heteroatom. dFx very likely recognizes the 3,5-dinitrobenzyl leaving group of acyl-DBE; that is, the substrate presumably binds to the active site of dFx in inverted orientation respect to Phe-EE (Phe-CME) to Fx3. Therefore, it is reasonably assumed that side chain of acyl-DBE is directed away from the active site as similar to the orientation of the ethyl group of ester moiety seen in the crystal structure of Fx3. This results in granting dFx the ability to catalyze aminoacylation toward acyl-DBE with various side chains. The reason why Fx3 does not accept acyl-DBE is that the 3,5-dinitrobenzyl group is too large to be accommodated in the pocket of Fx3. In fact, dFx acquires a single base insertion in J2/1a. A reasonable assumption is that this base insertion along with base mutations enlarges the pocket to accommodate the 3,5-dinitrobenzyl group, though we currently have no clear vision how this alteration of bases could constitute the active cavity.

Structural Basis of Recognition of tRNA Acceptor End

All flexizymes accept virtually any kind of tRNA irrespective of their anticodon and body sequences, since tRNA binding requires the formation of only three base pairs between flexizyme's G55–U57 and tRNA's tC75–tG73 (Figure 7C). These three base pairs, consisting of two Watson–Crick and one G•U wobble pairs, initially predicted by biochemical experiments are also observed in the crystal structure. All tRNAs have the CCA-3′ end without exception, but the base at position 73, namely, discriminator base, potentially has all possible four bases. It has been shown that U57 in Fx3 tolerates pairing with tA73 (Watson–Crick) as well as tU73 (U•U wobble).[9] tRNA with tC73 is the worst substrate for Fx3, but prolonging the incubation time with Fx3 still yields the desired aminoacyl-tRNAs. Alternatively, U57G mutation in Fx3 pairs with tC73 would increase the efficiency like that of the normal Fx3-tRNA pair (unpublished data).

In addition to these three base pairs, A54 was fully conserved in all flexizymes. Although it was unclear if a noncanonical base pair of A54 with tA76 could be formed, our earlier biochemical data showed that tA76 was critical to maintain the full ribozyme activity where mutation of tA76 to C, G, or U decreases its aminoacylation activity by 20%, 15%, and 10%, respectively; and also A54 mutation diminishes aminoacylation activity.[9] The crystal structure of Fx3 has revealed that tA76 makes van der Waals contact with the ribose of G24 and its N1 makes a hydrogen bond with 2′-OH of A23.

Moreover, flexizyme utilizes the J1a/3 hairpin-shaped turn to precisely position the acceptor end of tRNA. Notably, U52–A54 form the J1a/3 turn by the following specific interactions: (1) A54 makes multiple hydrogen bonds with the C50•G21 pair of P1a from the minor groove of the helix, making the so-called type I A-minor motif,[25] (2) U52 stacks on A54 and also packs against G51•C20 of P1a, and (3) U53 is extruded from the turn. This turn structure helps to place G55–U57 perpendicular to the main helical stack in Fx3. The importance of this unique hairpin U-turn structure in activity was supported by the biochemical data where U52 was completely conserved in all flexizymes. Further, comparison of crystal structures of amino acid substrate bound and unbound conformers has revealed that this J1a/3 fully docks with P1a only when amino acid substrate (in this case Phe-EE) is bound. Together with the result from early chemical probing data on r24 where U52–53 were protected from Pb^{2+} cleavage upon binding of amino acid substrate,[9] the precise positioning of the acceptor end of tRNA by J1a/3 seems to be closely coupled with amino acid substrate binding.

An intriguing mechanistic question arisen from our earlier biochemical data that Fx3 exclusively aminoacylates on the 3′-OH group of tA76, not on the 2′-OH group.[10] The crystal structure has revealed that this selectivity is directed by two major interactions. First, the 3′-OH is positioned close to the carbonyl carbon of Phe-EE through the interactions between tA76 and J1a/2, where tA76 interacts with G24 and A23 as described above. Second, in the short helix consisting of G55–G56 and tC75–tC74, the 2′-OH is buried in the minor groove and less accessible than 3′-OH.

Conclusion

We have here comprehensively summarized the evolutionary history of a family of ARS ribozymes, flexizymes. Despite the fact that our attempt toward the generation of such ribozymes began from evolutionary interests of the origin of aminoacylation and translation chemistries, we have ended up developing a useful and practical catalytic system with almost no structural limitation of usable amino/hydroxy acids and tRNA, thus providing a versatile tool for the synthesis of acyl-tRNAs. This system offers us a facile method for the preparation of various mischarged tRNAs that allows for performing a new dimension of biochemical studies on the translation system, such as ribosome[26] and elongation factors. Moreover, acyl-tRNAs prepared by flexizymes could be integrated with a custom-made in vitro translation system, enabling us to reprogram the genetic

code assigning from proteinogenic amino acids to nonproteinogenic amino acids.[12-20,27-33] Thus, this methodology opens new means to express nonstandard peptides and their libraries for the discovery of therapeutic peptides.

Regardless of the potential applications of flexizymes to translation, the present studies on flexizymes imply a possible existence of similar ribozymes in the RNA world. We have shown that in a small ribozyme consisting of only 45 nt, a few mutations and insertions in the active site have successfully altered the specificity toward amino acid substrates with different aromatic leaving groups. This may imply that ribozymes set in the 5′-leader region of tRNAs could have evolved against amino acids with particular signatures of side chains or leaving groups (or their combinations). Because it has been shown that M1 RNA is able to cleave 5′-leader flexizyme away from tRNA after aminoacylation, in the RNA world such collaborations of catalytic precursor tRNAs with M1 RNA-like ribozyme could have established the genetic code currently used in all organisms. Yet, this is a still scientific imagination until we witness such a molecular fossil discovered in nature.

BIOGRAPHICAL INFORMATION

Jumpei Morimoto was born in Takasago City, Japan in 1983. He received his B. Eng. and M. Eng. from the University of Tokyo and entered the doctoral course in the Department of Chemistry and Biotechnology, Graduate School of Engineering in 2009 under the supervision of H. Suga. His current research interests are in the field of biotechnology related to translation and discovery of peptidomimetics against epigenetic targets.

Yuuki Hayashi was born in Tsukushino City, Japan in 1978. After receiving his B. Eng. and M. Eng. from Osaka University, he received his D. Eng. from the Department of Information Science, Graduate School of Engineering at the same University in 2006 under the supervision of Prof. T. Yomo. He has been working as a post-doctoral fellow under the supervision of H. Suga since 2006. His present research interest is to identify novel selective peptide inhibitors against kinases, which are responsible for human diseases such as pancreatic and ovarian carcinomas.

Kazuhiro Iwasaki was born in Osaka City, Japan in 1983. He received his B. Sci. from Tokyo University of Science and M. Eng. from the University of Tokyo. He is in the doctoral course of the Department of Chemistry and Biotechnology, Graduate School of Engineering under the supervision of H. Suga. His present research interest is in vitro selection of nonstandard peptides against cancer markers.

Hiroaki Suga was born in Okayama City, Japan in 1963. He received his B. Eng. and M. Eng. from Okayama University and Ph. D. in Chemistry from the Massachusetts Institute of Technology in 1994 under the supervision of S. Masamune. After three years of post-doctoral work in Massachusetts General Hospital working in the laboratory of J. W. Szostak, he was appointed as a tenure-track Assistant Professor in the Department of Chemistry in the State University of New York at Buffalo to carry his independent research group, and he became tenured Associate Professor in 2002. In 2003, he moved to the Research Center for Advanced Science and Technology at the University of Tokyo as Associate Professor, and soon after was promoted to Full Professor. In 2010, he changed his affiliation to the Department of Chemistry, Graduate School of Science. His research interests are in the field of chemical biology and biotechnology related to RNA, translation, and peptides. He also conducts drug discovery programs in collaboration with industrial partners, including his own biotechnology company, PeptiDream Inc. Tokyo.

We thank Drs. H. Saito and H. Murakami for early contributions to the development of flexizymes, Drs. H. Xiao and A. R. Ferré-D'Amaré for contributions to solving the crystal structure, and all members of the Suga laboratory. This work was supported by grants from the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (S) (16101007), Specially Promoted Research (21000005), and a research and development projects of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization to H.S.; and, in part, Grants-in-Aid for JSPS fellows to J.M. (21-9079) and K.I. (22-9695).

FOOTNOTES

*To whom correspondence should be addressed. Telephone/Fax: +81-3-5841-8372. E-mail: hsuga@chem.s.u-tokyo.ac.jp.

REFERENCES

1 Illangasekare, M.; Sanchez, G.; Nickles, T.; Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **1995**, *267*, 643–647.
2 Illangasekare, M.; Yarus, M. Specific, rapid synthesis of Phe-RNA by RNA. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5470–5475.
3 Jenne, A.; Famulok, M. A novel ribozyme with ester transferase activity. *Chem. Biol.* **1998**, *5*, 23–34.
4 Lohse, P. A.; Szostak, J. W. Ribozyme-catalysed amino-acid transfer reactions. *Nature* **1996**, *381*, 442–444.
5 Lee, N.; Bessho, Y.; Wei, K.; Szostak, J. W.; Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nat. Struct. Biol.* **2000**, *7*, 28–33.
6 Lee, N.; Suga, H. A minihelix-loop RNA acts as a trans-aminoacylation catalyst. *RNA* **2001**, *7*, 1043–1051.
7 Bessho, Y.; Hodgson, D. R.W.; Suga, H. A tRNA aminoacylation system for non-natural amino acids based on a programmable ribozyme. *Nat. Biotechnol.* **2002**, *20*, 723–728.
8 Saito, H.; Kourouklis, D.; Suga, H. An in vitro evolved precursor tRNA with aminoacylation activity. *EMBO J.* **2001**, *20*, 1797–1806.
9 Saito, H.; Watanabe, K.; Suga, H. Concurrent molecular recognition of the amino acid and tRNA by a ribozyme. *RNA* **2001**, *7*, 1867–1878.
10 Saito, H.; Suga, H. A ribozyme exclusively aminoacylates the 3′-hydroxy group of the tRNA terminal adenosine. *J. Am. Chem. Soc.* **2001**, *123*, 7178–7179.
11 Murakami, H.; Saito, H.; Suga, H. A versatile tRNA aminoacylation catalyst based on RNA. *Chem. Biol.* **2003**, *10*, 655–662.
12 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.
13 Goto, Y.; Murakami, H.; Suga, H. Initiating translation with D-amino acids. *RNA* **2008**, *14*, 1390–1398.
14 Goto, Y.; Ohta, A.; Sako, Y.; Yamagishi, Y.; Murakami, H.; Suga, H. Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol.* **2008**, *3*, 120–129.

15 Kawakami, T.; Murakami, H.; Suga, H. Messenger RNA-programmed incorporation of multiple N-methyl-amino acids into linear and cyclic peptides. *Chem. Biol.* **2008**, *15*, 32–42.

16 Kawakami, T.; Murakami, H.; Suga, H. Ribosomal synthesis of polypeptoids and peptoid–peptide hybrids. *J. Am. Chem. Soc.* **2008**, *130*, 16861–16863.

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

18 Niwa, N.; Yamagishi, Y.; Murakami, H.; Suga, H. A flexizyme that selectively charges amino acids activated by a water-friendly leaving group. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3892–3894.

19 Goto, Y.; Suga, H. Translation initiation with initiator tRNA charged with exotic peptides. *J. Am. Chem. Soc.* **2009**, *131*, 5040–5041.

20 Ohshiro, Y.; Nakajima, E.; Goto, Y.; Fuse, S.; Takahashi, T.; Doi, T.; Suga, H. Ribosomal synthesis of backbone-macrocyclic peptides containing γ-amino acids. *ChemBioChem* **2011**, *12*, 1183–1187.

21 Goto, Y.; Katoh, T.; Suga, H. Flexizymes for genetic code reprogramming. *Nat. Protoc.* **2011**, *6*, 779–790.

22 Xiao, H.; Murakami, H.; Suga, H.; Ferré-D'Amaré, A. R. Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature* **2008**, *454*, 358–361.

23 Ferré-D'Amaré, A. R. Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. *Methods* **2010**, *52*, 159–167.

24 Saito, H.; Suga, H. Outersphere and innersphere coordinated metal ions in an aminoacyl-tRNA synthetase ribozyme. *Nucleic Acids Res.* **2002**, *30*, 5151–5159.

25 Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4899–4903.

26 Effraim, P. R.; Wang, J.; Englander, M. T.; Avins, J.; Leyh, T. S.; Gonzalez, R. L.; Cornish, V. W. Natural amino acids do not require their native tRNAs for efficient selection by the ribosome. *Nat. Chem. Biol.* **2009**, *5*, 947–953.

27 Nakajima, E.; Goto, Y.; Sako, Y.; Murakami, H.; Suga, H. Ribosomal synthesis of peptides with C-terminal lactams, thiolactones, and alkylamides. *ChemBioChem* **2009**, *10*, 1186–1192.

28 Sako, Y.; Goto, Y.; Murakami, H.; Suga, H. Ribosomal synthesis of peptidase-resistant peptides closed by a nonreducible inter-side-chain bond. *ACS Chem. Biol.* **2008**, *3*, 241–249.

29 Sako, Y.; Morimoto, J.; Murakami, H.; Suga, H. Ribosomal synthesis of bicyclic peptides via two orthogonal inter-side-chain reactions. *J. Am. Chem. Soc.* **2008**, *130*, 7232–7234.

30 Yamagishi, Y.; Ashigai, H.; Goto, Y.; Murakami, H.; Suga, H. Ribosomal synthesis of cyclic peptides with a fluorogenic oxidative coupling reaction. *ChemBioChem* **2009**, *10*, 1469–1472.

