Moreover, peptides **13** and **16** were composed of not only different amino acid sequences but also different bicyclic structures with eight/eight- and seven/seven-membered ring fusions, respectively. To confirm the occurrence of azide–alkyne cycloaddition, the precursor as well as cyclized peptides were treated with TCEP, and these peptides were analyzed by MALDI-TOF (Figure 4A–C, right panels). As expected, the precursor peptides **8**, **11**, and **14** were reduced by TCEP to yield the corresponding monocyclic amino peptides, **9**, **12**, and **15**. In contrast, no change in molecular weight was observed for the Cu(I)-treated peptides, indicating that the cyclization of **8**, **11**, and **14** took place to yield bicyclic peptides **10**, **13**, and **16**, respectively. These results have clearly shown the versatility of this methodology to the synthesis of bicyclic peptides with various sequences and ring sizes in the 7/7, 8/7, and 8/8 fusions.

In conclusion, we have developed a new methodology to construct bicyclic scaffolds by the combination of peptide translation under a reprogrammed genetic code with post-translational cyclizations involving two pairs of amino acids, Cys–Cab and Aha–Pgl. Because the sulfhydryl–chloroacetyl reaction and the azide–alkyne cycloaddition are orthogonal, they act as the specific pairs to yield only the desirable bicyclic peptide with fairly clean conversions of each step. Most importantly, we have shown the versatility of this methodology to the peptide synthesis of various bicyclic peptides bearing fused ring scaffolds, which are controlled by the side chain length of the azide-containing nonproteinogenic amino acids and even the amino acid sequence compositions between the cyclizing

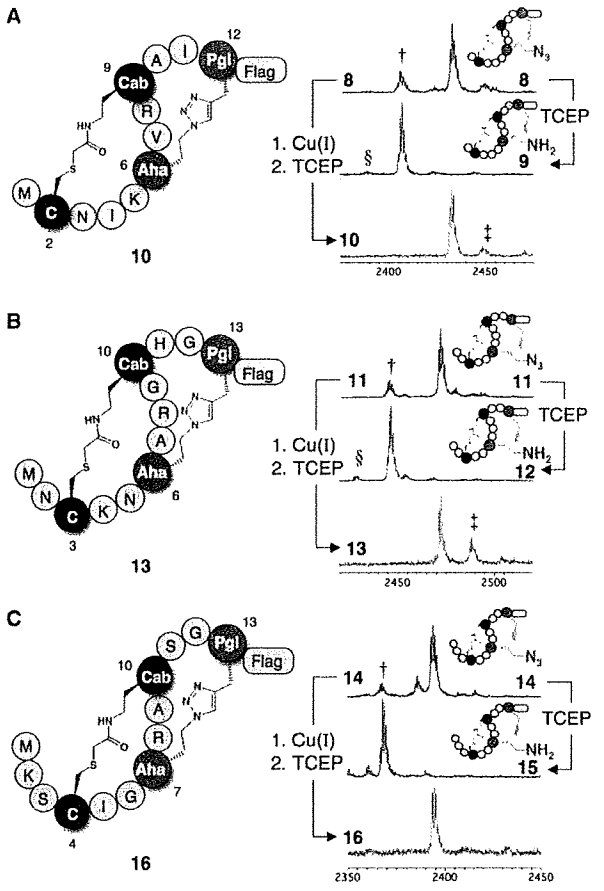


Figure 4. Ribosomal synthesis of bicyclic peptides from three different template DNAs. Left panel shows the schematic structures of bicyclic peptides **10**, **13**, and **16**. Right panel shows MALDI-TOF analysis of the precursor monocyclic peptides **8**, **11**, and **14**, the TCEP-reduced peptides **9**, **12**, and **15**, and Cu(I) followed by TCEP-treated peptides **10**, **13**, and **16**. (A) Formation of bicyclic peptide **10**. Molecular weights of the observed peaks are as follows: **8**, m_{obs} = 2430.58 Da (m_{calc} = 2431.12); **9**, m_{obs} = 2405.82 Da (m_{calc} = 2405.13); bicyclic peptide **10**, m_{obs} =2431.46 Da (m_{calc} = 2431.12). The minor peak indicated by † (m_{obs} = 2404.57 Da) was possibly generated due to the reduction of the azide group by DTT and/or mercaptoethanol contained in the translation buffer.[11] The minor peak indicated by ‡ (m_{obs} = 2448.52 Da) was likely generated by the oxidation of Met of **10** during the MALDI-TOF analysis; § denotes an imido ester produced by a similar mechanism reported by Back et al.[12] (B) Formation of bicyclic peptide **13**. Molecular weights of the observed peaks are as follows: **11**, m_{obs} = 2470.67 Da (m_{calc} = 2471.02); **12**, m_{obs} = 2444.85 Da (m_{calc} = 2445.03); bicyclic peptide **13**, m_{obs} =2471.23 Da (m_{calc} = 2471.02). The minor peaks shown by † and ‡ are m_{obs} = 2444.58 and 2487.75 Da, respectively. (C) Formation of bicyclic peptide **16**. Molecular weights of the observed peaks are as follows: **14**, m_{obs} = 2392.76 Da (m_{calc} = 2393.03); **15**, m_{obs} = 2366.95 Da (m_{calc} = 2367.04); bicyclic peptide **16**, m_{obs} =2393.16 Da (m_{calc} = 2393.03). The minor peak indicated by † corresponds to m_{obs} = 2366.79 Da, whereas the other minor peak observed near 2390 Da was unknown.

residues. Notably, the libraries of bicyclic peptides can be easily constructed by synthesis of the corresponding mRNA libraries. Screening of such peptide libraries with uniform bicyclic scaffolds, perhaps in combination with an appropriate in vitro peptide-display technique, will quickly lead us to discover unique peptidyl drugs against various therapeutic targets.

Acknowledgment. We thank Dr. P. C. Reid for critical proofread. This work was supported by grants of Japan Society for the promotion of Science Grants-in-Aid for Scientific Research (S) (16101007) to H.S., Grants-in-Aid for JSPS Fellows to Y.S. (17-11918), and a research and development projects of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (NEDO).

Supporting Information Available: Experimental details and additional data. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Forte, L. R.; London, R. M.; Freeman, R. H.; Krause, W. J. *Am. J. Physiol. Renal. Physiol.* **2000**, *278*, F180–F191. (b) Sakurai, T.; Amemiya, A.; Ishii, M.; Matsuzaki, I.; Chemelli, R. M.; Tanaka, H.; Williams, S. C.; Richardson, J. A.; Kozlowski, G. P.; Wilson, S.; Arch, J. R.; Buckingham, R. E.; Haynes, A. C.; Carr, S. A.; Annan, R. S.; McNulty, D. E.; Liu, W. S.; Terrett, J. A.; Elshourbagy, N. A.; Bergsma, D. J.; Yanagisawa, M. *Cell* **1998**, *92*, 573–585. (c) Yanagisawa, M.; Kurihara, H.; Kimura, S.; Tomobe, Y.; Kobayashi, M.; Mitsui, Y.; Yazaki, Y.; Goto, K.; Masaki, T. *Nature* **1988**, *332*, 411–415.

