Table 1. Kinetic and Equilibrium Constants of Macrocyclic *N*-Methyl-Peptides against E6AP HECT Domain

Peptide	Sequence	k_{on} (x10^6) $M^{-1}s^{-1}$	k_{off} (x10^{-3}) s^{-1}	K_d nM
CM_{11}-1	Ac^{D}WCDV^{Me}SGR^{Me}F^{Me}GY^{Me}FPCG-NH_2 (S–C)	2.66	1.60	0.60
LM_{11}-1	Ac^{D}WCDV^{Me}SGR^{Me}F^{Me}GY^{Me}FPCG-NH_2	0.047	8.46	180
CM_{11}-1S_1	Ac^{D}WSDV^{Me}SGR^{Me}F^{Me}GY^{Me}FPCG-NH_2 (S–C)	1.33	15.5	11.7
CP_{11}-1	Ac^{D}WCDV S GR F G Y F PCG-NH_2 (S–C)	not detectable		>1000
LP_{11}-1	Ac^{D}WCDV S GR F G Y F PCG-NH_2	not detectable		>1000
CM_{11}-3	Ac^{D}WCDV^{Me}SHY^{Me}F P Y ^{Me}FPCG-NH_2 (S–C)	0.257	0.320	1.24
LM_{11}-3	Ac^{D}WCDV^{Me}SHY^{Me}F P Y^{Me}FPCG-NH_2	0.014	4.46	325
CM_{11}-5	Ac^{D}WCSN^{Me}AGH^{Me}F P Y^{Me}FPCG-NH_2 (S–C)	2.19	2.87	1.31
LM_{11}-5	Ac^{D}WCSN^{Me}AGH^{Me}F P Y^{Me}FPCG-NH_2	0.089	20.6	231

Data were collected by the standard SPR method using Biocore T100 and the constants were generated by the equipped data fitting program. See also Figures S2 and S3.

analysis against the E6AP Avi-$(His)_6$-GB1-HECT domain immobilized on a SAv-sensor chip (Table 1). All macrocyclic CM_{11}-peptides have values of k_{on} with a range of 0.2–3 × 10^6 $M^{-1}s^{-1}$, k_{off} with a range of 0.3–3 × 10^{-3} s^{-1}, resulting in the dissociation constants (K_d) with subnanomolar to 1 nM range. Thus, the representative CM_{11}-peptides studied here have remarkably strong affinity to E6AP HECT domain, and particularly CM_{11}-1 has the lowest K_d values among the CM_{11}-peptides. In addition to the CM_{11}-peptides, we synthesized the respective linear peptides bearing N-terminal acetyl group, referred to as LM_{11}-peptides. All LM_{11}-peptides lost affinity over 170-fold. This suggests that the macrocyclic structure closed by the Ac-S-C_{16} thioether bond in CM_{11}-peptides is crucial to exhibit high binding activity to the HECT domain.

In addition to the above peptides, we synthesized three more mutants of CM_{11}-1 to further validate the importance of the specific structure. One was the aforementioned $C_1 \rightarrow S_1$ mutant peptide, CM_{11}-1S_1, and the others are a macrocyclic peptide without N-methyl backbone but having the same side chains, CP_{11}-1, and its linear peptide, LP_{11}-1 (Table 1). CM_{11}-1S_1 had a 9-fold faster k_{off} rate than CM_{11}-1, implying that C_1 somehow contributes to slowing the dissociation from the target site but is not absolutely essential for binding activity. On the other hand, the other two mutants completely lost binding ability, indicating that *N*-methylated backbone with the N-terminal Ac-C_{16} thioether macrocycle is crucial to maintain the full binding capability of CM_{11}-1 against E6AP HECT domain.

Furthermore, we verified the binding specificity of CM_{11}-peptides against E6AP HECT domain using a different HECT-type E3 Ub ligase, Smurf2 (Ogunjimi et al., 2005). None of CM_{11}-peptides had any SPR signature of binding against Smurf2 Avi-$(His)_6$-GB1-HECT domain immobilized on a SAv-sensor chip, indicating that CM_{11}-peptides have high selectivity toward the E6AP HECT domain over Smurf2 HECT domain nor Avi-$(His)_6$-GB1-tag region (data not shown). Moreover, we investigated human plasma stability of CM_{11}-peptides compared with other control peptides (see more details in Supplemental Information and Figure S3), indicating that a CM_{11}-peptide (CM_{11}-1S_1) is very stable in plasma. Taken together, CM_{11}-peptides have remarkable binding activity and specificity to E6AP HECT domain and plasma stability.