31 Kang, T. J.; Yuzawa, S.; Suga, H. Expression of histone H3 tails with combinatorial lysine modifications under the reprogrammed genetic code for the investigation on epigenetic markers. *Chem. Biol.* **2008**, *15*, 1166–1174.

32 Kawakami, T.; Ohta, A.; Ohuchi, M.; Ashigai, H.; Murakami, H.; Suga, H. Diverse backbone-cyclized peptides via codon reprogramming. *Nat. Chem. Biol.* **2009**, *5*, 888–890.

33 Goto, Y.; Iwasaki, K.; Torikai, K.; Murakami, H.; Suga, H. Ribosomall synthesis of dehydrobutyrine- and methyllanthionine-containg peptides. *Chem. Commun.* **2009**, *23*, 3419–3421.

ChemComm

Cite this: *Chem. Commun.*, 2011, **47**, 9946–9958

www.rsc.org/chemcomm

HIGHLIGHT

Ribosomal synthesis of backbone macrocyclic peptides

Takayuki Katoh,[a] Yuki Goto,[a] Md. Shamim Reza[b] and Hiroaki Suga*[ab]

DOI: 10.1039/c1cc12647d

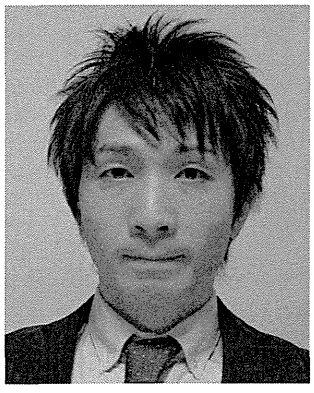
A wealth of knowledge has been accumulated on ribosomal synthesis of macrocyclic peptides in the past decade. In nature, backbone cyclization of the translated linear peptides is generally catalyzed by specific enzymes, giving them peptidase resistance, thermodynamic stability and various other physiological activities. Due to these biochemical traits, backbone cyclic peptides have become an attractive resource for the discovery of drug leads. Recently, various new methodologies have also been established to generate man-made cyclic peptides. Here, we describe the biosynthetic mechanisms of naturally occurring backbone macrocyclic peptides focusing on cyclotides, sunflower trypsin inhibitors (SFTIs) and cyanobactins as well as several new emerging methodologies, such as sortase mediated ligation, protein splicing method and genetic code reprogramming.

[a] *Department of Chemistry, Graduate School of Science, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: hsuga@chem.s.u-tokyo.ac.jp; Fax: +81-3-5841-8372; Tel: +81-3-5841-8372*

[b] *Department of Advanced Interdisciplinary Studies, Graduate School of Engineering, The University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8904, Japan*

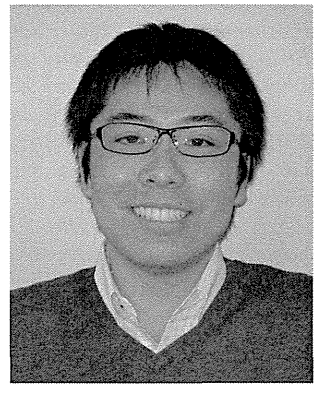
Introduction

The translation system plays a central role in synthesizing polypeptides (proteins) in all organisms on the earth. The ribosome, that is the catalytic machinery of the translation system, decodes the sequence information stored in mRNA (messenger RNA) template based on the genetic code and accurately polymerizes amino acids. The genetic code consists of the combinations of trinucleotides, referred to as codons, and each codon assigns one of 20 proteinogenic amino acids or termination of translation (we here use the word "proteinogenic" rather than "natural" for the amino acids assigned in the genetic code, since many



Takayuki Katoh

Takayuki Katoh performed his doctoral research at the University of Tokyo under the mentorship of Prof. K. Watanabe and Prof. T. Suzuki. He studied biogenesis of mammalian microRNAs and mechanism of RNA interference. After spending five years as a postdoctoral fellow at the University of Tokyo and Japan Biological Informatics Consortium (JBIC), he was appointed as an assistant professor at the Department of Chemistry, Graduate School of Science, the University of Tokyo in 2009. His research is now focused on the development of cyclic peptide drugs utilizing the genetic code reprogramming method.



Yuki Goto

Yuki Goto obtained his MS in bioorganic chemistry under the guidance of Prof. K. Nakatani and Prof. I. Saito at Kyoto University (Japan) in 2005. He then joined the group of Prof. H. Suga at the University of Tokyo (Japan) and received his PhD in the field of chemical biology in 2008. After a postdoctoral stay in the University of Illinois at Urbana-Champaign (USA) with Prof. W. A. van der Donk focusing on biosynthesis of lantibiotics, he was appointed as an assistant professor at the University of Tokyo in 2009. His research interest is engineering of the biological synthesis system to develop artificial biosynthetic machineries and to obtain novel functional molecules.

non-proteinogenic amino acids, such as *N*-methyl amino acids discussed in this review, are also naturally occurring amino acids). This decoding system of mRNA provides unique and robust machinery for the template-dependent synthesis of long polypeptides (could be more than 1000 residues) achieved with high fidelity and purity.

Although the translation system is often used to express proteins because it has an obvious advantage for the synthesis of longer polypeptides over chemical synthesis, it is also applicable to express shorter polypeptides (here simply referred to as peptides) in lengths with less than 50 residues. In such a case, the major advantage to use the translation system over chemical synthesis is the facility of preparation of peptide libraries with high diversities ranging from 10^7 to 10^{13}. Since peptides expressed by the translation system are mRNA template-directed, their sequences are encoded in cognate mRNA sequences. Generally, mRNAs can be prepared from the DNA templates by transcription, and custom synthesis of such DNAs is commercially available inexpensively. For library synthesis in the translating region of mRNAs, the initiation codon followed by random codon NNK (N = G, A, T, or C; K is T or G) is repeated in designated times, and then one of the termination codons is installed. When a certain strategy to link between genotype mRNA templates and phenotype peptides is implemented into the screening processes, amplification of the genotypes of active phenotypes could be achieved by an appropriate method, such as RT-PCR (reverse transcription-polymerase chain reaction) of the selected genomes. Thus, it is possible for genetic encoding peptide expression to not only generate such libraries with a high quality but also deconvolute the sequences of interest identified by screening if there were an extremely low abundance or quantity of active peptides, as low as a few copies, in the initial library.

Although short peptides consisting of proteinogenic amino acids are not stable under physiological environments, *e.g.* in human plasma, due to their susceptibility to peptidase/proteases, nature installs unique chemical features into not only side chains but also the peptide backbone to improve the proteolytic stability of peptides. One such feature is the macrocyclic structure composed of a covalent linkage between the amino group at the N-terminus and the carboxylate at the C-terminus, the N-terminal amino group and a side chain of an internal residue, or a side chain of an internal residue and the C-terminal carboxylate. These macrocyclic structures increase not only the physiological (serum) stability of peptides but also affinity to binding partners derived from the conformational rigidity.

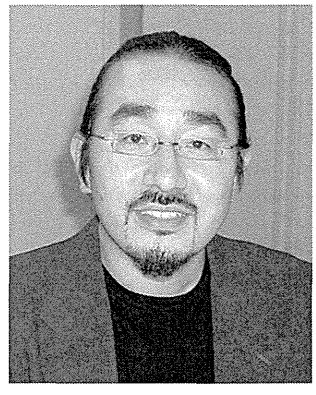
While nature uses diverse kinds of chemical strategies for macrocyclization, here we will focus our discussions on "backbone macrocyclization". The backbone macrocyclization seen in naturally occurring peptides is generally employed by specific enzymes that modify the precursor linear peptides to ligate the two ends by certain mechanisms. The non-charged termini in such peptides protect from processing by cellular exopeptidases, granting them physiological stability *in vivo*. More recently, new versatile methodologies have been developed to obtain "non-naturally occurring" backbone macrocyclic peptides in conjunction with their library construction and screening for desired biologically active species. This review also comprehensively summarizes such methodologies.

Naturally occurring backbone cyclic peptides

To date, various kinds of backbone cyclized peptides have been identified in nature from various organisms including both prokaryotes and eukaryotes.[1] In the case of the peptides produced by ribosomal synthesis, the genes coding cyclic peptides are first transcribed and translated into precursor linear peptides, typically having some accessory sequences at both N- and C-termini. These sequences usually function as a recognition motif(s) for a specific enzyme(s) involved in a process of peptide maturation, and are cleaved off from the mature peptides by a specific protease(s).[2] At this cleavage step, some families of protease can also catalyze cyclization of cognate peptides. For instance, in the case of cyanobactin peptides, a specific serine-protease catalyzes the cleavage of the C-terminal recognition sequence and simultaneously macrocyclizes cognate peptide between the

Md. Shamim Reza

Md. Shamim Reza studied Biochemistry and Molecular Biology at the University of Dhaka, Bangladesh, where he received his MS degree in 2005. He is currently performing his doctoral studies at the University of Tokyo under the supervision of Prof. Hiroaki Suga. His research is focused on the development of novel peptide therapeutics. A key aspect of his research is the implementation of genetic code reprogramming technology for achieving cell permeable peptides.



Hiroaki Suga

Professor, Department of Chemistry, Graduate School of Science, The University of Tokyo. He received his BEng and MEng from Okayama University, PhD in Chemistry from the Massachusetts Institute of Technology in 1994. After his post-doctoral work in Massachusetts General Hospital, he was appointed as a tenure-track Assistant Professor in the Department of Chemistry in the State University of New York at Buffalo in 1997, becoming Associate Professor in 2002. In 2003, he moved to the Research Center for Advanced Science and Technology in the University of Tokyo, and moved to the present position in 2010.

N-terminus and newly formed C-terminus.[3] It has been reported that macrocyclization of some bacterial peptides, such as microcins and microviridins, is also cyclized by a similar family of proteases, but the macrocyclization does not take place at the N- and C-termini; instead, the cyclization occurs between the N- or C-terminus and a side chain group at an internal residue,[4,5] suggesting that the catalytic mechanism could be similar but not exactly the same as the head-to-tail cyclic peptides. Below, we will discuss biosynthesis of some representative macrocyclic peptides.

Backbone cyclization of cyclotides and other related peptides

Cyclotides[6] and sunflower trypsin inhibitors (SFTIs)[7] produced by plants are a family of backbone cyclic peptides. A 29-mer compact peptide, known as Kalata B1, is the most well characterized cyclotide, which is originally isolated from African plant *Oldenlandia affinis* (Fig. 1). This product has been historically known as an ingredient of a natural drug, kalata-kalata, used for accelerating contractions and childbirth. One of the remarkable features of this peptide is high thermostability, where the activity is retained even after boiling. This property is derived from its compactly folded structure, composed of a triple stranded β-sheet structure and six cysteine residues, interlocking three disulfide bonds in addition to the head-to-tail cyclization.

Precursors of the cyclotides generally consist of an endoplasmic reticulum (ER) signal sequence, an N-terminal propeptide (NTPP), a mature cyclotide domain and a C-terminal propeptide (CTPP). Excision of the mature cyclotide domain and *de novo* ligation of the newly formed N- and C-termini occur during the maturation, but yet precise mechanisms of the cyclization are not fully understood. Due to the highly conserved asparagine residue located at the cleavage sites, it is likely that asparaginyl endopeptidases (AEP) are utilized in the maturation of cyclotides.[8] In fact, the AEP was shown to have both peptidase and ligase activity *in vivo*, so that both reactions can be catalyzed by a single enzyme. Moreover, Gillon *et al.* recently reported that the mature cyclic peptide could not be produced when the conserved asparagine was replaced by alanine or aspartate residues in the transgenic *Arabidopsis thaliana* and *Nicotiana tabacum* that carry cyclotide genes.[9] It has been recently suggested by Mylne *et al.* that SFTI peptides are adopting a similar cyclization mechanism by means of the AEP; however, the precise chemical reaction mechanism still remains to be solved.[10] Therefore, more biochemical as well as genomic studies of these peptide families would be awaited for revealing the exact mechanism of processing actions.

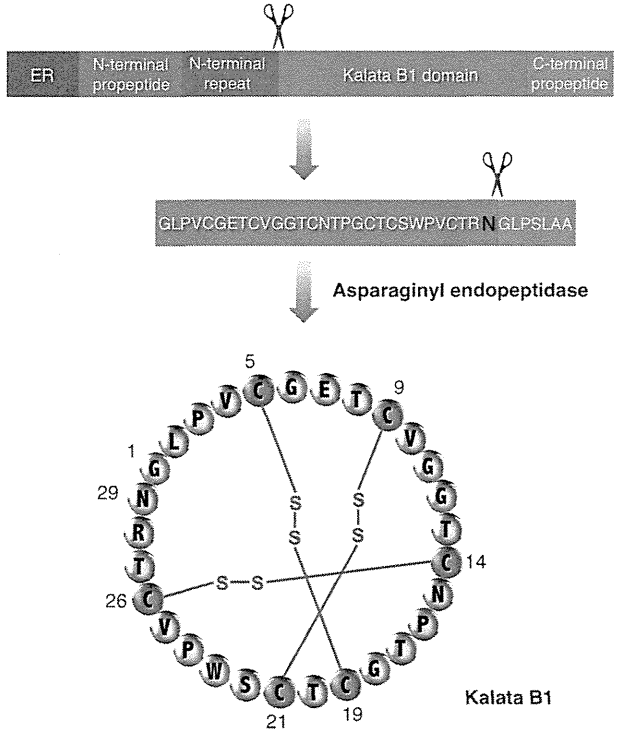


Fig. 1 A proposed mechanism for the cyclization of Kalata B1. The N-terminal flanking sequences including ER signal, N-terminal propeptide and N-terminal repeat are first excised from the precursor by as-yet unidentified enzyme(s). Then, the cleavage of the C-terminal propeptide and macrocyclization of the mature peptide are catalyzed by a single asparaginyl endopeptidase (AEP).