(2) Badock, V.; Raida, M.; Adermann, K.; Forssmann, W. G.; Schrader, M. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1952–1956.

(3) (a) Thornton, J. M. J. *Mol. Biol.* **1981**, *151*, 261–287. (b) Richardson, J. S. *Adv. Protein Chem.* **1981**, *34*, 167–339.

(4) (a) Ellgaard, L.; Ruddock, L. W. *EMBO Rep.* **2005**, *6*, 28–32. (b) Gilbert, H. F. *J. Biol. Chem.* **1997**, *272*, 29399–29402.

(5) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001**, *19*, 751–755.

(6) Sako, Y.; Goto, Y.; Murakami, H.; Suga, H. *ACS Chem. Biol.* **2008**, *3*, 241–249.

(7) (a) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J. Am. Chem. Soc.* **2003**, *125*, 3192–3193. (b) Speers, A. E.; Adam, G. C.; Cravatt, B. F. *J. Am. Chem. Soc.* **2003**, *125*, 4686–4687. (c) Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16793–16797. (d) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9482–9487. (e) Link, A. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 11164–11165.

(8) (a) Link, A. J.; Vink, M. K.; Tirrell, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 10598–10602. (b) Forster, A. C.; Tan, Z.; Nalam, M. N.; Lin, H.; Qu, H.; Cornish, V. W.; Blacklow, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6353–6357.

(9) (a) Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. *Nat. Methods* **2006**, *3*, 357–359. (b) Ohuchi, M.; Murakami, H.; Suga, H. *Curr. Opin. Chem. Biol.* **2007**, *11*, 135–144. (c) Kang, T.-J.; Suga, H. *Biochem. Cell Biol.* **2008**, *86*, 92–99. (d) Ohta, A.; Yamagishi, Y.; Suga, H. *Curr. Opin. Chem. Biol.* **2008**, *12*, 159–167. (e) Goto, Y.; Ohta, A.; Sako, Y.; Yamagishi, Y.; Murakami, H.; Suga, H. *ACS Chem. Biol.* **2008**, *3*, 120–129. (f) Kawakami, T.; Murakami, H.; Suga, H. *Chem. Biol.* **2008**, *15*, 32–42. (g) Ohta, A.; Murakami, H.; Higashimura, E.; Suga, H. *Chem. Biol.* **2007**, *14*, 1315–1322.

(10) Our general method for the peptide quantification was as follows: Radioisotope (RI) counts of [^{14}C]-Asp were plotted against its known concentrations to determine the calibration line. Then, RI counts of the expressed wild-type peptide and cyclic peptides were fitted to the calibration line to estimate the individual concentration of peptides. We found that this quantification method was reliable to assess the quantity of peptide previously proven by performing activity assays using a known bioactive peptide.[6]

(11) During the translation, the azide group of Aha, Anv, and Anl was very likely reduced by DTT and/or mercaptoethanol included in the translation buffer. See the following reference as an example of chemistry: Meinjohanns, E.; Meldal, M.; Jensen, T.; Werdelin, O.; GalliStampino, L.; Mouritsen, S.; Bock, K. *J. Chem. Soc., Perkin Trans. 1* **1997**, 871–884.

(12) Back, J. W.; David, O.; Kramer, G.; Masson, G.; Kasper, P. T.; de Koning, L. J.; de Jong, L.; van Maarseveen, J. H.; de Koster, C. G. *Angew. Chem., Int. Ed.* **2005**, *44*, 7946–7950. We found that the reported mechanism herein also proceeds in the translated peptides containing appropriate nonproteinogenic amino acids (Nakajima, E.; et al. unpublished results). More detailed studies of this work will be reported elsewhere.

JA800953C

J|A|C|S
COMMUNICATIONS

Published on Web 11/19/2008

Ribosomal Synthesis of Polypeptoids and Peptoid–Peptide Hybrids

Takashi Kawakami,[†] Hiroshi Murakami,[‡] and Hiroaki Suga*[,†,‡]

Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan, and Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan

Received September 4, 2008; E-mail: hsuga@rcast.u-tokyo.ac.jp

Peptoids are artificially designed synthetic oligomers composed of N-substituted glycine building blocks.[1] It has been shown that peptoids exhibit unique pharmacokinetic properties, different from ordinary peptides, with greater protease resistance and membrane permeability being attributed to the lack of protons on the amide bond.[2] The N-substituted groups represent the secondary structural diversity of peptoids like peptide sidechains, allowing for construction of libraries and leading to the discovery of aptamers (or inhibitors) against therapeutic targets.[3] However, peptoids have intrinsic poor plasticity to form a variety of 3D structures, and therefore the 3D structural diversity in the peptoid library would be limited. Alternatively, a peptoid–peptide hybrid would be an attractive framework to improve the plasticity perhaps without sacrificing the major pharmacokinetics of the peptoid. However, such an approach to date relied on a semirational design of hybrids based on an active peptide followed by their activity validation,[4] giving only a limited number of successes. Clearly, a new synthetic strategy of peptoids or peptoid–peptide hybrids, which is readily extendable to their diverse library construction and screening (or selection) of active ligands, is awaited for the discovery of potent molecules. We here report a novel method to express polypeptoids and peptoid–peptide hybrids by means of translation machinery under the reprogrammed genetic code, allowing for the mRNA-directed synthesis of such oligomers with linear and cyclic scaffolds.

Despite the fact that *N*-methyl-α-amino acids are difficult to efficiently incorporate without prematurely truncated products or oligomerize into a nascent peptide chain using the classical amber suppression method,[5] we and other groups recently found that nearly a dozen *N*-methyl-α-amino acids derived from proteinogenic or nonproteinogenic amino acids could be incorporated by means of translation machinery using genetic code reprogramming.[6] Success, at least in our case, relied on the integration of two systems. One is a reconstituted *E. coli* cell-free translation system in which certain proteinogenic amino acids and/or cognate aminoacyl-tRNA synthetases are withdrawn (referred to as *withdrawn* PURE system; *w*PURE), allowing us to diminish the competing background incorporation of the proteinogenic amino acids in the translation elongation event.[7] The other is a system involving artificial tRNA acylation ribozymes, referred to as flexizymes.[8] This has greatly facilitated the preparation of a wide variety of tRNAs charged with nonproteinogenic amino acids.[6g,8c,9] During our studies, we found the flexizyme system was able to charge N-substituted glycines (rGly) onto tRNAs efficiently (Supporting Information, Figure S1). We thus became interested in investigating if rGly could be incorporated into a nascent peptide chain under the reprogrammed genetic code.