CM_{11}-1 Inhibits Ubiquitination of Target Proteins Catalyzed by E6AP

Despite the observation of strong binding activity of CM_{11}-peptides to E6AP HECT domain, it does not necessarily mean that they are able to inhibit the E6AP ubiquitination activity. To assess whether CM_{11}-peptides have an ability to inhibit the ubiquitination activity, we chose the most active CM_{11}-1 and performed an in vitro assay to monitor ubiquitin thioester formation. It is known that E6AP or even its HECT domain alone forms the ubiquitin thioester intermediate (Beaudenon et al., 2005; Scheffner et al., 1990; Scheffner et al., 1995) in the presence of Ub, E1, and E2 in vitro (Figure 2). When an in vitro translation system from rabbit reticulocyte lysate (RRL), which contains these essential Ub-related components, was used for translating the E6AP HECT domain from the appropriate mRNA template in the presence of [^{35}S]-Met (Huibregtse et al., 1995),

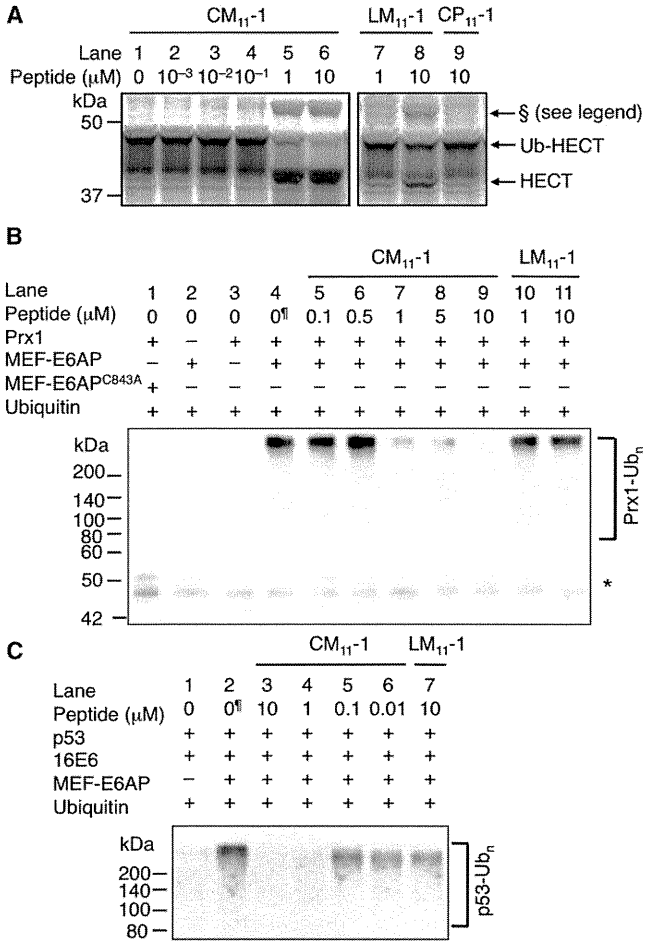


Figure 4. Inhibition of Ub-Thioester Formation with E6AP HECT Domain and Polyubiqutination of Target Proteins by CM_{11}-1 and Its Derivatives

(A) Inhibition of Ub-thioester formation with E6AP HECT domain. ^{35}S-labeled E6AP HECT (43 kDa) was translated in a rabbit reticulocyte lysate cell-free translation system. Because the translation lysate supposedly contained ubiquitin (8 kDa), E1, and E2, the expressed E6AP HECT would be endogenously converted to the ubiquitin-adduct (E6AP HECT-Ub, 51 kDa). The translation mixture was treated with various concentrations of peptides (10^{-3}–10 μM) at room temperature for 30 min. The resulting mixtures were analyzed by SDS-PAGE without DTT. §When the Ub-thioester formation on E6AP was inhibited by CM_{11}-1 (as well as 10 μM LM_{11}-1), a slower migrating band than Ub-E6AP appeared on the gel. Although the product of this band has not yet been defined, this band disappeared upon addition of a free thiol reagent such as DTT (data not shown), suggesting a possibility of dimer formation of HECT domain via a disulfide bond. Alternatively, the free cysteine in HECT domain formed a disulfide bond with other proteins containing a free cysteine residue.