Sortase-mediated peptide cyclization

Sortase-mediated ligation is an emerging technique that enables us to perform ligation of two distinct peptides/proteins using a transpeptidase sortase A (SrtA) isolated from a Gram-positive *Staphylococcus aureus*.[11] The native role of SrtA *in vivo* is to ligate cell-surface proteins having a LPXTG motif and cell wall peptidoglycans. Cys184 in SrtA acts as a nucleophile for the cleavage of a T–G peptide bond in the LPXTG motif and forms a protein–LPXT–SrtA thioester intermediate. This intermediate further reacts with the amino group of N-terminal glycine oligomers (oligo-Gs) on the cell wall peptidoglycans, producing the ligated product of protein–LPXT–(oligo-Gs)–peptidoglycans. While the LPXTG motif in the acceptor protein is strictly recognized by SrtA, oligo-Gs in the donor peptide is not absolutely required. In an extreme case, the donor peptide having only a single G residue at the N-terminus can be a substrate, although such a donor peptide is less efficient compared with that with multiple G residues.[12,13] Therefore, since users are able to design any acceptor protein simply having the LPXTG motif, the SrtA-mediated ligation technique is applicable to various donor proteins and acceptor peptides. In a technical side, SrtA lacking the N-terminal domain has given users better handling properties due to increased water-solubility and stability and therefore is often used for *in vitro* experiments.[14]

The peptide ligation capability of SrtA also can be applied for intramolecular ligation of polypeptides, *i.e.* macrocyclization of proteins. It has been

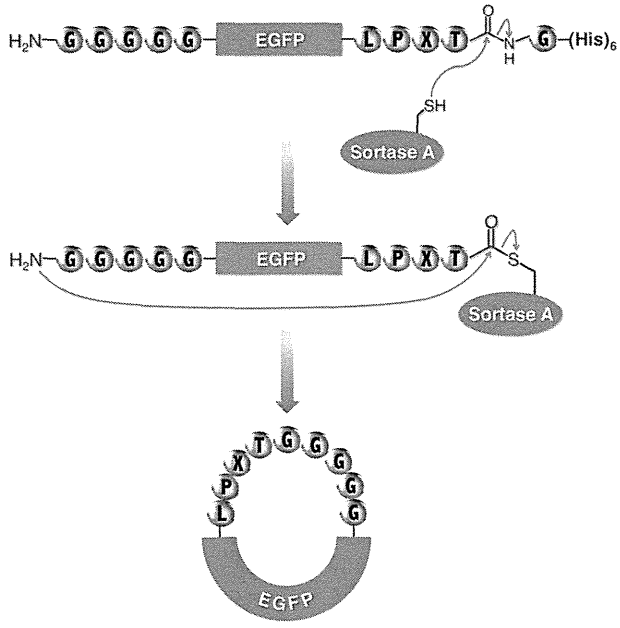


Fig. 2 Macrocyclization of polypeptides (proteins) by utilizing the Sortase-mediated ligation method. EGFP having a N-terminal oligoglycine tag and C-terminal LPXTG tag can be cyclized by a single transpeptidase Sortase A (SrtA).

demonstrated that EGFP (enhanced-green fluorescent protein) having a penta-G motif at the N-terminus and a LPETG motif at the C-terminus could be a good substrate of SrtA and both the ends were successfully ligated to yield cyclic EGFP[14,15] (Fig. 2). On the other hand, this approach requires rather long motifs of oligo-Gs for efficient ligation, and therefore limits the technical application to the expression of longer macrocyclic proteins rather than shorter macrocyclic peptides.

Biosynthesis of patellamide, a cyanobactin peptide

Patellamide found in marine ascidians[16] turned out to be a product of a symbiont cyanobacteria, *Prochloron didemni*, containing a set of genes necessary for patellamide biosynthesis.[17] A cyanobactin peptide family consists of more than 100 various backbone-cyclized peptides including patellamide, trunkamide,[18] microcyclamide,[19] aerucyclamide,[20] nostocyclamide,[21] tenuecyclamide,[22] venturamide,[23] dendroamide[24] and others, and thus cyanobactins represent an attractive source of naturally occurring macrocyclic peptides. The *pat* gene cluster encodes proteins or peptides from PatA to PatG, among which *patA*, *D*, *E*, *F* and *G* are the essential genes for the patellamide biosynthesis, whereas *patB* and *C* are dispensable[17] (Fig. 3). The *patE* gene, on the other hand, encodes the pre-peptide that contains a 37-mer N-terminal leader sequence and two core peptides, named cassette I (VTACITFC) and II (ITVCISVC), both of which are flanked by N- and C-terminal recognition sequences consisting of G(L/V)E(A/P)S and AYDG(E), respectively (Fig. 3). A serine protease PatA plays a role in cleavage of the pre-peptide at the N-terminus of cassettes I and II. The other subtilisin-like serine protease PatG removes C-terminal recognition sequences of both cassettes I and II, and also catalyzes macrocyclization. A proposed chemical reaction mechanism of the patellamide macrocyclization is that a serine residue in the active site of PatG and the carboxyl group of the C-terminal amino acid in the core peptide form a covalent acyl–enzyme intermediate, followed by nucleophilic attack of the N-terminal amino group to the activated acyl moiety in the intermediate.[3] Subsequently, the cyclic peptide is released from the enzyme, and the cassettes I and II are thus converted into mature patellamides C and A, respectively.

Patellamide and other related cyanobactin families consist of not only a macrocyclic structure but also heterocycles such as thiazole and oxazoline, which are derived from Cys, Ser and Thr (C/S/T). These heterocycles likely contribute to expanding the chemical space and acquiring various biological activities, exemplified by cytotoxic, antibacterial, anti-malarial and anti-tumor activities of various cyanobactins.[25] In patellamide biosynthesis, the heterocyclase domain of PatD catalyzes thiazoline and oxazoline formation from C/S/T residues, and then the oxidase domain of PatG catalyzes oxidation of the thiazoline into thiazole; thus PatG plays a dual role in the biosynthesis.[26,27]

Despite the fact that the core peptide sequences in cyanobactins are variable, the enzymes and their recognition motifs are fully conserved.[28] This indicates that the core peptide region of PatE can be genetically engineered and used as a platform for constructing a cyclic peptide library by randomizing the sequence. Although cyanobactin-producing cyanobacteria are not culturable as yet, a recent study showed that the patellamide biosynthetic genes could be transferred to *Escherichia coli* and the mature cyclic peptides were successfully produced.[17,26,28] Macrocyclization of core peptides by PatG can be also reconstituted *in vitro*.[27] The minimal structure of the substrate peptide is the core sequence along with a C-terminal hexamer or pentamer recognition motif, AYDG(E).[17] Schmidt and co-workers have recently demonstrated *in vitro* processing activity of PatG for various synthetic core peptides of which tolerable length was in a range from 6 to 11 amino acids while their C-terminus was fixed with proline.[27] Interestingly, peptides containing alkyl spacers such as aminohexanoic or aminoheptanoic acid in the core region could also serve as substrates for cyclization by PatG. Likewise, D-amino acids are also tolerated when located at some central positions of the core peptide. When a nucleophilic group was present at the side chain at the N-terminus, for instance Lys, cyclization *via* the side chain could be observed. Taken all together, these results suggest engineering potentials of the PatG system and substrate peptides that may allow researchers to perform the syntheses of various sequences of cyclic peptides. Yet, our present knowledge of this catalytic system implies a limitation that the producible size of cyclic peptides

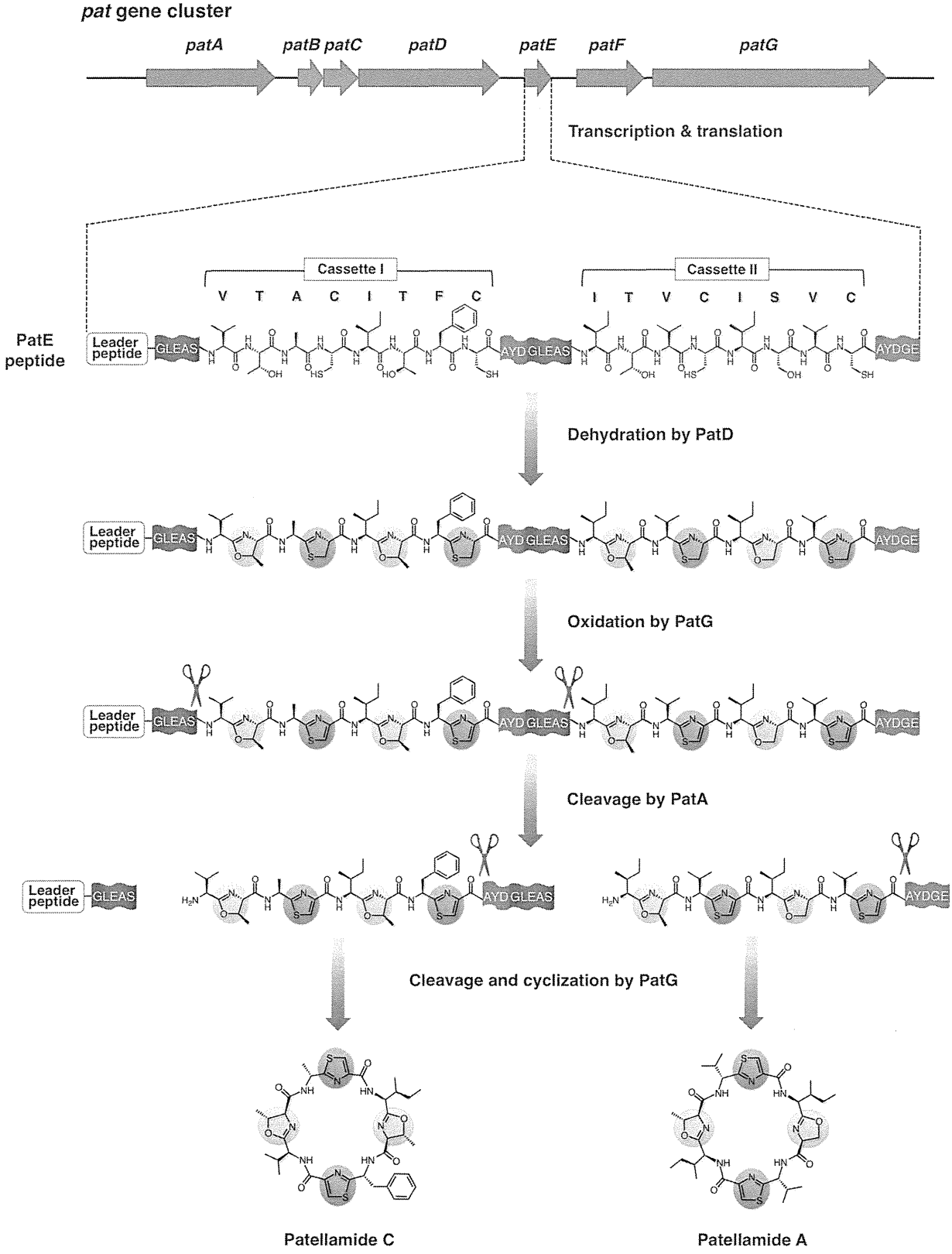


Fig. 3 A model for the biosynthesis of patellamides C and A. Genes indicated by the red arrows are essential for patellamide biosynthesis, whereas the blue arrows indicate unnecessary genes. The precursor peptide PatE contains two core peptide regions (cassettes I and II), both of which have flanking N- and C-terminal recognition motifs. PatD and PatG catalyze dehydration and oxidation of C/S/T residues, respectively, to yield thiazole (purple circles) and oxazoline (green circles) structures. PatA cleaves off the N-terminal recognition motifs of both cassettes I and II. PatG catalyzes the cleavage of the C-terminal recognition motifs and macrocyclization of the peptides simultaneously.

probably may fall in a relatively narrow range (6–11-mers).

Protein splicing methods

Protein splicing is a posttranslational processing of proteins, in which an internal protein sequence is removed from a protein precursor and its N- and C-terminal flanking fragments, termed exteins, are ligated to each other. The internal part of protein, called intein,[29] indeed possesses the catalytic activity to carry out the protein splicing, *i.e.* cleavage of two amide bonds and formation of a new amide bond[30] (Fig. 4a). Protein splicing by inteins is initiated by (thio)ester formation *via* N → S/O acyl shift at the conserved C/S/T residue located at the first position of the intein domain. The intein then catalyzes trans(thio)-esterification between the (thio)ester and the side chain of C/S/T present in the first position of C-terminal extein, resulting in transfer of the N-terminal extein to the side chain of the C-terminal extein residue. The last catalytic step by intein involving the cyclization of C-terminal asparagine of intein domain leads to the cleavage of the amide bond between intein and C-terminal extein and generates free amino group in the C/S/T residue. The following spontaneous S/O-N acyl rearrangement yields a new amide bond connecting the two exteins.

Because of the unique catalytic ability, intein chemistry has been widely applied as a biotechnological tool in the protein engineering field, including protein semi-synthesis *via* expressed protein ligation,[31,32] site-specific segmental isotopic labeling of proteins enabling for their NMR studies,[33] and the removal of affinity or hydrophilic tags from recombinant proteins.[34,35]

Methods using C-terminal intein to generate a thioester group

The ability of intein, which specifically conducts cleavage and formation of backbone-amide bonds, has been also utilized for the synthesis of small backbone cyclic peptides. A simple way to produce macrocyclic peptides using intein chemistry involved the method in which an intein domain is arranged at the C-terminal region of the objective precursor peptide and an N-terminal Cys bearing a free amino group is generated by an appropriate sequence-specific peptidase, such as methionine aminopeptidase and factor Xa[36,37] (Fig. 4b). N → S acyl shift induced by the C-terminal intein triggers macrocyclization *via* thioesterification with the sulfhydryl group of the N-terminal Cys, and then the subsequent S → N acyl rearrangement yields backbone-cyclic peptides. A similar approach utilizing two inteins, called the TWIN system, was also reported.[38] In this system, a target peptide sequence is located between the N- and C-terminal inteins, which are responsible for the generation of N-terminal free Cys amine and C-terminal thioester, respectively (Fig. 4c).