To investigate the incorporation efficiencies of rGly into a peptide chain by translation machinery, we chose a systematic set of rGly bearing various *N*-substituted groups (Figure 1A). The N-substituted groups were divided into three families, (1) nonbranched alkyl chains (**1**–**8**), (2) branched alkyl chains (**9**–**14**), and (3) functionalized alkyl chains (**15**–**30**). These rGly were derived to 3,5-dinitrobenzyl esters or cyanomethyl esters, making them compatible to the flexizyme system,[8c] and then charged onto suppressor $tRNA^{Asn\text{-}E2}_{GGU}$[9a] (Figure S1). The individual rGly-$tRNA^{Asn\text{-}E2}_{GGU}$ were added to a *w*PURE system lacking threonine (Thr) and surveyed for their single incorporation into a model peptide upon

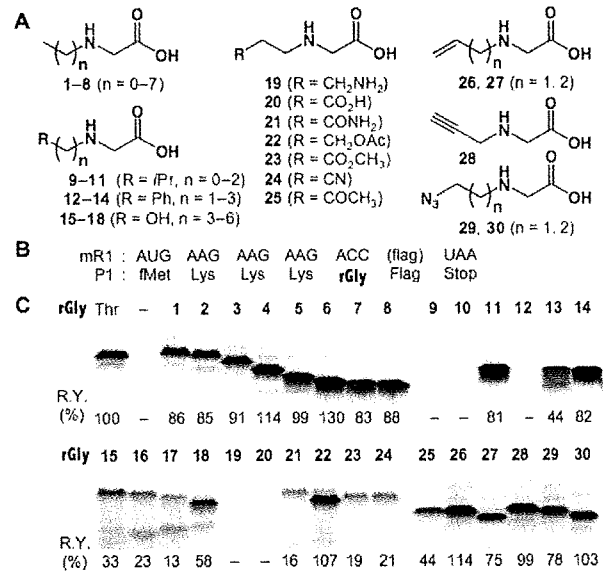


Figure 1. Tolerance of N-substituted glycines (rGly) in ribosomal peptide synthesis. (A) Chemical structure of rGly used in this study. **1**, *N*-methylglycine (***N*me**); **2**, *N*-ethyl-glycine (***N*et**); **3**, *N*-*n*-propyl-glycine (***N*pr**); **4**, *N*-*n*-butyl-glycine (***N*bu**); **5**, *N*-*n*-pentyl-glycine; **6**, *N*-*n*-hexyl-glycine; **7**, *N*-*n*-heptyl-glycine; **8**, *N*-*n*-octyl-glycine; **9**, *N*-isopropyl-glycine; **10**, *N*-i-butyl-glycine; **11**, *N*-isopentyl-glycine; **12**, *N*-benzyl-glycine; **13**, *N*-(2-phenylethyl)-glycine; **14**, *N*-(3-phenylpropyl)-glycine; **15**, *N*-(3-hydroxypropyl)-glycine; **16**, *N*-(4-hydroxybutyl)-glycine; **17**, *N*-(5-hydroxypentyl)-glycine; **18**, *N*-(6-hydroxyhexyl)-glycine; **19**, *N*-(3-aminopropyl)-glycine; **20**, *N*-(2-carboxylethyl)-glycine; **21**, *N*-(2-carbamoylethyl)-glycine; **22**, *N*-(3-acetoxypropyl)-glycine (***N*aco**); **23**, *N*-[2-(methoxycarbonyl)ethyl]-glycine; **24**, *N*-(2-cyanoethyl)-glycine; **25**, *N*-(3-oxobutyl)-glycine; **26**, *N*-allyl-glycine (***N*all**); **27**, *N*-3-butenyl-glycine; **28**, *N*-propargyl-glycine; **29**, *N*-(2-azidoethyl)-glycine; **30**, *N*-(3-azidopropyl)-glycine. (B) Sequences of mRNA (mR1) that express an rGly-containing-peptide (P1). Flag in parentheses indicates the RNA sequence encoding the Flag peptide (DYKDDDDK). (C) Tricine-SDS-PAGE analysis of the expressed peptides labeled with [^{14}C]-Asp detected by autoradiography. Lane 1, the wildtype peptide expressed in the Thr-containing ordinary PURE system where ACC assigns Thr; lane 2, a negative control using Thr-withdrawn PURE (*w*PURE) system in the presence of uncharged $tRNA^{Asn\text{-}E2}_{GGU}$; lanes 3–32, expression of the peptide containing a single rGly in the presence of designated rGly-$tRNA^{Asn\text{-}E2}_{GGU}$ prepared by flexizyme system. Each expression yield (R.Y.) relative to wildtype was determined by a mean score of triplicates.

[†] Department of Chemistry and Biotechnology.
[‡] Research Center for Advanced Science and Technology.

10.1021/ja806998v CCC: $40.75

suppression of the Thr ACC codon in the designed mRNA sequence (Figure 1B, mR1 and P1). Tricine-SDS-PAGE analysis of the resulting [^{14}C]-Asp-labeled peptides showed that the first family of rGly bearing a nonbranched alky chain were all incorporated as efficiently as Thr (Figure 1C, **1**−**8**). The incorporation of the designated rGly was also confirmed by MALDI-TOF analysis, giving the expected molecular mass in all cases (Figure S2).

In the second family, three rGly (Figure 1C, **11**, **13**, and **14**) were efficient elongators similar to those in the first family, whereas the other three (Figure 1C, **9**, **10**, and **12**) having the branched methyl or phenyl group closer to the α-amino group were not. Most rGly in the third family exhibited acceptable incorporation efficiencies (Figure 1C, **15**−**18** and **21**−**30**), in which the MALDI-TOF was consistent with the expected molecular mass (Figure S2). The exceptions were those with a negatively or positively charged alkyl group (Figure 1C, **19** and **20**). In summary, our survey has shown that a wide variety of rGly, in which the *N*-substituent has no bulky group near the α-amino group and no charged functional group, act as elongators for translation machinery. This observation is similar to that for *N*-methyl-α-amino acids in which bulky or charged side chains were unfavorable elongators;[6g] a short list of unfavorable elongators recently reported by Zhang et al. also fell into in this trend, where *N*-substituted α-amino acids were bulky near the α-amino group due to the combination between a N-substituent and α-carbon side chain.[6f] Therefore, this seems to be a common trend for N-substituted α-amino acids.

We next attempted consecutive elongations of rGly to express a short stretch of polypeptoid sequences fused with a KK-Flag peptide tag to facilitate its isolation. In this study, we chose three rGly, ***N*et**, ***N*aco**, and ***N*all** (Figure 1A, **2**, **22**, and **26**) that were assigned to arbitrarily chosen ACC, UUC, and CUC codons, respectively, with aid of the flexizyme system. Five mRNA templates (Figure 2A, mR2−6) were designed to elongate the above rGly in succession ranging from 2 to 6 times (Figure 2A, P2−P6). We also expressed peptides (Figure 2A, wt2−6) from the respective mRNA templates using the ordinary PURE system to compare the expression level of peptoids. Tricine-SDS-PAGE analysis of [^{14}C]-Asp-labeled peptoids and peptides indicated that the expression level of peptoids gradually decreased with increasing peptoid length (lanes 2, 4, 6, 9, and 11), in contrast to the observation that peptides were expressed at a nearly consistent level independent from the length (Figure 2B, lanes 1, 3, 5, 8, and 10). Although this observation is somewhat similar to the successive elongation of *N*-methyl-α-amino acids,[6g] the expression level of peptoids was hampered more seriously by an increase in length. In fact, when we expressed a tetra-*N*-methyl-peptide using the mR4 template (Me4, see the box in Figure 2A and B), the expression level of Me4 was nearly double that of P4 (Figure 2B, lane 6 vs 7). Both ribosome and EF-Tu may have an effect on their elongation efficiencies.[10] In particular, peptoidyl elongation may be attributed to steric hindrances of the peptoidyl-tRNA in ribosome peptidyl-tRNA (P) site or/and the rGly-tRNA in aminoacyl-tRNA (A) site, probably making the peptoidyl elongation less efficient than *N*-methyl-peptidyl elongation.