(B) Inhibition of E6-independent polyubiquitination on Prx1. $(His)_6$-Prx1 was incubated with 250 nM MEF-E6AP or inactive mutant MEF-E6AP^{C843A}, $(His)_6$-E1, $(His)_6$-UbcH7, ubiquitin and peptide at 37°C for 30 min. Reaction products were immunoprecipitated with anti-Prx1 pAb and visualized by antiubiquitin mAb immunoblotting. Asterisk denotes immunoglobulin heavy chain. ¶As a negative control, DMSO used as a cosolvent for CM_{11}-1 inhibition was added.

(C) Inhibition of E6-dependent polyubiquitination on p53. p53 was incubated with 170 nM MEF-E6AP, $(His)_6$-E1, $(His)_6$-UbcH7, ubiquitin, HPV16 E6, and peptide at 37°C for 30 min. Reaction products were immunoprecipitated with anti-p53 pAb and visualized by antiubiquitin mAb immunoblotting. See also Figure S4.

a Ub-thioester intermediate of the ^{35}S-Met-labeled HECT domain (Ub-HECT) was distinguished from the parental HECT domain by means of a SDS-PAGE mobility shift assay (Figure 4A, lane 1). When CM_{11}-1 was added to this translation assay system, the migration shift was suppressed at 1 μM or higher concentrations (lanes 5 and 6). As a negative control, LM_{11}-1 and CP_{11}-1 were also tested for the same inhibition assay, in which an approximately 50% inhibition was observed at 10 μM LM_{11}-1, whereas neither 1 μM LM_{11}-1 nor 10 μM CP_{11}-1 exhibited the inhibition. Although the observed potency by CM_{11}-1 seemed not as strong as expected from the SPR data, this could be attributed to that the RRL translation system might contain endogenous label-free E6AP (Huibregtse et al., 1991; Scheffner et al., 1993) that might interact with some fractions of CM_{11}-1, resulting in a reduction of the apparent inhibitory potency. Nevertheless, this result suggests that CM_{11}-1 is able to inhibit the charge of Ub onto the HECT domain of E6AP.

We then have further pursued testing inhibition of E6AP-catalyzed polyubiquitination on target proteins. Peroxiredoxin 1 (Prx1) is an endogenous substrate of E6AP in human cells, and its polyubiquitination occurs independently from the presence of E6 (Nasu et al., 2010). To monitor the inhibitory action of CM_{11}-1 against E6AP, we used an in vitro-reconstituted Prx1 polyubiquitination assay system, in which purified His-tagged Prx1 was incubated with E6AP tagged with MEF (Myc-TEV protease site-flag) and ubiquitin in the presence of purified His-tagged E1 and E2 (UbcH7). The resulting polyubiquitinated Prx1 (Prx1-Ub_n) and free Prx1 were immunoprecipitated by anti-Prx1 polyclonal antibodies and were immunoblotted by an anti-Ub monoclonal antibody to visualize in SDS-PAGE (Figure 4B, lanes 1–9). As negative controls, LM_{11}-1 and CP_{11}-1 were also included in this examination. Clearly, polyubiquitination of Prx1 was inhibited by CM_{11}-1 in a dose-dependent manner (lanes 5–9), where an approximately 1 μM of CM_{11}-1 nearly shut down the E6AP activity. On the other hand, neither LM_{11}-1 (lanes 10 and 11) nor CP_{11}-1 (data not shown) was able to inhibit polyubiquitination. We also tested Prx1-polyubiquitination inhibition by CM_{11}-1S_1, showing a weaker inhibitory activity than CM_{11}-1 (Figure S4); the result seemed consistent with the K_d values for both peptides observed in SPR experiments.