Since a Cys reside is required at the N-terminus of a linear precursor polypeptide, the method is suitable for the synthesis of cyclic peptides containing Cys residue(s). Camarero *et al.* ingeniously employed this method to demonstrate the production of cyclotides, aforementioned backbone cyclic peptide family, in *E. coli*.[39,40] Given the fact that cyclotides exhibit a wide variety of bioactivities such as insecticidal, anti-tumor, hemolytic, trypsin inhibition, and anti-HIV,[41–43] endogenous expression of cyclotide analogs in bacteria using the intein-based methodology potentially opens a new avenue for development of artificial cyclotides with novel biological activities.

Methods using split-intein (SICLOPPS)

SICLOPPS (split-intein circular ligation of peptides and proteins) was devised in order to effectively produce macrocyclic peptides using naturally occurring split intein (DnaE)[44] discovered from *Synechocystis* sp.[45,46] In the SICLOPPS constructs, the C-terminal (In^C) and N-terminal (In^N) domains of DnaE are embedded at the N-terminus and C-terminus of the core peptide sequence, respectively. After translation of a gene encoding the entire In^c–core-peptide–In^N, association of the In^C and In^N regions reconstitutes the *trans*-activity of DnaE intein, resulting in cleavages and ligation of the N-terminus and C-terminus of the core peptide upon splicing of In^c–In^N, yielding a macrocyclic peptide (Fig. 4d).

It has been applied to the expression of small bioactive cyclic peptides. The research team of Benkovic who devised SICLOPPS as well as other teams applied this technology to the expression of genetically encoded cyclic peptide libraries where random sequences were embedded in the core peptide region.[47,48] In this strategy, ribosomal synthesis of the precursor In^c–random-peptide–In^N and subsequent split-intein mediated backbone cyclization were executed to produce macrocyclic peptides in cells, such as bacteria, yeast, and human cells. Cyclic peptides having specific biological activities could be selected from such a SICLOPPS macrocyclic peptide library with which an appropriate cellular reporter system was integrated. For instance, a reverse two-hybrid system (RTHS) was utilized to select cyclic peptide inhibitors against hetero-dimerization of ribonucleotide reductase (RR) subunits.[49] The RTHS was constructed using two chimeric repressor fusions, which consists of a RR subunit fused with a 434 repressor and a RR subunit fused with a p22 repressor; and they were coupled with reporter genes essential for cell growth in the downstream region of the chimeric 434–p22 promoter sequence. In this system, when a heterodimer of the repressor fusions bound to the promoter sequence the expression of reporters was repressed, resulting in no cell growth. When RTHS bacteria cells were transformed by the SICLOPPS peptide library plasmids, the corresponding macrocyclic peptides were expressed in the cells. If peptides could inhibit the heterodimerization of RR subunits, the repression of reporter genes was disabled, thus rescuing the growth of host cells. SICLOPPS plasmids were then harvested from the survived cells and cloned for DNA sequencing to successfully determine sequences of active species. This genetic selection approach has been applied for the isolation of several backbone cyclic peptides inhibiting a specific protein–protein interaction, such as homodimerization of aminoimidazole-4-carboxamide ribonucleotide transformylase[50] or HIV Gag protein with TSG101 (tumor susceptibility gene 101).[51] Moreover, alternative selection or screening methods have been also developed.[52–54]

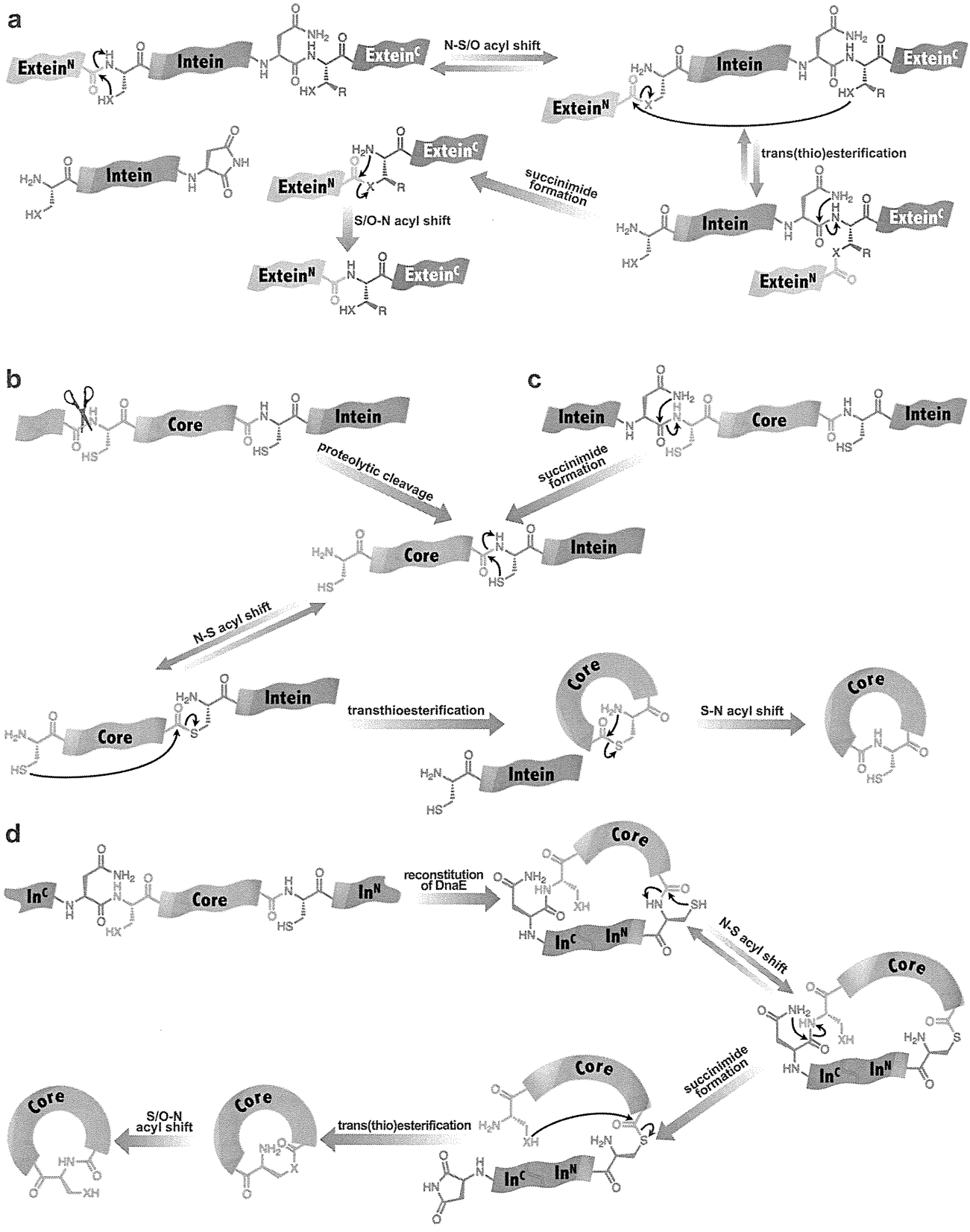


Fig. 4 Mechanism of intein-mediated protein splicing and its applications for the synthesis of backbone cyclic peptides. (a) Protein splicing by inteins. X and R represent S/O and H/Me, respectively, in the side chain of serine, threonine, or cysteine. (b, c) Intein-based macrocyclization of peptides *via* formation of a C-terminal thioester and *trans*-thioesterification involving a N-terminal cysteine residue. The C-terminal thioester is produced by the activity of an intein present at the C-terminal region. The N-terminal free cysteine residue is generated by (b) an appropriate sequence-specific peptidase; or (c) another intein domain at the N-terminal region. (d) Synthesis of backbone cyclic peptides using split-intein in SICLOPPS technology.

Cell-based selection or screening of the cyclic peptide libraries generated by SICLOPPS or other analogous methods has obvious advantages over the methods in screening of chemically synthesized peptide libraries or panning of expressed peptide libraries using a phage display. The most significant advantage is that since both target protein(s) as well as cyclic peptide library were encoded in the DNA plasmids, sophisticated molecular biology techniques are applicable to design selection/screening systems. Moreover, the entire system is compartmentalized in cells, allowing us to carry out high throughput genetic selection or functional screening rather than binding selection or the individual functional screening. On the other hand, there are some shortcomings. Because intein requires for having a C/S/T residue (generally C is preferred over T and S) at the N-terminal position of core peptide, these residues must be designed into the library sequence. More serious limitation would come from the C-terminal position of the core peptide; it favors a residue with a sterically non-demanding side chain for the efficient ligation. This fact reflects to the observation in the SICLOPPS peptide library that an approximately 30% of the library members could not be cyclized.[47,54] The diversity of peptide libraries is also limited by the transformation efficiency of the random SCILOPPS plasmids, generally in a range of 10^7. Moreover, genetic selection strategies might lead to false positives, and therefore an appropriate validation whether the isolated cyclic peptides are the true origin of active species, *i.e.* other alternative mechanisms contribute to the outcome of the genetic selection, is necessary to confirm the activity.[52] Regardless of these shortcomings, it is no doubt that SICLOPPS or others is a very powerful method that allows us to express backbone macrocyclic peptides in cells and screen active species by the integration with appropriate in-cell genetic selection or functional screening.

Genetic code reprogramming methods

Genetic code reprogramming is an emerging new method that enables us to ribosomally express peptides containing multiple non-proteinogenic amino acids. Among the available methods of genetic code reprogramming, the most flexible and preeminence methodology involves two sophisticated catalytic systems, flexizyme[55–57] and custom-made reconstituted *in vitro* translation systems.[57,58] Flexizymes are *de novo* tRNA acylation ribozymes capable of charging virtually any amino acids onto desired tRNAs with any body and anticodon sequences, and thus facilitate the preparation of desired tRNAs charged with non-proteinogenic amino (and hydroxy) acids.[59–65] The aforementioned translation system is composed of purified ribosome, essential recombinant translation factors, tRNAs, ribonucleotides triphosphate, amino acids, and other small organic and inorganic molecules required for translation.[57,58] When certain amino acids and cognate aminoacyl-tRNA synthetases are not included to reconstitute the system, these "aminoacyl-tRNAs" are unavailable for translation, *i.e.* the corresponding codons become vacant. These vacant codons can be filled with non-proteinogenic amino acids by the addition of corresponding aminoacyl-tRNAs prepared by flexizymes. Thus, the integration of these two systems, referred to as the FIT (Flexible *In vitro* Translation[57]) system, facilitates expression of non-standard peptides from designed mRNA templates according to the newly designated genetic table.

Thioester formation followed by macrocyclization of peptides by means of genetic code reprogramming

It has been well established that α-hydroxy acids could be incorporated into the peptide backbone by a classically nonsense suppression method[66,67] and more recently a genetic code reprogramming method.[68] It has been shown that the incorporation of an α-hydroxy acid in the peptide backbone, yielding an ester bond at a specific site, is useful for mapping a role of hydrogen bonding in the peptide bond of interest[67,69,70] or altering the susceptibility against alkaline hydrolysis[71] However, to the best of our knowledge, no chemistry aggressively using the ester bond generated by the ribosomal incorporation of an α-hydroxy acid had been developed until 2009. In 2007, Kawakami and Aimoto reported an ingenious synthetic methodology in which a cysteinyl-prolinyl (C-P) ester moiety was embedded at the C-terminus of chemically prepared peptides on a synthetic resin and rearranged into the corresponding peptides with a diketopiperazine (dkp) thioester at the C-terminus.[72] Stimulated by this work, we have developed a method that expresses C-P-^{HO}G (glycolic acid) in the peptide C-terminus by reprogramming ^{HO}G into a codon using the FIT system, thus generating a peptide with the C-terminal dkp thioester.[73] Clearly, the process is driven by the conjugation of C and P *via* the non-enzymatic nitrogen → sulfhydryl (N → S) equilibrium shift and the spontaneous cleavage of the ^{HO}G residue (Fig. 5), like Aimoto's method.