To demonstrate the expression of designer peptoids, we prepared two sequence variants of the tetrapeptoid P4 (Figure 3A, P4a and P4b) and expressed them from the respective mRNA templates (Figure 3A, mR4a and mR4b). The expression levels of these tetrapeptoids were similar to each other, and MALDI-TOF analysis of each peptoid showed a single major peak that is consistent with its expected molecular mass according to the corresponding mRNA templates (Figure 3B). To the best of our knowledge, this result

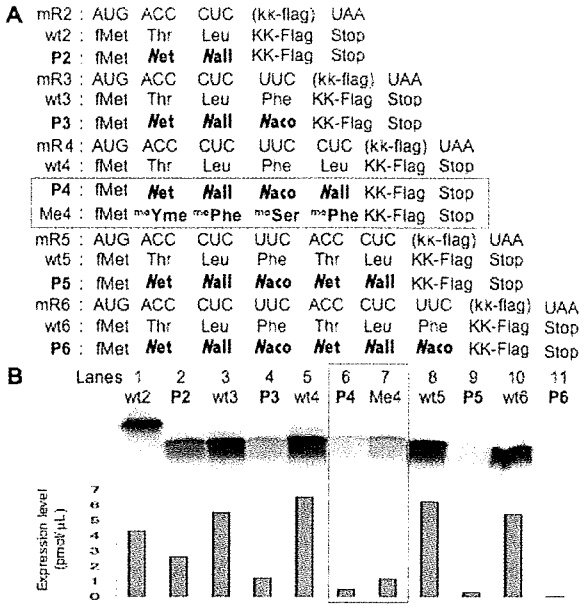


Figure 2. mRNA-programmed synthesis of di-, tri-, tetra-, penta-, and hexapeptoid. (A) Sequences of mRNA templates (mR2−6), control wildtype peptides (wt2−6), a tetra-*N*-methyl-peptide (Me4), and peptoids containing *N*et, *N*all, and *N*aco (P2−6). ^{me}Yme, *N*-methyl-L-*p*-methoxyphenylalanine; ^{me}Phe, *N*-methyl-L-phenylalanine; ^{me}Ser, *N*-methyl-L-serine. (B) Tricine-SDS-PAGE analysis of control wildtype peptides, a tetra-*N*-methyl-peptide, and peptoids expressed from the respective mRNA. The wildtype peptides were expressed in the ordinary PURE system, while peptoids were expressed in the *w*PURE system containing *N*et-tRNA$^{Asn\text{-}E2}_{GGU}$, *N*all-tRNA$^{Asn\text{-}E2}_{GAG}$, and *N*aco-tRNA$^{Asn\text{-}E2}_{GAA}$. Expression level of each peptide labeled with [^{14}C]-Asp based on its observed radioisotope counts is shown in the graph.

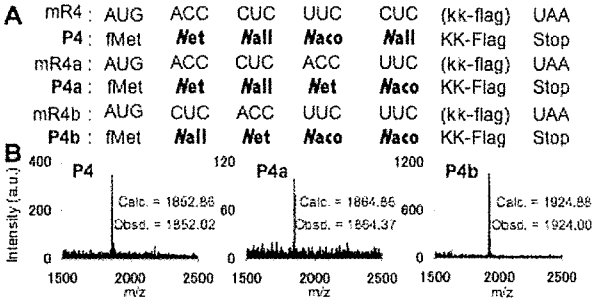


Figure 3. mRNA-programmed synthesis of linear tetrapeptoids (A) Sequences of mRNA templates (mR4, mR4a, and mR4b) and encoded linear peptoids containing *N*et, *N*all, and *N*aco (P4, P4a, and P4b). The kk-flag in parentheses indicates the RNA sequence encoding a KK-Flag peptide (KKDYKDDDDK). The expression level of each tetrapeptoid was estimated by tricine-SDS-PAGE as follows: P4, 0.52 pmol/μL; P4a, 0.37 pmol/μL; P4b, 0.51 pmol/μL. (B) MALDI-TOF-MS spectra of the Flag-purified translation products. The calculated mass (Calc) and observed mass (Obsd) for singly charged species $[M+H]^{+1}$ are shown in each spectrum.

records the first demonstration of mRNA-directed expression of polypeptoids using the translation machinery.

Due to the above success, we extended our methodology to the ribosomal synthesis of peptoid−peptide hybrids. With a future application to drug discovery in mind, we attempted to express peptoid−peptide hybrids with a framework in which peptoid bonds are scattered in cyclic peptides closed by a nonreducible thioether bond.[9b] We reprogrammed the genetic code in which CGC, ACC, CAC, and UCC were assigned to ***N*me**, ***N*et**, ***N*pr**, and ***N*bu** (Figure 1A, **1**−**4**), respectively, for the peptoid synthesis and AUG was

assigned to *N*-chloroacetyl-α-^{L}Phe or *N*-chloroacetyl-α-^{D}Phe[9d] (ClAc-^{L}Phe or ClAc-^{D}Phe) for the formation of cyclic scaffolds (Figure S3B). Four mRNA templates (Figure S3A, mR7−10) were designed to express cyclic peptoid−peptide hybrids with different sequence compositions attached to the C-terminal KK-Flag (Figure 4A, P7−P10). MALDI-TOF analysis of each cyclic peptoid−peptide