Finally, we examined the inhibitory activity of CM_{11}-1 against polyubiquitination on an E6-dependent substrate, p53, using a reconstituted p53 polyubiquitination assay system. Immunoprecipitation of poly- and nonubiquitinated p53 using anti-p53 pAb followed by immunobolotting using anti-Ub mAb enabled us to detect the polyUb-p53 on SDS-PAGE (Figure 4C). Again, 1 μM CM_{11}-1 was able to inhibit polyubiquitination of p53 in a dose-dependent manner (lanes 3–6), whereas the control peptide, LM_{11}-1, was not (lane 7). The result shows that CM_{11}-1 acts as an E6AP inhibitor that prevents polyubiquitination of Prx1 and p53 in E6-independent and E6-dependent manner, respectively. The trend of K_d values of CM_{11}-1 and its mutant peptides against E6AP HECT domain determined by the SPR experiments well reflected to their observed inhibitory behaviors against ubiquitination of target substrates (Table 1 and Figure 4). Because the present assay method allowed us to detect polyubiquitination instead of monoubiquitination of the substrate proteins catalyzed by an excess amount of E6AP (greater than

two orders of magnitude) over the inherent K_d value of CM_{11}-1, the observed inhibitory potency of CM_{11}-1 was only qualitatively assessed. Most importantly, CM_{11}-1 was capable of inhibiting Ub ligase activity of E6AP even though it was simply selected by binding to E6AP.

SIGNIFICANCE

Here, we have demonstrated RaPID selection of "natural product-like" peptides consisting of thioether-macrocyclic and N-methylated backbone. The selection against E6AP HECT domain has yielded such desired peptides with remarkable binding abilities, falling in a range of K_d values from a subnanomolar to a single-digit nanomolar. One of the representative peptides, CM_{11}-1, chosen for further studies has displayed inhibitory activity against E6AP-catalyzing polyubiquitination on the target proteins, Prx1 and p53. The present work provides the proof-of-technology of RaPID system that enables for the discovery of potent inhibitors against a previously nondruggable ubiquitin ligase, thus opening a wide range of possibilities in the discovery of inhibitors against other ubiquitin ligase families. Most importantly, the natural product-like macrocyclic *N*-methyl-peptides have larger interaction surfaces compared with small organic molecules, as well as elevated stability under physiological conditions compared with ordinary peptides; therefore, they would provide tremendous potentials for the development of drug leads that disrupt not only enzyme activities but also protein-protein interactions.