To apply this method for the ribosomal synthesis of macrocyclic peptides, our FIT system reported in 2009 was customized to include two additional recombinant enzymes, peptide deformylase (PDF) and methionine aminopeptidase (MAP) (note that in the original report,[73] we referred this particular translation system to as the *bc*PURE system; but because we have constructed a more customized system we have decided to rename the system as the "flexizyme-integrated" flexible *in vitro* translation, FIT, system). Since the translation reaction is generally initiated with formylated methionine (f Met), expressed peptides have f Met at the N-terminus. PDF cleaves the formyl group on Met and MAP cleaves the resulting NH_2-Met residue to generate an N-terminal free amino group on the second residue *in situ* of the translation system (Fig. 5a). The efficiency of MAP cleavage depends on the kinds of the second residue, and it was found (also consistent with the previous knowledge[74–77]) that 7 proteinogenic amino acids (Gly, Ala, Pro, Ser, Thr, Val and Cys) are well compatible to MAP.[73] In addition to this collection, we have used the initiation reprogramming approach to assign another 7 proteinogenic amino acids (Met, Phe, Leu, Tyr, Ile, Gln and Trp) at the first position[65] and remove the formyl group by PDF to generate the N-terminal free amino group. All together, a total of 14 amino acids could be utilized as the first residue.[73] With this technique in hand, peptide sequences with a free amino

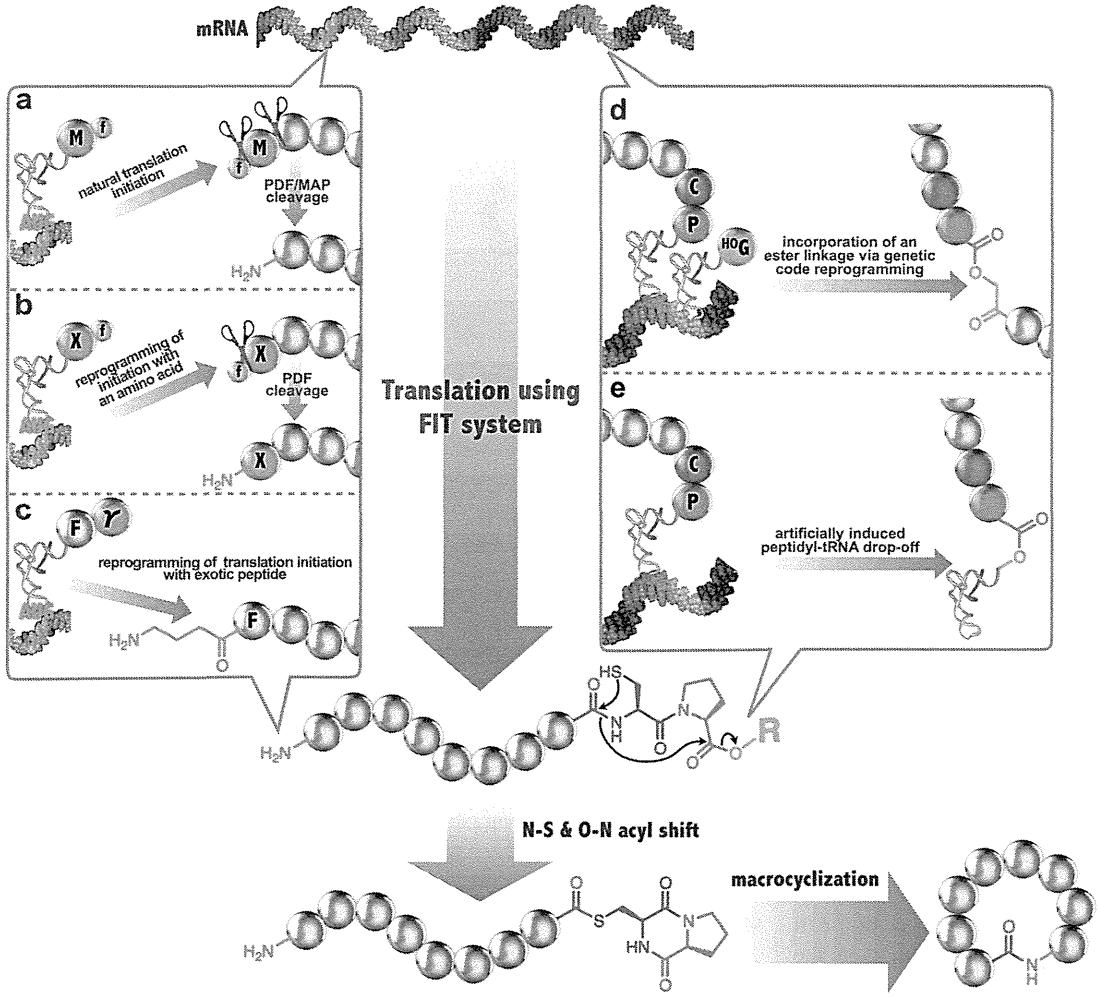


Fig. 5 Ribosomal synthesis of backbone cyclic peptides using the FIT system in which the genetic code is artificially reprogrammed. (a, b, c) Methods for the preparation of linear peptides bearing the N-terminal free amino group. The N-terminal free amino group is produced in the expressed peptides by (a) the normal translation initiation with formyl-methionine (fM) and subsequent cleavage of the formyl group and methionine by PDF and MAP, respectively; (b) reprogrammed initiation with a desirable formylated amino acid (fX) and the subsequent cleavage of the formyl group by PDF; or (c) reprogrammed initiation with exotic peptides having a N-terminal free amino group such as γ-amino acids. (d, e) Methods for the generation of an ester linkage onto the nascent peptide chain. Peptides bearing an ester bond are expressed by genetic code reprogramming with (d) glycolic acid (^{HO}G); or (e) artificially induced peptidyl-tRNA drop-off. The resulting peptides containing the N-terminal free amino group and the C-P-ester motif at the C-terminus, respectively, undergo macrocyclization *via* diketopiperazine formation at the C-P-ester motif.

group at the N-terminus and a dkp thioester at the N-terminus could be readily generated in the FIT system *in situ* (Fig. 5d). Importantly, such constructs self-cyclize to produce macrocyclic peptides. We indeed demonstrated ribosomal expression of monocyclic backbone-cyclic peptides such as eptidemnamide[28] and scleramide,[78] as well as bicyclic and tetracyclic peptides interbridged by disulfide bonds, such as sunflower trypsin inhibitor (SFTI-1)[7] and rhesus θ defensin-1 (RTD-1)[79] (Fig. 6), in which the internal C residues actively assist the backbone cyclization *via* the formation of thioester intermediates between the sulfhydryl side chain and C-terminus.[73]

Backbone macrocyclic peptides containing non-proteinogenic amino acids

The genetic code reprogramming approach combining with the self-backbone cyclization mediated by C-P-^{HO}G can be also coupled with codon reassignments to non-proteinogenic amino acids. Taking such an advantage, RTD-1 containing *N*-methyl-amino acids has been expressed.[73] The residues chosen for codon assignments were those of side chains not involved in activity, *i.e.* substitution of such residues with *N*-methyl-amino acids would not affect on the RTD-1 wild type activity. In fact, RTD^{Me} (Fig. 6) containing three *N*-methyl-amino acids, ^{Me}G, ^{Me}A, and ^{Me}F, inhibits protease activity of anthrax lethal factor (LF) with a nearly identical activity to the wild type RTD-1. Because the incorporation of *N*-methyl-amino acid residues potentially increases the protease resistance of peptides, this approach in combination with an appropriate screening method (*vide infra*) will provide a new methodology to grant peptides with increased physiological stabilities.

Initiation codon reprogramming can also be coupled with the C-P-^{HO}G macrocyclization method. We have previously demonstrated the initiation reprogramming using not only short

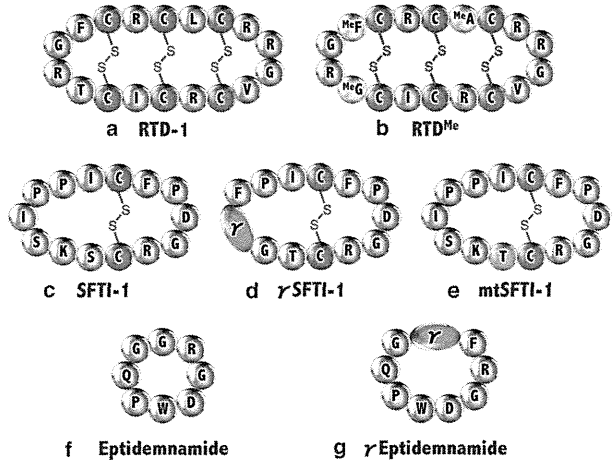


Fig. 6 Examples of backbone cyclic peptides expressed in the FIT system. (a) Wild type RTD-1 and (b) RTD-1 containing N-methyl amino acids (RTD^{Me}). (c) Wild type SFTI-1, (d) SFTI-1 containing aγ-amino acid(γSFTI-1) and (e) a mutant SFTI-1 involving a single mutation of position 2 from threonine to serine (mtSFTI-1). (f) Wild type eptidemnamide and (g) eptidemnamide containing a γ-amino acid(γ-eptidemnamide).

peptides, ranging from dipeptides to pentapeptides, but also those containing exotic amino acids such as D-α-amino acids (Fig. 5b) and β/γ-amino acids[65] (Fig. 5c). Despite the fact that such an exotic amino acid is notoriously difficult to incorporate into a peptide chain *via* elongation, we envision that a peptide containing the exotic amino acid(s) can be incorporated at the N-terminus by the initiation codon reprogramming and then the resulting peptide can be cyclized by the C-P-^{HO}G macrocyclization method. As a result, an exotic amino acid in the head-to-tail peptide is installed like in the middle of the sequence.

As a proof-of-concept experiment, we recently reported ribosomal synthesis of backbone macrocyclic peptides containing various γ-amino acids.[80] Dipeptides consisting of various γ-amino acids ligated with phenylalanine (γ-aa-F) were charged onto tRNA$^{fMet}_{CAU}$ by flexizyme, and used as initiators for the expression of peptides. In the downstream of the peptide sequence, the C-P-^{HO}G segment was continuously expressed, and thus the resulting peptides self-rearranged into the dkp-thioester at the C-terminus. The free γ-amino group in the initiator γ-aa-F was able to attack the dkp-thioester to yield backbone macrocyclic peptides containing various γ-amino acids. Unfortunately, peptides containing sterically demanding γ-amino acids poorly self-cyclized due to the inefficient nucleophilic attack of such a γ-amino group by the competition with dkp-thioester hydrolysis. However, the intramolecular sulfhydryl-assisting cyclization, *i.e.* expression of cysteine–cysteine inter-bridging bicyclic peptides by mimicking the structure of SFTI-1 (γSFTI-1) (Fig. 6), was successful in all cases of γ-aa-F's, giving desired bicyclic peptides containing γ-amino acids.

Induced peptidyl-tRNA drop-off promoting backbone cyclization

We recently reported an alternative method for ribosomal synthesis of inter-bridging backbone macrocyclic peptides promoted by the induced peptidyl-tRNA drop-off.[81] Peptidyl-tRNA drop-off could occur at any position during elongation, but much more frequently occurs at the last sense codon before the termination codon. This is because the rate of termination event promoted by release factors is known to be slower than the elongation, which in turn results in stalling the ribosome at the last sense codon and thus increases a chance of peptidyl-tRNA drop-off.[82] With this knowledge, one can propose that peptidyl-tRNA drop-off can be induced by artificial stalling of the ribosome during the elongation process. In fact, we were able to observe an efficient peptidyl-tRNA drop-off event at a vacant codon created by missing the corresponding amino acid in the FIT system.

Such a peptidyl-tRNA drop-off is generally considered to be an equivalent to the translation termination as an event, but as a product we have noticed that the peptidyl C-terminus is an ester bond ligating to the 3′-terminal hydroxyl group of tRNA. Significantly, this is chemically equivalent to the incorporation of ^{HO}G at the C-terminus of peptide.[73] On the basis of this concept, we designed a model mRNA that expressed the peptidyl-C-P sequence followed by a vacant codon (Fig. 5e). As expected, peptidyl-C-P-tRNA was dropped off during the elongation, and the self-rearrangement took place to yield the corresponding peptidyl-dkp thioester. When this strategy was applied to the expression of SFTI-1, we were able to observe the desired backbone cyclization.[81] Importantly, the vacant codon could be made not only by missing amino acid assignment, but also by missing a release factor in the FIT system. Therefore, it is possible that this methodology could apply to the expression of longer polypeptides, *i.e.* proteins, with a C-terminal thioester.

Rapid screening and deconvolution of functional peptides from the crude translated products

A unique aspect of mRNA-directed synthesis of peptides is the facile construction of diverse peptide libraries required for screening of active species. To build a technical foundation for the library construction as well as screening strategy, we demonstrated expression of a library of SFTI-1 (Fig. 7a) and rapid screening of bioactive sequences integrated with a limiting-dilution PCR deconvolution methodology[73] (Fig. 7b). The preliminary experiment showed that the crude product of wild-type SFTI-1 (wtSFTI-1) expressed in a FIT system under the assignment of the CUC leucine codon to ^{HO}G was able to inhibit trypsin inhibitory activity using a fluorogenic substrate, suggesting that the FIT system itself did not interfere the assay sensitivity.

A DNA library was constructed encoding SFTI-1 mutants in which three amino acid residues at positions 2, 4 and 6 embedded in the trypsin-binding loop were randomized by inserting NNK triplets (N = G, A, C or T and K = T or G) into the corresponding positions in the parental DNA template[73] (Fig. 7a).