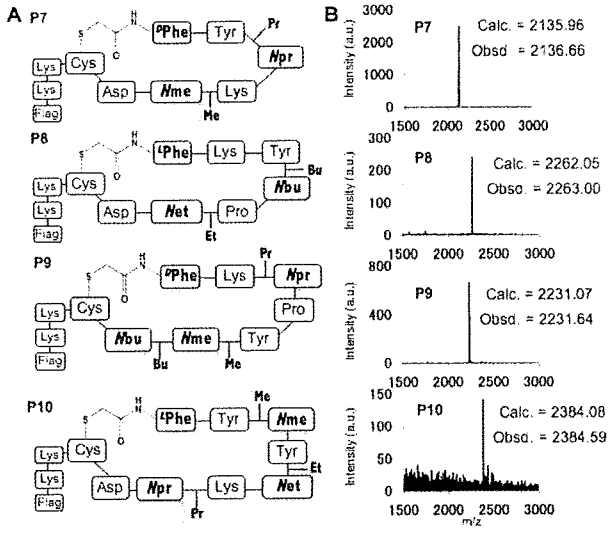


Figure 4. mRNA-programmed synthesis of cyclic peptoid−peptide hybrids. (A) Schematic structure of cyclic peptoid−peptide hybrids containing ***N*me**, ***N*et**, ***N*pr**, and ***N*bu** (P7−10). The locations of *N*-substituted groups on the peptoid bond are highlighted in gray. The expression level of each peptoid−peptide hybrid was estimated by tricine-SDS-PAGE as follows: P7, 1.97 pmol/μL; P8, 2.63 pmol/μL; P9, 0.21 pmol/μL, and P10, 0.84 pmol/μL. (B) MALDI-TOF-MS spectra of the Flag-purified translation products. The calculated mass (Calc) and observed mass (Obsd) for singly charged species $[M+H]^{+1}$ are shown in each spectrum.

hybrid expressed from the respective mRNA template exhibited the expected molecular mass as a single major peak (Figure 4B, P7−P10). This firmly demonstrates the mRNA-directed expression of cyclic peptoid−peptide hybrids under the reprogrammed genetic code.

In conclusion, we have demonstrated mRNA-directed synthesis of linear polypeptoids and cyclic peptoid−peptide hybrids. The translation machinery surprisingly accepts a variety of rGly for elongation, including those with alkyl chains and functional groups that are orthogonal[7d,9c,11] to proteinogenic side chains, such as the ketone, alkene, azide, and alkyne. We have shown consecutive elongations of rGly and synthesis of cyclic peptoid−peptide hybrids. Particularly, the latter hybrids contain two or more peptoid bonds dispersed in the sequence and consist of cyclic structures closed by a physiologically stable thioether bond. Since we have already shown that the thioether cyclization is spontaneously formed and applicable to a variety of ring sizes independent from the sequence compositions of peptide,[6g,9b,c,12] the methodology reported herein offers us a new means of the mRNA-programmed library synthesis of cyclic peptoid−peptide hybrids. This has already directed our current research to running screenings or selections of a new class of peptidic drugs against therapeutic targets by the integration with appropriate techniques.[13] Such efforts are underway in our laboratory.

Acknowledgment. We thank Dr. P. C. Reid for critical proofread. This work was supported by grants of Japan Society for the promotion of Science Grants-in-Aid for Scientific Research (S) (16101007) to H.S., Grants-in-Aid for JSPS Fellows to T.K. (20-664), and a research and development projects of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (NEDO).

Supporting Information Available: Experimental details and additional data. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Simon, R. J.; Kania, R. S.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9367–9371. (b) Patch, J. A.; Barron, A. E. *Curr. Opin. Chem. Biol.* **2002**, *6*, 872–877. (c) Shin, S. B.; Yoo, B.; Todaro, L. J.; Kirshenbaum, K. *J. Am. Chem. Soc.* **2007**, *129*, 3218–3225. (d) Lee, B. C.; Chu, T. K.; Dill, K. A.; Zuckermann, R. N. *J. Am. Chem. Soc.* **2008**, *130*, 8847–8855.
(2) Kwon, Y. U.; Kodadek, T. *J. Am. Chem. Soc.* **2007**, *129*, 1508–1509.
(3) (a) Zuckermann, R. N.; et al. *J. Med. Chem.* **1994**, *37*, 2678–2685. (b) Alluri, P. G.; Reddy, M. M.; Bachhawat-Sikder, K.; Olivos, H. J.; Kodadek, T. *J. Am. Chem. Soc.* **2003**, *125*, 13995–14004. (c) Wrenn, S. J.; Weisinger, R. M.; Halpin, D. R.; Harbury, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 13137–13143. (d) Udugamasooriya, D. G.; Dineen, S. P.; Brekken, R. A.; Kodadek, T. *J. Am. Chem. Soc.* **2008**, *130*, 5744–5752.
(4) (a) Nguyen, J. T.; Turck, C. W.; Cohen, F. E.; Zuckermann, R. N.; Lim, W. A. *Science* **1998**, *282*, 2088–2092. (b) Kruijtzer, J. A.; Nijenhuis, W. A.; Wanders, N.; Gispen, W. H.; Liskamp, R. M.; Adan, R. A. *J. Med. Chem.* **2005**, *48*, 4224–4230.
(5) (a) Bain, J. D.; Wacker, D. A.; Kuo, E. E.; Chamberlin, A. R. *Tetrahedron* **1991**, *47*, 2389–2400. (b) Ellman, J. A.; Mendel, D.; Schultz, P. G. *Science* **1992**, *255*, 197–200. (c) Short, G. F., 3rd; Laikhter, A. L.; Lodder, M.; Shayo, Y.; Arslan, T.; Hecht, S. M. *Biochemistry* **2000**, *39*, 8768–8781.
(6) (a) Frankel, A.; Millward, S. W.; Roberts, R. W. *Chem. Biol.* **2003**, *10*, 1043–1050. (b) Merryman, C.; Green, R. *Chem. Biol.* **2004**, *11*, 575–582. (c) Tan, Z.; Forster, A. C.; Blacklow, S. C.; Cornish, V. W. *J. Am. Chem. Soc.* **2004**, *126*, 12752–12753. (d) Tan, Z.; Blacklow, S. C.; Cornish, V. W.; Forster, A. C. *Methods* **2005**, *36*, 279–290. (e) Hartman, M. C.; Josephson, K.; Lin, C. W.; Szostak, J. W. *PLoS ONE* **2007**, *2*, e972. (f) Zhang, B.; Tan, Z.; Dickson, L. G.; Nalam, M. N.; Cornish, V. W.; Forster, A. C. *J. Am. Chem. Soc.* **2007**, *129*, 11316–11317. (g) Kawakami, T.; Murakami, H.; Suga, H. *Chem. Biol.* **2008**, *15*, 32–42. (h) Ohta, A.; Yamagishi, Y.; Suga, H. *Curr. Opin. Chem. Biol.* **2008**, *12*, 159–167. (i) Kang, T. J.; Suga, H. *Biochem. Cell. Biol.* **2008**, *86*, 92–99. (j) Subtelny, A. O.; Hartman, M. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2008**, *130*, 6131–6136. (k) Sando, S.; Masu, H.; Furutani, C.; Aoyama, Y. *Org. Biomol. Chem.* **2008**, *6*, 2666–2668.
(7) (a) Forster, A. C.; Weissbach, H.; Blacklow, S. C. *Anal. Biochem.* **2001**, *297*, 60–70. (b) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001**, *19*, 751–5. (c) Forster, A. C.; Tan, Z.; Nalam, M. N.; Lin, H.; Qu, H.; Cornish, V. W.; Blacklow, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6353–6357. (d) Josephson, K.; Hartman, M. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 11727–11735. (e) Shimizu, Y.; Kanamori, T.; Ueda, T. *Methods* **2005**, *36*, 299–304. (f) Shimizu, Y.; Kuruma, Y.; Ying, B. W.; Umekage, S.; Ueda, T. *FEBS J.* **2006**, *273*, 4133–4140.
(8) (a) Xiao, H.; Murakami, H.; Suga, H.; Ferre-D'Amare, A. R. *Nature* **2008**, *454*, 358–361. (b) Ohuchi, M.; Murakami, H.; Suga, H. *Curr. Opin. Chem. Biol.* **2007**, *11*, 537–542. (c) Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. *Nat. Methods* **2006**, *3*, 357–359.
(9) (a) Ohta, A.; Murakami, H.; Higashimura, E.; Suga, H. *Chem. Biol.* **2007**, *14*, 1315–1322. (b) Goto, Y.; Ohta, A.; Sako, Y.; Yamagishi, Y.; Murakami, H.; Suga, H. *ACS Chem. Biol.* **2008**, *3*, 120–129. (c) Sako, Y.; Morimoto, J.; Murakami, H.; Suga, H. *J. Am. Chem. Soc.* **2008**, *130*, 7232–7234. (d) Goto, Y.; Murakami, H.; Suga, H. *RNA* **2008**, *14*, 1390–1398.
(10) LaRiviere, F. J.; Wolfson, A. D.; Uhlenbeck, O. C. *Science* **2001**, *294*, 165–168.
(11) (a) Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652–5653. (b) Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 11782–11783. (c) Link, A. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 11164–11165. (d) Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56–61. (e) Chang, P. V.; Prescher, J. A.; Hangauer, M. J.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2007**, *129*, 8400–8401. (f) Tam, A.; Soellner, M. B.; Raines, R. T. *J. Am. Chem. Soc.* **2007**, *129*, 11421–11430. (g) Slavoff, S. A.; Chen, I.; Choi, Y. A.; Ting, A. Y. *J. Am. Chem. Soc.* **2008**, *130*, 1160–1162.
(12) Sako, Y.; Goto, Y.; Murakami, H.; Suga, H. *ACS Chem. Biol.* **2008**, *3*, 241–249.
(13) (a) Hanes, J.; Pluckthun, A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 4937–4942. (b) He, M.; Taussig, M. J. *Nucleic Acids Res.* **1997**, *25*, 5132–5134. (c) T Nemoto, N.; Miyamoto-Sato, E.; Husimi, Y.; Yanagawa, H. *FEBS Lett.* **1997**, *414*, 405–408. (d) Roberts, R. W.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12297–12302. (e) Mastrobattista, E.; Taly, V.; Chanudet, E.; Treacy, P.; Kelly, B. T.; Griffiths, A. D. *Chem. Biol.* **2005**, *12*, 1291–1300. (f) Wrenn, S. J.; Harbury, P. B. *Annu. Rev. Biochem.* **2007**, *76*, 331–349.