EXPERIMENTAL PROCEDURES

In Vitro Translation and Selection

Translation of the first round selection was performed using 100 pmol mRNA-puromycin (initial complexity is 6×10^{13}) and 150 μl of translation mixture in the presence of 3750 pmol of ClAc^{D}W-tRNA$^{fMet}_{CAU}$, ^{Me}G-tRNA$^{Asn\text{-}E2}_{GAU}$, ^{Me}A-tRNA$^{Asn\text{-}E2}_{GGC}$, ^{Me}S-tRNA$^{Asn\text{-}E2}_{GAG}$, and ^{Me}F-tRNA$^{Asn\text{-}E2}_{GAA}$ (25 μM each), at 37°C for 30 min. Subsequently, the translation mixture was incubated at room temperature for 12 min to conjugate the translated peptide with the corresponding mRNA-puromycin. This solution was incubation with 15 μl of 200 mM EDTA (pH 8.0) at 37°C for 30 min in order to dissociate ribosomes from mRNA-peptide complexes. For the first-round selection, 11 μl of E6AP HECT immobilized streptavidin magnetic beads (Dynabeads M-280, Invitrogen) was used at a concentration of 200 nM target protein, and mixed with the solution of mRNA-displayed *N*-methyl-peptides. The binding reaction was performed at 4°C for 30 min with rotation. After supernatant was removed, the bead was washed with 300 μl of cold wash buffer (100 mM Tris-HCl [pH 7.5], 300 mM NaCl, 0.05% [v/v] tween 20). To the bead was added 40 μl of RT reaction buffer (50 mM Tris-HCl [pH 8.3], 75 mM KCl, 3 mM $MgCl_2$, 10 μM DTT, 0.5 mM dNTPs, 2 μM CGS3an13.R39) containing 200 units of M-MLV reverse transcriptase (Promega) and 8 units of RNase inhibitor (Promega), and reverse transcribed at 42°C for 60 min with rotation. The selected cDNA was eluted with 360 μl of PCR buffer (10 mM Tris-HCl [pH 7.5], 50 mM KCl, 0.1% [v/v] Triton X-100, 2.5 mM $MgCl_2$, 0.25 mM dNTPs, 0.25 μM T7g10M.F48, 0.25 μM CGS3an13.R39) at 95°C for 5 min. After addition of *Taq* DNA polymerase to the eluate, the mixture was used for PCR amplification. The amplified DNA was purified by the extraction with phenol/chloroform and ethanol purification and was used for the next round of selection. Since the second round, 10 μl scale of transcription and 7.5 μl of ligation with puromycin linker were carried out. The resulting mRNA-puromycin of the second round was translated using 5 μl of the translation mixture in the presence of each 25 μM acyl-tRNAs at 37°C for 30 min, followed by incubation at room temperature for 12 min. After incubation with 1 μl of 100 mM EDTA (pH 8.0) at 37°C for 30 min, the reverse transcription of the mRNA-displayed peptides was performed by RT reaction buffer in the presence of M-MLV reverse transcriptase without RNase H activity (Promega), at 42°C for 60 min with rotation. After quenching the reaction with 1 μl of 100 mM EDTA (pH 8.0), the solution was neutralized with 1.1 μl of 0.2 M HCl. The complexes with cDNA- and mRNA-displayed peptides were subjected to 2.4 μl of the magnetic bead without target and were incubated at 4°C for 30 min for negative selection at once. Subsequently, the supernatant was mixed with 0.8 μl of the magnetic bead with E6AP HECT and was incubated with 4°C for 30 min for positive selection, followed by thrice washing with 10 μl of cold wash buffer. After addition of 100 μl of PCR buffer to the bead, the cDNA were eluted at 95°C for 5 min and amplified by *Taq* DNA polymerase. In the third round and all subsequent rounds, the all experiments were performed by a half of the reaction scale of the second round. Moreover, negative selection was performed at 4°C for 20 min at three times in the third and fourth round, and at nine times in the fifth and sixth round. On the other hand, positive selections after second round were performed by mixing with the complexes of cDNA and mRNA-displayed peptides and 200 nM E6AP HECT (not immobilized on streptavidin magnetic bead) at 37°C for 30 min, followed by pull down by streptavidin magnetic bead at 37°C for 5 min with rotation. After thrice washing with 5 μl of wash buffer at room temperature, the selected cDNA was eluted with 100 μl of PCR buffer at 95°C for 5 min. After addition of *Taq* DNA polymerase to the eluate, the mixture was used for PCR amplification. To monitor the convergence of the selection process, real-time PCR (RT-PCR) was used to quantify the amounts of input and output DNA in every round. For input cDNA, 0.25 μl aliquot of the RT mixture was diluted with 150 μl of a dilution solution (10 mM Tris-HCl [pH 8.0] and 300 mM NaCl), and 1 μl of the diluted cDNA was mixed with 19 μl of PCR buffer containing SYBR Green I (Molecular Probe) and *Taq* DNA polymerase. For output cDNA, 1 μl aliquots of the eluates from the beads of positive and negative selections were mixed with 10 μl of PCR buffer containing SYBR Green I and *Taq* DNA polymerase. The reverse transcribed $(NNU)_{10}$ mRNA mixture was serial-diluted and used for the templates as standards.

MALDI-TOF Analysis of Translated Clone Peptides

To identify the expressed cyclic *N*-methylated peptides, a 5 μl scale translation reaction was performed using FIT system with 40 nM of clone DNA, 25 μM each of ClAc^{D}W-tRNA$^{fMet}_{CAU}$, ^{Me}G-tRNA$^{Asn\text{-}E2}_{GAU}$, ^{Me}A-tRNA$^{Asn\text{-}E2}_{GGC}$, ^{Me}S-tRNA$^{Asn\text{-}E2}_{GAG}$, and ^{Me}F-tRNA$^{Asn\text{-}E2}_{GAA}$ at 37°C for 30 min. After quenching with 0.2% TFA, the crude peptide mixture was desalted with C-Tip (C18 desalting SPE, Nikkyo technos) and eluted with 80% acetonitrile and 0.5% acetic acid solution saturated with the matrix α-cyano-4-hydroxycinnamic acid (Bruker Daltonics). MALDI-TOF analysis was performed using an Autoflex TOF/TOF (Bruker Daltonics) and peptide calibration standard II (Bruker Daltonics) as external standards.