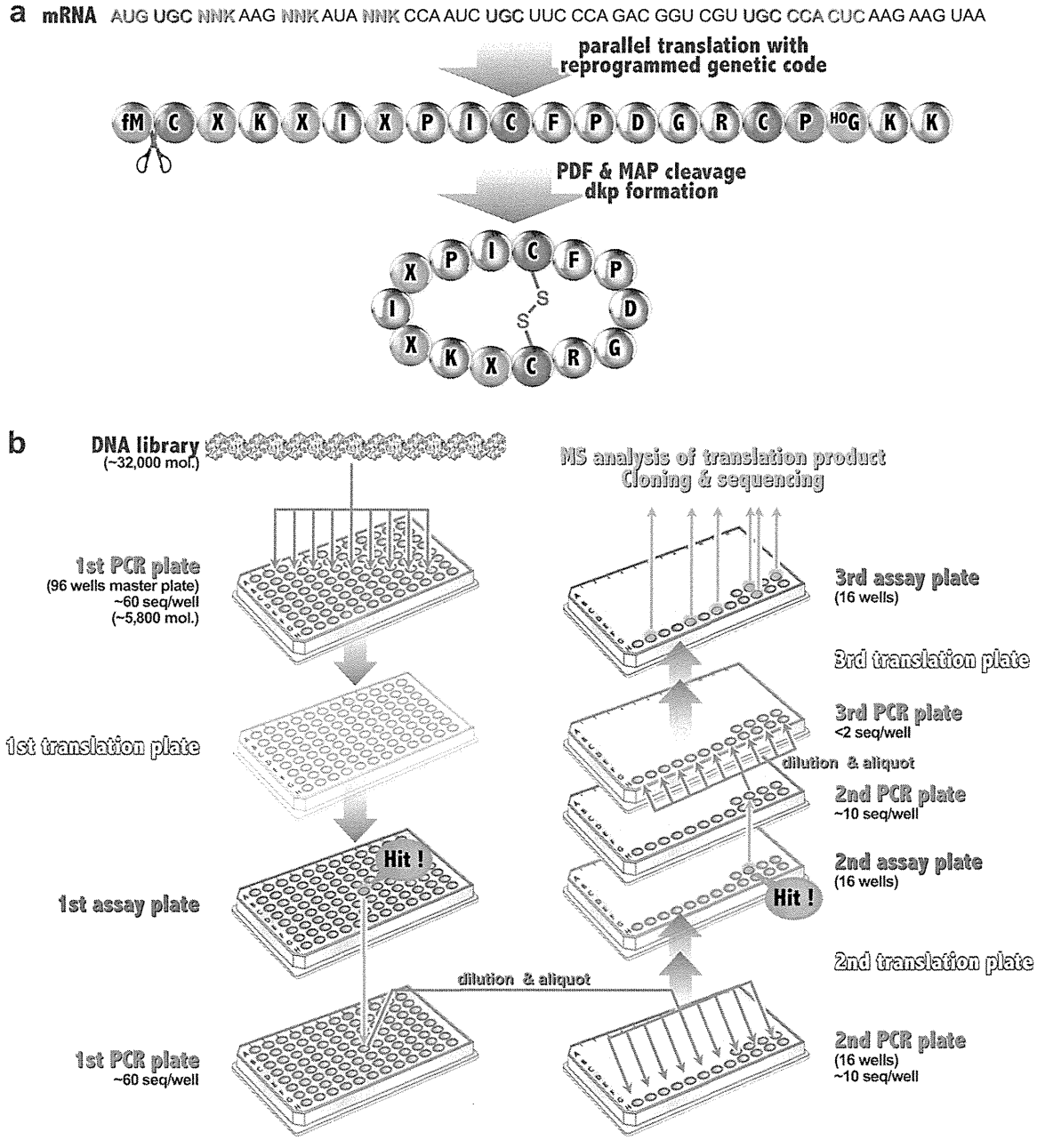


Fig. 7 Construction of a backbone cyclic peptide library in the FIT system and rapid screening of functional species. (a) Sequences of the mRNA library expressing the precursors of the SFTI-1 mutant library. ^{HO}G is assigned to the CUC codon by genetic code reprogramming. X denotes random amino acid residues assigned by a random nucleotide triplet consisting of NNK (N = G, A, C, or T and K = T or G). (b) Schematic representation of the procedures for rapid screening and deconvolution of active sequence(s) by the limiting-dilution PCR procedure.

A 96-well parallel one-pot expression was performed from the PCR-amplified DNA libraries originating from approximately 60 unique sequences, and the crude products were directly analyzed using the above assay system. A total of three rounds of screening and deconvolution of active sequences in the hit wells resulted in identifying a mutant SFTI-1 (mtSFTI-1) involving a single mutation of position 2 from threonine to serine, which was comparable to that of wtSFTI-1. Although the observed mutation was not surprising, the integration of the ribosomal expression of a SFTI-1 library with activity screening and limiting-dilution PCR deconvolution proved our strategy for rapid determination of active sequences.

The genetic code reprogramming method is novel and unique in terms of flexibility of peptide ring sizes and composition of amino acids, even with a non-proteinogenic side chain as well as main chain, for the synthesis of macrocyclic peptide libraries. Unlike intein-mediated macrocyclization, the ligation site would be not be completely restricted by the kind of residues, but yet a cysteine residue is occasionally helpful to cyclize the ring efficiently. The virtue of this method is facility to couple with existing *in vitro* assay systems and active species

are rapidly deconvoluted by a simple limiting dilution PCR method. On the other hand, because of this deconvolution strategy, the complexity of library would be limited with less than 10^5. Thus, the best use of this method would be integrated with a selection method using an appropriate *in vitro* display technique, in which a highly diverse library, as high as 10^{12} complexity, is used to select macrocyclic peptides that strongly bind a target; and then identify those with desired functions, *i.e.* antagonists or agonist activity, from a smaller "binding active" library using the limiting dilution PCR deconvolution method.

Conclusion

In recent years, peptide therapeutics have become an integrated part of pharmaceutical industry due to the higher selectivity and binding affinity against both intra- and extracellular targets. Although peptide therapeutics have greater potency compared to the small molecule therapies, they often suffer reduced bioavailability due to their less constrained conformation which in turn decreases their clinical efficacy. To address and overcome these drawbacks, a variety of peptide macrocyclization methods have become an important part of strategies. The respective methods disclosed in this review will give us new opportunities to prepare not only unique collections of macrocyclic peptide but also randomly composed libraries with variety levels of sequence and chemical diversities. Since in these methods the macrocyclic peptide libraries are genetically encoded, deconvolution of active sequences found from the libraries is much easier than that from chemical libraries. We have witnessed herein that they are proven to integrate with screening strategies, allowing researchers to quickly identify desired bioactive peptides.

Although none of these methods described in this review are yet perfect having some limitations, their further elaborations will probably develop more matured and reliable technologies. Particularly, a combination of one of these methods with another method or a brand new approach will likely overcome the limitations, giving a better or even the best way to discover therapeutically relevant peptides.

Acknowledgements

This work was supported by grants of Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (S) (16101007), Specially Promoted Research (21000005), a research and development projects of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (NEDO) to H.S., grants of Japan Society for the Promotion of Science Grants-in-Aid for Young Scientists (B) (22710210) to T.K., and (B) (22750145) to Y.G.

References

1 L. Cascales and D. J. Craik, *Org. Biomol. Chem.*, 2010, **8**, 5035–5047.
2 T. J. Oman and W. A. van der Donk, *Nat. Chem. Biol.*, 2010, **6**, 9–18.
3 J. Lee, J. McIntosh, B. J. Hathaway and E. W. Schmidt, *J. Am. Chem. Soc.*, 2009, **131**, 2122–2124.
4 S. Duquesne, V. Petit, J. Peduzzi and S. Rebuffat, *J. Mol. Microbiol. Biotechnol.*, 2007, **13**, 200–209.
5 N. Ziemert, K. Ishida, A. Liaimer, C. Hertweck and E. Dittmann, *Angew. Chem., Int. Ed.*, 2008, **47**, 7756–7759.
6 D. J. Craik and A. C. Conibear, *J. Org. Chem.*, 2011, **76**, 4805–4817.
7 M. L. Korsinczky, H. J. Schirra and D. J. Craik, *Curr. Protein Pept. Sci.*, 2004, **5**, 351–364.
8 I. Saska, A. D. Gillon, N. Hatsugai, R. G. Dietzgen, I. Hara-Nishimura, M. A. Anderson and D. J. Craik, *J. Biol. Chem.*, 2007, **282**, 29721–29728.
9 A. D. Gillon, I. Saska, C. V. Jennings, R. F. Guarino, D. J. Craik and M. A. Anderson, *Plant J.*, 2008, **53**, 505–515.
10 J. S. Mylne, M. L. Colgrave, N. L. Daly, A. H. Chanson, A. G. Elliott, E. J. McCallum, A. Jones and D. J. Craik, *Nat. Chem. Biol.*, 2011, **7**, 257–259.
11 S. K. Mazmanian, G. Liu, H. Ton-That and O. Schneewind, *Science*, 1999, **285**, 760–763.
12 X. Huang, A. Aulabaugh, W. Ding, B. Kapoor, L. Alksne, K. Tabei and G. Ellestad, *Biochemistry*, 2003, **42**, 11307–11315.
13 H. Mao, S. A. Hart, A. Schink and B. A. Pollok, *J. Am. Chem. Soc.*, 2004, **126**, 2670–2671.
14 S. Tsukiji and T. Nagamune, *ChemBioChem*, 2009, **10**, 787–798.
15 R. Parthasarathy, S. Subramanian and E. T. Boder, *Bioconjugate Chem.*, 2007, **18**, 469–476.
16 C. M. Ireland, A. R. Durso, Jr., R. A. Newman and M. P. Hacker, *J. Org. Chem.*, 1982, **47**, 1807–1811.
17 E. W. Schmidt, J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood and J. Ravel, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7315–7320.
18 A. R. Carroll, J. C. Coll, D. J. Bourne, J. K. Macleod, T. M. Zabriskie, C. M. Ireland and B. F. Bowden, *Aust. J. Chem.*, 1996, **49**, 659–667.
19 K. Ishida, H. Nakagawa and M. Murakami, *J. Nat. Prod.*, 2000, **63**, 1315–1317.
20 C. Portmann, J. F. Blom, K. Gademann and F. Jüttner, *J. Nat. Prod.*, 2008, **71**, 1193–1196.
21 F. Jüttner, A. K. Todorova, N. Walch and W. von Philipsborn, *Phytochemistry*, 2001, **57**, 613–619.
22 R. Banker and S. Carmeli, *J. Nat. Prod.*, 1998, **61**, 1248–1251.
23 R. G. Linington, J. Gonzalez, L. D. Ureña, L. I. Romero, E. Ortega-Barría and W. H. Gerwick, *J. Nat. Prod.*, 2007, **70**, 397–401.
24 J. Ogino, R. E. Moore, G. M. Patterson and C. D. Smith, *J. Nat. Prod.*, 1996, **59**, 581–586.
25 W. E. Houssen and M. Jaspars, *ChemBioChem*, 2010, **11**, 1803–1815.
26 M. S. Donia, J. Ravel and E. W. Schmidt, *Nat. Chem. Biol.*, 2008, **4**, 341–343.
27 J. A. McIntosh, C. R. Robertson, V. Agarwal, S. K. Nair, G. W. Bulaj and E. W. Schmidt, *J. Am. Chem. Soc.*, 2010, **132**, 15499–15501.
28 M. S. Donia, B. J. Hathaway, S. Sudek, M. G. Haygood, M. J. Rosovitz, J. Ravel and E. W. Schmidt, *Nat. Chem. Biol.*, 2006, **2**, 729–735.
29 F. B. Perler, E. O. Davis, G. E. Dean, F. S. Gimble, W. E. Jack, N. Neff, C. J. Noren, J. Thorner and M. Belfort, *Nucleic Acids Res.*, 1994, **22**, 1125–1127.
30 S. Chong, Y. Shao, H. Paulus, J. Benner, F. B. Perler and M. Q. Xu, *J. Biol. Chem.*, 1996, **271**, 22159–22168.
31 T. W. Muir, D. Sondhi and P. A. Cole, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6705–6710.
32 D. Schwarzer and P. A. Cole, *Curr. Opin. Chem. Biol.*, 2005, **9**, 561–569.
33 D. Cowburn, A. Shekhtman, R. Xu, J. J. Ottesen and T. W. Muir, *Methods Mol. Biol. (Totowa, N. J.)*, 2004, **278**, 47–56.
34 S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus and M. Q. Xu, *Gene*, 1997, **192**, 271–281.
35 S. Chong, G. E. Montello, A. Zhang, E. J. Cantor, W. Liao, M. Q. Xu and J. Benner, *Nucleic Acids Res.*, 1998, **26**, 5109–5115.
36 J. A. Camarero and T. W. Muir, *J. Am. Chem. Soc.*, 1999, **121**, 5597–5598.
37 H. Iwai and A. Pluckthun, *FEBS Lett.*, 1999, **459**, 166–172.
38 T. C. J. Evans, J. Benner and M. Q. Xu, *J. Biol. Chem.*, 1999, **274**, 18359–18363.
39 R. H. Kimura, A. T. Tran and J. A. Camarero, *Angew. Chem., Int. Ed.*, 2006, **45**, 973–976.
40 J. A. Camarero, R. H. Kimura, Y. H. Woo, A. Shekhtman and J. Cantor, *ChemBioChem*, 2007, **8**, 1363–1366.
41 D. J. Craik, N. L. Daly, J. Mulvenna, M. R. Plan and M. Trabi, *Curr. Protein Pept. Sci.*, 2004, **5**, 297–315.