JA806998V

Chemistry & Biology

Expression of Histone H3 Tails with Combinatorial Lysine Modifications under the Reprogrammed Genetic Code for the Investigation on Epigenetic Markers

Taek Jin Kang,[1,3] Satoshi Yuzawa,[2,3] and Hiroaki Suga[1,2,*]
[1]Research Center for Advanced Science and Technology, The University of Tokyo, 153-8904 Tokyo, Japan
[2]Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 113-8656 Tokyo, Japan
[3]These authors contributed equally to this work.
*Correspondence: hsuga@rcast.u-tokyo.ac.jp
DOI 10.1016/j.chembiol.2008.09.014

SUMMARY

We report the ribosomal synthesis of N-terminal peptides of histone H3, so-called H3 tail (H3t), with combinatorial methyl and acetyl modifications of selected lysine residues, and the application of such peptides to studying the influence of lysine modification on H3t binding to chromodomain of heterochromatin protein 1 (chromoHP1). Genetic code reprogramming was employed to reassign four codons to acetylated, mono-, di-, and trimethylated lysines, and 38-mer H3t peptides containing modified lysines at designated sites were expressed from the corresponding mRNA sequences. Using a series of H3t constructs, we show complex crosstalk among methylated lysine 9 and 27, and acetylated lysine 14 for binding to chromoHP1. This proof-of-concept study offers a unique means for the synthesis of not only an H3t library containing modified lysines but also other classes of peptides bearing posttranslational methylation and acetylation.

INTRODUCTION

A nucleosome consists of 146 base pairs of DNA wrapped around an octamer of core histone proteins, forming the basic repeating unit of chromatin. Each of the four core histones, H2A, H2B, H3, and H4, contains a structural domain consisting of three α helices and an unstructured domain at the N terminus or C terminus, referred to as the histone tail. Enzyme-catalyzed chemical modification of certain residues in the histone tails serves to regulate the local structure of chromatin, altering the accessibility of various proteins to histone tails as well as to DNA. Thus, histone tail modification represents one of the important epigenetic marks linked to various cellular processes, such as replication, transcription, and DNA repair (Kouzarides, 2007).

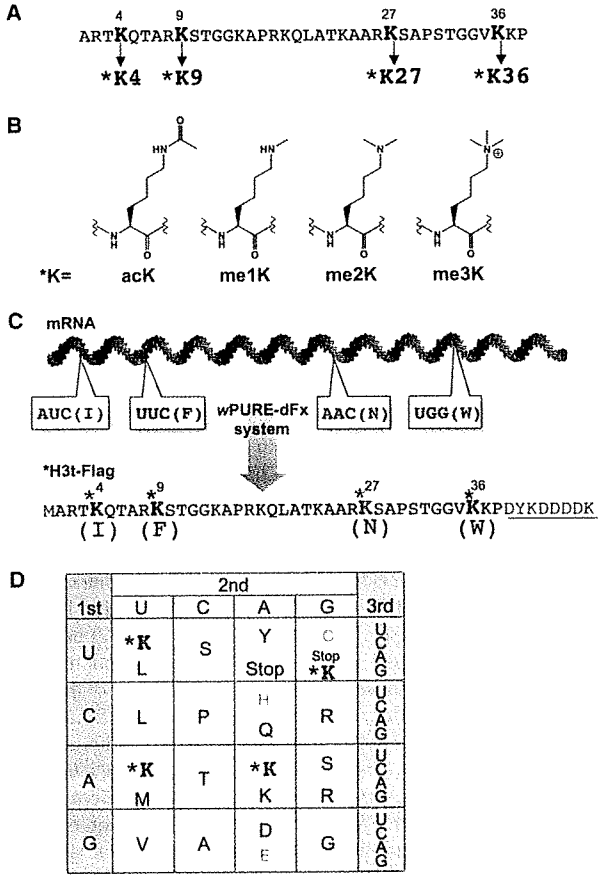
The histone tails of core histones, especially that of H3 (H3t), consists of about 40 amino acid residues and are subjected to various chemical modifications such as acetylation, methylation, phosphorylation, and possibly ubiquitylation (Kouzarides, 2007; Peterson and Laniel, 2004). Lysines (Ks) in H3t can be acetylated (ac; acetylated lysine, acK) or methylated (me), and the methylation can be one of mono-, di-, or trimethylation (me1, me2, and me3, respectively; Figures 1A and 1B). In general, hyper- and hypoacetylation of H3t are linked with transcriptional activation and repression, respectively. Methylation can influence transcription in various ways, depending on the site of the modification (Kouzarides, 2007; Martin and Zhang, 2005). Linking a particular modification or set of modifications to a particular cellular function involving effector proteins postulated a "histone code" hypothesis (Jenuwein and Allis, 2001; Strahl and Allis, 2000). Thus, modifications at H3t can be viewed as special marks modulating the interaction among histones or between the histone and nonhistone proteins. Certain modifications at H3t can also affect the consequences of others, and evidence of such crosstalk between modifications has been accumulating in the past few years (Fischle et al., 2005; Kouzarides, 2007; Latham and Dent, 2007; Seet et al., 2006). Moreover, recent findings of the combinatorial linkages among modifications in H3t also suggested possible long-range crosstalk among modifications of H3t (Garcia et al., 2007; Taverna et al., 2007). However, our current understandings of their interplay at the molecular level are limited, partly because there is no reliable and readily accessible method to prepare the full-length H3t with multiple modifications in a defined and controlled manner.

Genetic code reprogramming is a technique where codons usually assigned to proteinogenic amino acids are reassigned to nonproteinogenic amino acids, including those with unique side chains (Forster et al., 2003; Josephson et al., 2005; Murakami et al., 2006). This enables the expression of peptides containing multiple nonproteinogenic amino acids using a translation system. We envisioned that this technique is suitable for the synthesis of H3t with sparsely positioned multiple modifications. We here report the mRNA-directed synthesis of 38-mer H3t containing Ks modified with N^{ε}-mono-, di-, trimethyl and/or acetyl group(s) and its application to studies on the specific interaction between chromoHP1 and H3t modified at positions 9, 14, and 27. This proof-of-concept study offers a unique means for the synthesis of an H3t library containing modified Ks to investigate the role of combinatorial modifications in histone tail structure and function.

RESULTS

Translatability of Methylated or Acetylated Lysine upon Genetic Code Reprogramming

The 38-mer H3t is composed of 10 members of the proteinogenic amino acids (alanine, arginine, glutamine, glycine, leucine,



1st	2nd U	2nd C	2nd A	2nd G	3rd
U	*K L	S	Y Stop	C Stop *K	U C A G
C	L	P	H Q	R	U C A G
A	*K M	T	*K K	S R	U C A G
G	V	A	D E	G	U C A G

Figure 1. Genetic Code Reprogramming for the Preparation of *H3t

(A) The primary sequence of H3t used in this study. *Ks were incorporated into positions indicated.

(B) *Ks used in this study. Note that both me1K and me2K are protonated in physiological conditions, and thus all the meKs are positively charged.

(C) An example of codon assignments for the genetic code reprogramming. Codons for K4, K9, K27, and K36 were changed to those for I, F, N, and W, respectively. Color codes used here represent the codons in use throughout the paper.

(D) A representative reprogrammed genetic code table. Four reprogrammed codons are in colors, whereas three unused codon boxes, shown in gray, were left vacant.

lysine, proline, serine, threonine, and valine), including eight Ks. Among these Ks, four Ks at positions 4, 9, 27, and 36 (K4, K9, K27, and K36, respectively) were often found methylated in vivo; therefore, we were interested in synthesizing H3t peptides containing Ks specifically modified at these positions (abbreviated to *Ks; Figure 1A). In addition to the these 10 amino acids, methionine, aspartate, and tyrosine (M, D, Y, respectively) were included in the peptide chain to assign the initiation (M) and Flag peptide tag (DYKDDDDK; D could be [^{14}C]-D) at the C terminus for purification, immobilization, and radioisotope labeling of the expressed H3t peptides. Consequently, seven amino acids were yet unused, and four of these codons were reassigned to *Ks modified with me1, me2, me3, or ac on the ε-amino group (Figure 1B). We chose codons of AUC (isoleucine, I), UUC (phenylalanine, F), AAC (asparagine, N), and UGG (tryptophan, W) and reassigned these codons to code for *Ks (Figures 1C and 1D). Thus, these amino acids as well as other unused amino acids were withdrawn from the PURE (protein synthesis using recombinant elements) system to generate a *w*PURE system that contained all the aminoacyl-tRNA synthetases but only the essential 13 amino acids, for genetic code reprogramming. Derivatization of *Ks to 3,5-dinitrobenzyl esters (DBE) made these amino acids compatible with a dinitro-flexizyme, a ribozyme-based tRNA acylating catalyst (dFx; see Murakami et al., 2006). dFx was able to aminoacylate $tRNA^{Asn\text{-}E2}_{GAU}$ with the respective modified *K-DBE (see Figure S1 available online). Because dFx is inherently able to recognize tRNAs bearing any anticodon, this finding ensured that any pair of *K and $tRNA^{Asn\text{-}E2}_{NNN}$ (NNN denotes an anticodon complementary to the reprogrammed codon) could be prepared by the dFx technology.

Though it was known that acK could be incorporated into a nascent peptide chain efficiently by translation under the reprogrammed genetic code (Murakami et al., 2006), no information was available for the incorporation efficiency of meKs. To verify the translatability of the individual *Ks at each position, we designed a template containing codons for I F, N, and W at positions 4, 9, 27, and 36, respectively (Figure 1C). Prior to the suppression experiments, two control experiments were performed: as a positive control, the corresponding DNA template was transcribed and translated in the *w*PURE system that was supplemented with I, F, N, and W. An intense band of a [^{14}C]-labeled peptide was observed on SDS-tricine-PAGE (Figure 2A, lane 1), and MALDI-TOF analysis of the Flag-purified peptide revealed a molecular weight (MW) consistent with that of the expected peptide containing I, F, N, and W at the designated positions instead of Ks (Figure 2B, P-WT). As a negative control to verify the background expression at position 4 (I) without *K4 suppression, the same template was translated in a *w*PURE system supplemented with F, N, and W only. A faint band with nearly the same mobility as the P-WT was observed (Figure 2A, lane 2), and the MALDI-TOF analysis of the Flag-purified peptide suggested the misincorporation of other amino acids (Figure S2). However, this background expression could be readily competed out by suppression using *K-$tRNA^{Asn\text{-}E2}_{GAU}$; upon the addition of the respective *K-$tRNA^{Asn\text{-}E2}_{GAU}$ to the *w*PURE system supplemented with F, N, and W, an intense band appeared irrespective of *K used (lanes 3–6), and the MALDI-TOF analysis of the respective Flag-purified peptide confirmed a peak of the main product with the expected MW containing the designated *K (Figure 2B, P-K4me1, P-K4me2, P-K4me3, and P-K4ac). Likewise, the background peptide expressions at the other three positions, K9, K27, and K36 (lanes 8, 14, and 20), were also suppressed by the addition of the respective *K-$tRNA^{Asn\text{-}E2}_{NNN}$ (NNN denotes the corresponding anticodon) to afford an intense band in the tricine-SDS PAGE (Figure 2A, lanes 9–12, 15–18, and 21–24), and the MW of the respective peptides were also confirmed (Figure 2B, P-KXme1, P-KXme2, P-KXme3, and P-KXac; X = 9, 27, or 36). These findings showed all *Ks were compatible with the ribosomal peptide synthesis upon genetic code reprogramming, yielding the designated peptide as a main product.