42 U. Göransson, M. Sjögren, E. Svangård, P. Claeson and L. Bohlin, *J. Nat. Prod.*, 2004, **67**, 1287–1290.
43 K. R. Gustafson, T. C. McKee and H. R. Bokesch, *Curr. Protein Pept. Sci.*, 2004, **5**, 331–340.
44 H. Wu, Z. Hu and X. Q. Liu, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 9226–9231.
45 T. C. J. Evans, D. Martin, R. Kolly, D. Panne, L. Sun, I. Ghosh, L. Chen, J. Benner, X. Q. Liu and M. Q. Xu, *J. Biol. Chem.*, 2000, **275**, 9091–9094.
46 C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 13638–13643.
47 C. P. Scott, E. Abel-Santos, A. D. Jones and S. J. Benkovic, *Chem. Biol.*, 2001, **8**, 801–815.
48 T. M. Kinsella, C. T. Ohashi, A. G. Harder, G. C. Yam, W. Li, B. Peelle, E. S. Pali, M. K. Bennett, S. M. Molineaux, D. A. Anderson, E. S. Masuda and D. G. Payan, *J. Biol. Chem.*, 2002, **277**, 37512–37518.
49 A. R. Horswill, S. N. Savinov and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 15591–15596.
50 A. Tavassoli and S. J. Benkovic, *Angew. Chem., Int. Ed.*, 2005, **44**, 2760–2763.
51 A. Tavassoli, Q. Lu, J. Gam, H. Pan, S. J. Benkovic and S. N. Cohen, *ACS Chem. Biol.*, 2008, **3**, 757–764.
52 L. Cheng, T. A. Naumann, A. R. Horswill, S. J. Hong, B. J. Venters, J. W. Tomsho, S. J. Benkovic and K. C. Keiler, *Protein Sci.*, 2007, **16**, 1535–1542.
53 T. A. Naumann, A. Tavassoli and S. J. Benkovic, *ChemBioChem*, 2008, **9**, 194–197.
54 J. A. Kritzer, S. Hamamichi, J. M. McCaffery, S. Santagata, T. A. Naumann, K. A. Caldwell, G. A. Caldwell and S. Lindquist, *Nat. Chem. Biol.*, 2009, **5**, 655–663.
55 H. Murakami, H. Saito and H. Suga, *Chem. Biol.*, 2003, **10**, 655–662.
56 N. Niwa, Y. Yamagishi, H. Murakami and H. Suga, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 3892–3894.
57 Y. Goto, T. Katoh and H. Suga, *Nat. Protoc.*, 2011, **6**, 779–790.
58 Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa and T. Ueda, *Nat. Biotechnol.*, 2001, **19**, 751–755.
59 Y. Sako, J. Morimoto, H. Murakami and H. Suga, *J. Am. Chem. Soc.*, 2008, **130**, 7232–7234.
60 Y. Goto, H. Murakami and H. Suga, *RNA*, 2008, **14**, 1390–1398.
61 Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami and H. Suga, *ACS Chem. Biol.*, 2008, **3**, 120–129.
62 T. Kawakami, H. Murakami and H. Suga, *Chem. Biol.*, 2008, **15**, 32–42.
63 T. Kawakami, H. Murakami and H. Suga, *J. Am. Chem. Soc.*, 2008, **130**, 16861–16863.
64 A. Ohta, H. Murakami, E. Higashimura and H. Suga, *Chem. Biol.*, 2007, **14**, 1315–1322.
65 Y. Goto and H. Suga, *J. Am. Chem. Soc.*, 2009, **131**, 5040–5041.
66 J. D. Bain, E. S. Diala, C. G. Glabe, D. A. Wacker, M. H. Lyttle, T. A. Dix and A. R. Chamberlin, *Biochemistry*, 1991, **30**, 5411–5421.
67 J. A. Ellman, D. Mendel and P. G. Schultz, *Science*, 1992, **255**, 197–200.
68 A. Ohta, H. Murakami and H. Suga, *ChemBioChem*, 2008, **9**, 2773–2778.
69 J. T. Koh, V. W. Cornish and P. G. Schultz, *Biochemistry*, 1997, **36**, 11314–11322.
70 P. M. England, Y. Zhang, D. A. Dougherty and H. A. Lester, *Cell (Cambridge, Mass.)*, 1999, **96**, 89–98.
71 P. M. England, H. A. Lester and D. A. Dougherty, *Biochemistry*, 1999, **38**, 14409–14415.
72 T. Kawakami and S. Aimoto, *Chem. Lett.*, 2007, 76–77.
73 T. Kawakami, A. Ohta, M. Ohuchi, H. Ashigai, H. Murakami and H. Suga, *Nat. Chem. Biol.*, 2009, **5**, 888–890.
74 R. A. Bradshaw, W. W. Brickey and K. W. Walker, *Trends Biochem. Sci.*, 1998, **23**, 263–267.
75 F. Sherman, J. W. Stewart and S. Tsunasawa, *BioEssays*, 1985, **3**, 27–31.
76 S. Huang, R. C. Elliott, P. S. Liu, R. K. Koduri, J. L. Weickmann, J. H. Lee, L. C. Blair, P. Ghosh-Dastidar, R. A. Bradshaw and K. M. Bryan, *Biochemistry*, 1987, **26**, 8242–8246.
77 J. P. Boissel, T. J. Kasper and H. F. Bunn, *J. Biol. Chem.*, 1988, **263**, 8443–8449.
78 A. C. Whyte, B. K. Joshi, J. B. Gloer, D. T. Wicklow and P. F. Dowd, *J. Nat. Prod.*, 2000, **63**, 1006–1009.
79 M. E. Selsted, *Curr. Protein Pept. Sci.*, 2004, **5**, 365–371.
80 Y. Ohshiro, E. Nakajima, Y. Goto, S. Fuse, T. Takahashi, T. Doi and H. Suga, *ChemBioChem*, 2011, **12**, 1183–1187.
81 T. J. Kang, Y. Hayashi and H. Suga, *Angew. Chem., Int. Ed.*, 2011, **50**, 2159–2161.
82 J. Menez, V. Heurgué-Hamard and R. H. Buckingham, *Nucleic Acids Res.*, 2000, **28**, 4725–4732.

Flexizymes for genetic code reprogramming

Yuki Goto[1,4], Takayuki Katoh[1,4] & Hiroaki Suga[1–3]

[1]Department of Chemistry, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo, Japan. [2]Research Center for Advanced Science and Technology, The University of Tokyo, Meguro-ku, Tokyo, Japan. [3]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology and College of Pharmacy, Seoul National University, Seoul, Korea. [4]These authors contributed equally to this work. Correspondence should be addressed to H.S. (hsuga@chem.s.u-tokyo.ac.jp).

Published online 12 May 2011; doi:10.1038/nprot.2011.331

Genetic code reprogramming is a method for the reassignment of arbitrary codons from proteinogenic amino acids to nonproteinogenic ones; thus, specific sequences of nonstandard peptides can be ribosomally expressed according to their mRNA templates. Here we describe a protocol that facilitates genetic code reprogramming using flexizymes integrated with a custom-made *in vitro* translation apparatus, referred to as the flexible *in vitro* translation (FIT) system. Flexizymes are flexible tRNA acylation ribozymes that enable the preparation of a diverse array of nonproteinogenic acyl-tRNAs. These acyl-tRNAs read vacant codons created in the FIT system, yielding the desired nonstandard peptides with diverse exotic structures, such as *N*-methyl amino acids, D-amino acids and physiologically stable macrocyclic scaffolds. The facility of the protocol allows a wide variety of applications in the synthesis of new classes of nonstandard peptides with biological functions. Preparation of flexizymes and tRNA used for genetic code reprogramming, optimization of flexizyme reaction conditions and expression of nonstandard peptides using the FIT system can be completed by one person in ~1 week. However, once the flexizymes and tRNAs are in hand and reaction conditions are fixed, synthesis of acyl-tRNAs and peptide expression is generally completed in 1 d, and alteration of a peptide sequence can be achieved by simply changing the corresponding mRNA template.

INTRODUCTION

The translation apparatus polymerizes amino acids in accordance with the sequence information encoded in mRNA templates with remarkably high fidelity. In spite of the governance of the genetic code restricting the use of 20 proteinogenic amino acids in translation, it is possible to alter this rule by appropriate manipulations of the translation system, wherein some components are excluded in order to reassign the codons from proteinogenic to nonproteinogenic (or artificial) amino acids. Such a custom-reconstituted *in vitro* translation system could enable us to 'reprogram' the governance of the genetic code and thereby the translation apparatus could turn into machinery for the synthesis of nonstandard polypeptides consisting of nonproteinogenic amino acids[1–3].

Despite the fact that the above idea sounds simple, in practice it is technically very challenging. Two major practical stumbling blocks exist, in addition to a number of minor issues, which can substantially hinder the execution of such experiments. First, the preparation of mischarged acyl-tRNAs with desired nonproteinogenic amino acids (Xaa-tRNAs; Xaa denotes any amino acid including proteinogenic aminoacyl donors as well as nonproteinogenic ones) is not a straightforward process. The classical methodology relies on semienzymatic synthesis, in which a chemically prepared Xaa dinucleotide is enzymatically ligated to a tRNA lacking the 3′-terminal dinucleotide CA[4]. Although this approach is theoretically applicable to any type of amino acid, it is technically difficult and occasionally fails to yield sufficient quality and quantity of the desired mischarged Xaa-tRNA for translation. Thus, from a practical point of view, this step has made genetic code reprogramming technology very specialized and laborious, thereby seriously limiting its broad application. A more recent approach using appropriate mutants of a subset of aminoacyl-tRNA synthetases (ARSs)[5] is efficient and applicable *in vivo*, but nevertheless limits the choice of nonproteinogenic amino acids. Thus, for the genetic code reprogramming purpose, this approach is not necessarily the best.

The second major obstacle concerns how multiple codon reassignments can be achieved while maintaining high efficiency and fidelity. Although Forster's demonstration clearly showed the potential of genetic code reprogramming for the ribosomal synthesis of nonstandard peptides[1], questions still remained as to whether the reprogrammed codons could coexist with other elongation codons that assign proteinogenic amino acids and/or with other reprogrammed elongation codons that assign nonproteinogenic amino acids. Importantly, the nonproteinogenic amino acids in past studies using sense or nonsense suppression strategies[6–8] are 'good' ones that are well known to be easily incorporable into peptide chains; such amino acids are often aromatic amino acids (i.e., phenylalanine analogs) or those structurally similar to proteinogenic amino acids. On the other hand, 'bad' nonproteinogenic amino acids such as *N*-methyl amino acids or *N*-alkyl glycines are excluded from most studies, particularly from those involving multiple incorporations. This is because the inefficient incorporation of 'bad' ones tends to result in the termination of peptide synthesis due to unwanted competition with RFs or peptidyl-tRNA drop off, and it occasionally results in the misincorporation of proteinogenic amino acids. Therefore, further elaboration of the reconstituted *in vitro* translation system is required to realize the full potential of genetic code reprogramming for the ribosomal synthesis of nonstandard peptides.

In 2003, we introduced a prototype flexizyme (Fx3), flexible tRNA acylation ribozyme to facilitate the aminoacylation of tRNA using amino acid substrates activated by cyanomethyl ester; however, it was applicable to only aromatic amino acids[9]. We subsequently developed three next-generation flexizymes, referred to as dFx, eFx and aFx[3,10]. The complete set of flexizymes accepts virtually any amino acid as acyl-donor substrates by simply matching the flexizyme with the active group in the acyl-donor substrate (see Experimental design section for details). Indeed, the flexizyme system has been applied to the synthesis of a wide variety of Xaa-tRNAs,

where Xaa represents ordinary proteinogenic amino acids[3] as well as nonproteinogenic amino acids such as those with artificial side chains[3,11], D-configurations[3,12], *N*-acyl[13] and *N*-alkyl modifications[14,15]; that is, virtually any kind of amino acid. Moreover, it is also capable of charging hydroxyacyl[16] and peptidyl groups involving D-amino acid residues[17] onto tRNAs (for the convenience of description, nonproteinogenic amino and hydroxyl acyl as well as peptidyl groups are abbreviated to Xaa). In addition to the versatility to acyl-substrates, flexizymes are able to accept a wide array of tRNAs without depending on body and anticodon sequences, unlike ARSs, via simple base pair recognition of the 3′ end of tRNA[18]. As the greatest benefit of flexizymes is their technical simplicity and facility (i.e., mixing the appropriate flexizyme with the desired tRNA and acid substrate, followed by incubation on ice for a few hours, thereby yielding the desired Xaa-tRNAs), the flexizyme system eliminates the previous technical barrier in the preparation of Xaa-tRNA for genetic code manipulation.

The second obstacle could be resolved by developing a more sophisticated *in vitro* translation system optimized for the specific codon reassignments themselves. In this system, arbitrarily chosen codons are emptied by simply excluding the corresponding proteinogenic amino acids and cognate ARSs from the translation components, as a result of which the potential competitors for the aimed reprogrammed codons do not exist. These 'vacant' codons are then reassigned with desired nonproteinogenic amino acids by supplying Xaa-tRNAs whose anticodons are capable of reading the vacant codons. Moreover, the tRNA body sequences in Xaa-tRNAs are engineered so as not to be recognized by endogenous ARSs (see Experimental design section for details). This strategy has markedly improved the incorporation efficiency of 'bad' nonproteinogenic amino acids. For instance, various *N*-methyl amino acids and *N*-alkyl glycines are elongated in succession 12 and 6 times, respectively[14,15]. On the other hand, even though flexizymes facilitate the preparation of tRNAs charged with D-amino acids or β-amino acids, some of these amino acids could not be consecutively elongated because of poor compatibility of naturally occurring ribosomes (H. Murakami and H.S., unpublished data). It should be noted that the initiation event turned out to be more amenable to a wide variety of amino acids, including D-amino acids, β-amino acids and peptides containing such amino acids[12,13,17]; therefore, initiation reprogramming is an effective strategy to synthesize peptides containing such amino acids.

In this article, we describe protocols for genetic code reprogramming using flexizymes integrated into a customized reconstituted translation system[3,11–17], referred to as the FIT (flexible *in vitro* translation) system (**Fig. 1**). In the FIT system, unnecessary translation components such as ARSs, amino acids, methionyl-tRNA formyl transferase, 10-formyl-tetrahydrofolate, and/or release factors can be excluded depending on the desired codons to be reprogrammed. The minimal components of the translation apparatus are customized to match the use of any desired codon(s), while retaining maximum peptide expression levels and decoding fidelity. Each customized translation apparatus is then supplemented with Xaa-tRNAs prepared by the flexizyme technology to occupy the vacant codons, yielding the desired peptide product in high purity. If necessary, other enzymes, e.g., peptide deformylase and methionine aminopeptidase, can be added to the FIT system[19]. By means of the genetic code reprogramming, along with the unique adopting chemistry, the translation system turns into machinery for the synthesis of natural product-like nonstandard peptides consisting of *N*-methyl peptide backbone and macrocyclic scaffolds[13–15,19].

We describe (in ANTICIPATED RESULTS) two practical examples of genetic code reprogramming, wherein an *N*-methyl peptide and peptoid-peptide hybrid cyclized by a thioether bond were expressed in the FIT system. The *N*-methylated backbone is often found in peptidic natural products, which are generally synthesized by multi-enzyme clusters called non-ribosomal peptide synthetases[20]. Peptoid is a class of artificial peptidomimetic molecules composed of *N*-alkylated glycines[21]. In both cases, the substitutions on the α-amido nitrogen of the backbone potentially contribute to conformational rigidity, target affinity peptidase resistance and membrane permeability[22,23]. These functional advantages increase opportunities to turn such peptides into a novel class of therapeutic reagents. Although it has been known that *N*-substituted amino acids are difficult to incorporate into a peptide chain by conventional suppression methods using amber or four-base codons, we recently reported that *N*-methyl peptide and peptoid-peptide hybrid syntheses could be accomplished in an FIT system in which up to seven codons are assigned to α-*N*-methyl amino acids or *N*-substituted glycine[14,15].