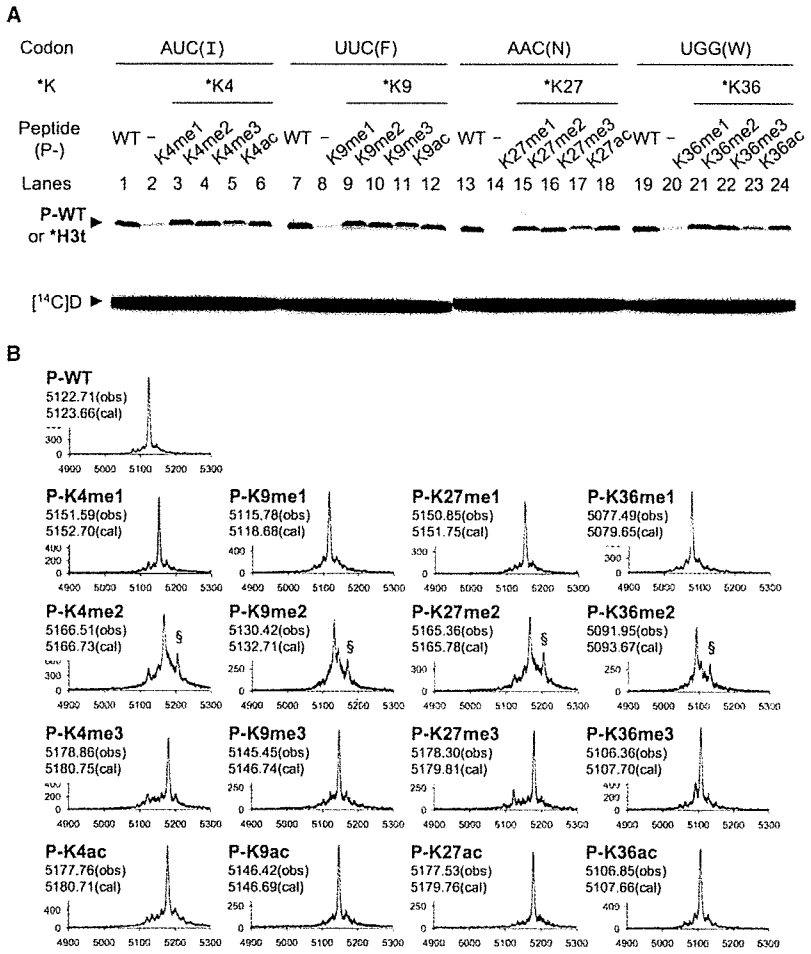


Figure 2. Single Incorporations of *Ks into H3t

(A) Analysis of expressed peptides containing a single *K substitution. Incorporation efficiencies were examined on the 15% SDS-tricine-PAGE using the wild-type peptide that has I4, F9, N27, and W36 as a control. Because the *w*PURE system used in this study did not contain any of these amino acids, a proper combination of amino acids were exogenously added to the translation; for example, F, N, and W were added to the translation for I codon suppressions. Unincorporated [^{14}C]-D is shown at the bottom of the gel.

(B) MALDI-TOF analysis of each *H3t peptide after anti-Flag antibody affinity purification. § indicates a peak of the potassium adduct. Note that an intense peak corresponding to the potassium adduct was present in the series of me2K-containing *H3t.

Interaction of HP1 with the Translated H3t Peptides Containing a Methylated Lysine at the Ninth Position

The next critical question was whether the translated H3t peptides containing the modified Ks (*H3t peptides) were able to exhibit the expected modification-specific interactions with effector proteins. To confirm this, we chose heterochromatin protein 1 (HP1), of which the chromodomain is known to interact with K9-modified *H3t peptides in a methylation-dependent manner (Jacobs et al., 2001; Lachner et al., 2001). Two well-established analytical methods for detecting protein-protein interactions, pull-down analysis and surface plasmon resonance (SPR), were employed to show the meK9-specific recruitment of the chromodomain of HP1 (chromoHP1) by the *H3t peptide.

A template sequence that contained the UUC(F) codon at position 9 was translated to Flag-tagged H3t, *H3tK9ac, *H3tK9me1, *H3tK9me2, and *H3tK9me3 in the *w*PURE system, in the presence of the corresponding *K-tRNA$^{Asn\text{-}E2}_{GAA}$ (Figure 3A). As a negative control, the same template was translated in the absence of aminoacyl-tRNA$^{Asn\text{-}E2}_{GAA}$ in the same *w*PURE system. *H3t peptides were displayed on an agarose resin through covalently immobilized anti-Flag antibody, and the resulting resin was used directly as bait to pull down chromoHP1. The immobilized molecules were acid eluted from the resin and analyzed by immunoblot using anti-Flag and anti-polyhistidine polyclonal antibodies (Figure 3B). The *H3t peptides were detected only with the addition of charged tRNA$^{Asn\text{-}E2}_{GAA}$ (Figure 3B, lanes 2–6 versus lane 1), confirming that the *H3t peptide expression was solely the consequence of genetic code reprogramming. Significantly, chromoHP1 was detected only when *H3t containing meK9 was used as bait (Figure 3B, lanes 2 and 3 versus lanes 4–6; see also Figure S3 for product verification by mass analysis). We clearly show that *H3tK9me3 was able to pull down chromoHP1 the most effectively among the *H3t peptides. Our findings are in good agreement with an earlier report in which the affinity of 15-mer synthetic H3t peptides (Ala1–Ala15) containing K9me1, k9me2, and k9me3 were in the order of K9me1 < K9me2 < K9me3 (Fischle et al., 2003).

To quantitatively assess the affinity of *H3t peptides, we next performed SPR where the expressed *H3t peptides were immobilized on a sensor chip that was covalently coupled with an anti-Flag antibody. When 2 μM of chromoHP1 was flowed over the respective *H3t peptides on the chip, we reproducibly observed that the *H3tK9me3 exhibited a resonance unit at equilibrium (ΔReq) value 5–20 times higher than other *H3t peptides, indicating its strong affinity to chromoHP1 (Figure 3C). We then varied the concentrations of chromoHP1 using *H3tK9me3 as a binding partner to obtain the dissociation constant of 0.76 ± 0.05 μM (Figure 3C, inset). This value was slightly lower but similar to the reported values 2.4–4.0 μM using the aforementioned short H3t peptide with the K9me3 modification. Collectively, these findings show that the *H3t peptides prepared by genetic code reprogramming can be used to detect interactions with a known effector protein qualitatively and quantitatively.