For cyclization of expressed peptides, translation initiated with an α-amino acid bearing an *N*-chloroacetamide group (e.g., α-*N*-chloroacetyl phenylalanine, ^{ClAc}Phe) allows us to express a wide variety of linear peptides having the chloroacetyl group on the N terminus from the corresponding DNA templates[13]. When a cysteine residue is embedded within the peptide sequence, the sulfhydryl group of the side chain spontaneously reacts with the N-terminal chloroacetamide group to yield a cyclic peptide closed by a thioether bond. This intramolecular cyclization spontaneously occurs in the translation mixture without need for any additional external reagents in a high yield (in most cases in a nearly quantitative manner). From extensive studies, it appears that the reaction selectively takes place between chloroacetamide and cysteine thiol, independent of sequence composition and the length of the peptide (up to at least a 15-mer length). Importantly, we have shown that such thioether-cyclized peptides show remarkable resistance against peptidases under reducing conditions[24].

Experimental design

The choice of flexizyme pairing with the acyl-donor leaving group. The three flexizymes developed by our laboratory are active in the specific combination with a leaving group (**Fig. 2**)[3,10]. Despite the fact that the differences in sequences between these flexizymes are only a few bases, they show the best activity toward their cognate leaving groups in acyl-donor substrates. Thus, we must choose an appropriate pair of flexizyme and leaving group depending upon the structure of objective acyl-donor substrates. For those containing aromatic side chains, e.g., Phe, Tyr, Trp and their analogs, eFx is the choice to pair with those activated by a cyanomethyl ester. For those bearing non-aromatic side chains, dFx is the primary choice pairing with the 3,5-dinitrobenzyl ester leaving group. Although this pair is the most versatile and generic choice for tRNA acylation, acyl-donor substrates with certain structures (e.g., α-*N*-acyl, β-branched or bulky side chains) occasionally result in sluggish reaction rates and poor charging yields. In such a case, an alternative choice is eFx paired with the 4-chlorobenzyl thioester leaving group, which often shows substantial improvement. The last pair is aFx and

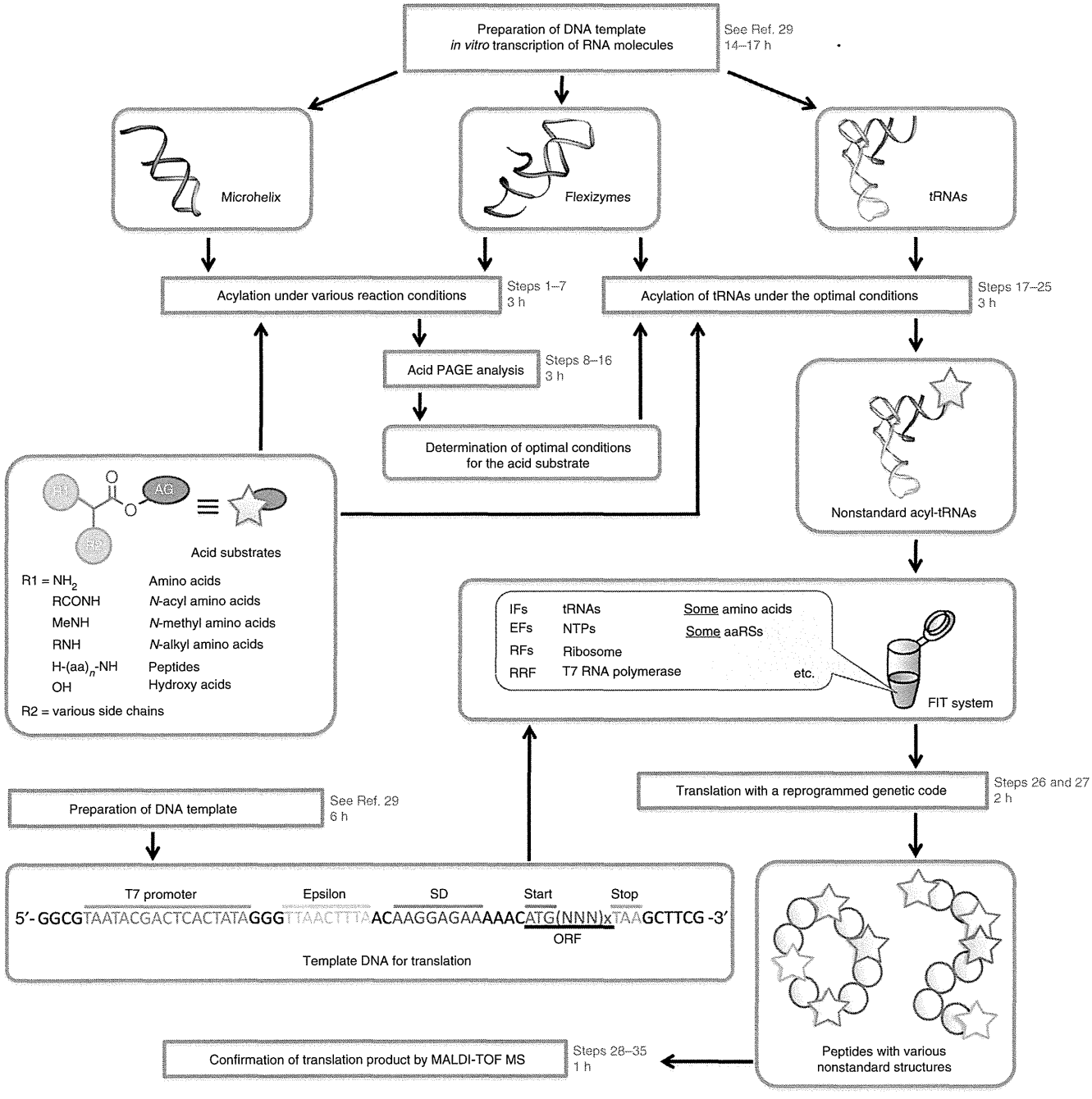


Figure 1 | *Schematic overview of genetic code reprogramming and expression of nonstandard peptides using the FIT system. See the Experimental design section for details.*

acyl-donors bearing the leaving group of 4-[(2-aminoethyl)carbam oyl]benzyl thioester. This pair has been developed for a special case in which the activation of acid with any of the above leaving groups results in poor solubility in the reaction buffer (note that flexizymes remain active in up to 25% (vol/vol) DMSO-containing buffer, with higher concentrations resulting in a gradual loss in their activity). The 4-[(2-aminoethyl)carbamoyl]benzyl thioester leaving group improves the water solubility of the acyl-donor, and therefore this combination overcomes such a limitation. All flexizymes are able to accept tRNAs bearing any choice of anticodon and body sequence as long as the 3′-terminal region (5′-(A/G/U)CCA-3′, where bases interacting with the 3′-terminal bases of flexizyme are underlined; see below) can be matched with that of flexizymes (5′-AGGU-3′, where A does not form a base pair).

Conditions for the flexizyme-catalyzing tRNA acylation. Optimal conditions for the flexizyme reactions vary depending on the acyl-donor substrates. The empirically determined optimal conditions for representative substrates are summarized in **Table 1**. Although the generic conditions, in which the mixture of tRNA and flexizyme with 25 mM acyl-donor substrate is incubated for 2 h at pH 7.5 in the presence of 600 mM $MgCl_2$, would work in most cases, it is occasionally necessary to optimize conditions depending on substrate structures; thus, the data in **Table 1** are useful

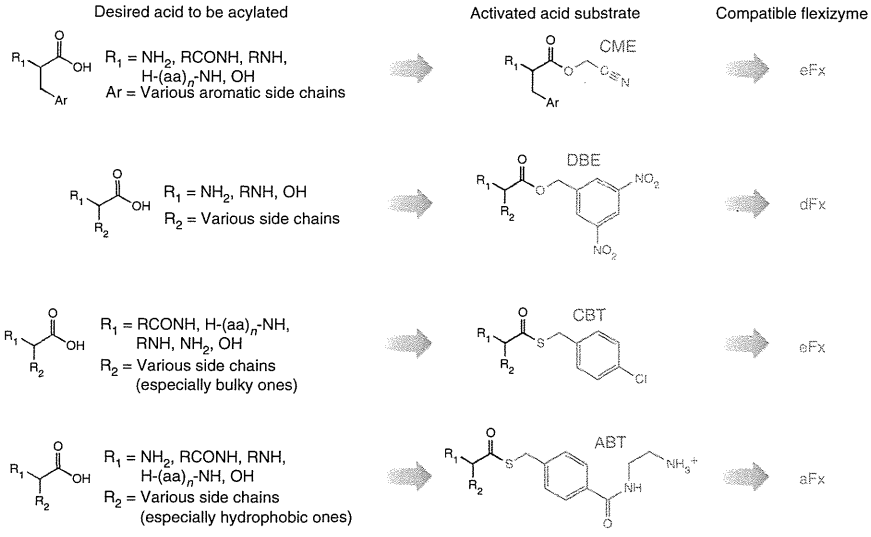


Figure 2 | Summary of the appropriate combination of active groups and flexizymes for various acid substrates.

as reference conditions that can be modified and optimized if necessary. The recommended method for optimizing the reaction conditions when using a new acid substrate (see Steps 1–16 in the PROCEDURE section) is to use a microhelix RNA instead of tRNA. The convenience of using this small RNA is that the acylation yield can be directly determined by the band intensity of Xaa-microhelix RNA separated from that of free RNA on acid PAGE[3,25]. The acylation reaction mixture containing the microhelix RNA, flexizyme and acyl-donor substrate is incubated on ice and sampled at several time points, with the resulting RNA analyzed by acid-PAGE to determine the optimal conditions (**Fig. 3**). Conditions resulting in a yield of greater than 10% in this analysis can be applied to the flexizyme-catalyzed tRNA acylation; thus, prepared Xaa-tRNAs are generally sufficient for the experiments of genetic code reprogramming. It is also possible to monitor the formation of Xaa-tRNAs by treating with an appropriate biotinylation reagent, followed by streptavidin gel-shift assay using PAGE, although this method is limited to the use of substrates with a free α-amino group[26,27]. If the observed yield of Xaa-RNA or tRNA is less than 10%, the pH of the reaction buffer and concentration of the acyl-donor substrate can be increased, which often results in a yield sufficient for the following experiments.

Transfer RNAs used for genetic code reprogramming. Naturally occurring tRNAs have specific body sequences along with various base modifications[28]. Such structural uniqueness existing in each tRNA dictates not only the respective tRNA identity to cognate ARSs and amino acids but also the best efficiency and fidelity to decode the cognate codon. On the contrary, the tRNAs used for genetic code reprogramming are *in vitro* transcripts. Even though amino acids and ARSs corresponding to the reprogrammed codons are absent, it cannot be completely ruled out that other endogenous ARSs unexpectedly mischarge the deacylated tRNAs (originating from Xaa-tRNAs) with their cognate amino acids. Even trace amounts of such mischarged aa-tRNAs are far more effective than Xaa-tRNAs for the incorporation into the peptide chain as they are naturally better substrates for ribosome/EF-Tu. Therefore, it is crucial to reduce the possibility of the formation of mischarged aa-tRNAs as much as possible. To this end, the body sequences of Xaa-tRNAs are artificially engineered to be inert against endogenous ARSs, so-called 'orthogonal' tRNAs. Investigation of several different tRNA constructs for orthogonal characteristics and simultaneously high decoding efficiency in translation has allowed us to identify $tRNA^{AsnE2}_{NNN}$ (NNN denotes anticodon triplet)[16], the body sequence of which was derived from an *Escherichia coli* (*E. coli*) asparaginyl $tRNA^{Asn}$. Notably, this engineered tRNA body sequence maintains

TABLE 1 | Recommended acylation conditions for representative amino acids and their reference yields of acyl-microhelix RNAs.

Amino acid	Active group	Flexizyme	Reaction time (h)	Yield (%)
^{ClAc}Phe	CME	eFx	2	89
^{Me}Tyr	CME	eFx	10[a]	40
dE-dK-dE-dK-F	CME	eFx	2	62
^{HO}Yme	CME	eFx	2[b]	91
^{Me}Ala	DBE	dFx	2	60
^{Me}Gly	DBE	dFx	2	64
^{Pr}Gly	DBE	dFx	6	73
^{Bu}Gly	DBE	dFx	6	70
D-Asp	DBE	dFx	2	47
β-Ala	DBE	dFx	22[c]	75
^{Ac}Arg	CBT	eFx	72	51
Ile	CBT	eFx	6	17
Opa	ABT	aFx	2	48

Note: ^{ClAc}Phe, α-*N*-chloroacetyl phenylalanine; ^{Me}Tyr, α-*N*-methyl tyrsine; dE-dK-dE-dK-F, D-glutamyl-D-lysinyl-D-glutamyl-D-lysinyl-phenylalanine; ^{HO}Yme, *p*-methoxyphenyllactic acid; ^{Me}Ala, α-*N*-methyl alanine; ^{Me}Gly, α-*N*-methyl glycine; ^{Pr}Gly, α-*N*-propyl glycine; ^{Bu}Gly, α-*N*-butyl glycine; D-Asp, D-Aspartic acid; β-Ala, β-Alanine; ^{Ac}Arg,α-*N*-acetyl arginine; Ile, isoleucine; Opa, 2-amino-3-(octanamido)propanoic acid.
[a]Final concentration of acid substrate is increased to 40 mM. [b]Reaction was carried out in HEPES-KOH buffer (pH 8.0). [c]Reaction was carried out in Bicine-KOH buffer (pH 9.0).