SPR Analysis of Peptides

The interaction between E6AP HECT and peptides was assessed using a BIACORE T100 instrument (GE Healthcare) equipped with research-grade streptavidin sensor chip at 25°C. Biotinylated E6AP HECT was immobilized to a surface density of approximate 1,500 response units (RU) using standard immobilization protocols (GE Healthcare). HBS-EP+ (10 mM HEPES [pH 7.4], 150 mM NaCl, 3 mM EDTA, and 0.05% [v/v] surfactant P20) containing 1.0% (v/v) DMSO was used as the running buffer for all experiments. Peptide binding was tested by injecting varying concentrations (0.3 nM to 1,000 nM) at a flow rate of 30 μl min^{-1} and measured by single cycle kinetics method. Raw data were analyzed by the BIACORE T100 evaluation software 2.01 and fitted to the standard 1:1 interaction model.

In Vitro Ubiquitin Transfer Assay by ^{35}S-Labeled E6AP HECT

E6AP HECT cDNA was subcloned into pTNT vector (Promega) at XhoI and SalI sites. ^{35}S-Methionine-labeled E6AP HECT was synthesized in vitro by TNT T7 coupled rabbit reticulocyte lysate system (Promega) at 30°C for 90 min, by following the standard procedure (Huibregtse et al., 1995). After translation, 0.5 μl of peptides in 10% (v/v) DMSO with 10× concentrations shown in Figure 4A were added to the translation reaction mixture (4.5 μL). The resulting mixture was incubated at room temperature at 30 min and

quenched with 2 × SDS-polyacrylamide gel laemli sample buffer (125 mM Tris-HCl [pH 6.8], SDS 4%, glycerol 20%, 0.002% bromophenol blue) in the absence of dithiothreitol. Samples were subjected to SDS-PAGE on a 10% polyacrylamide gel.

In Vitro Polyubiquitination Assay for Prx1 and p53

The plasmid pGEM p53 was used for in vitro translation (Werness et al., 1990). In vitro translation was performed using TNT T7 coupled rabbit reticulocyte lysate system. Recombinant baculovirus for HPV16 E6 was produced using the BaculoGold system (PharMingen) as described previously (Shirakura et al., 2007). Hi5 cells (Invitrogen) were infected with the recombinant baculovirus to produce HPV16 E6 protein. HPV16 E6 Protein was partially purified by anion-exchange chromatography as previously described (Huibregtse et al., 1993). In vitro polyubiquitination assays for Prx1 were performed essentially as described previously (Nasu et al., 2010). Hi5 cells were infected with recombinant baculoviruses AcMEF-E6AP and Ac MEF-E6AP^{C843A} to produce MEF-E6AP and MEF-E6AP^{C843A}, respectively (Shirakura et al., 2007). MEF-E6AP and MEF-E6AP^{C843A} were purified on anti-FLAG M2 agarose beads (Sigma) according to the manufacturer's instructions. Assays were done in 40 μl volumes containing 20 mM Tris (pH 7.6), 50 mM NaCl, 5 mM $MgCl_2$, 100 μM DTT, 5 mM ATP, 250 nM MEF-E6AP or inactive mutant MEF-E6AP^{C843A}, 62.5 nM $(His)_6$-E1, 1.1 μM $(His)_6$-UbcH7, 25 μM ubiquitin, 8 μM $(His)_6$-Prx1 and peptide, and incubation at 37°C for 30 min. Reactions were performed at 37°C for 30 min. The ubiquitination reaction was terminated by freezing the samples with liquid nitrogen. To dissociate proteins, 1% SDS was added to lysates, which were then heated at 90°C for 15 min. The samples were diluted 10-fold with a dissociation dilution buffer containing 1% NP-40, 0.5% deoxycholate, 120 mM NaCl, 50 mM HEPES, 1 mM EDTA, and complete protease inhibitor cocktail (Roche). Samples were immunoprecipitated with anti-Prx1 PAb and analysis by immunoblotting with antiubiquitin mouse monoclonal antibody (anti-Ubi-1, Millipore) to detect ubiquitinated Prx1. In vitro polyubiquitination assays for p53 were performed essentially as described previously (Nakagawa and Huibregtse, 2000). Assays were done in 75 μl volumes containing 25 mM Tris-HCl (pH 8.0), 125 mM NaCl, 2 mM $MgCl_2$, 50 μM DTT, 5 μM ubiquitin, 2 mM ATP, 170 nM MEF-E6AP, 33 nM $(His)_6$-E1, 0.6 μM $(His)_6$-UbcH7, 2 μl of partially purified HPV16 E6, and 5 μl of in vitro translated p53. Peptides inhibitors were added to the samples as indicated. The reaction mixtures were incubated at 37°C for 30 min. The ubiquitination reaction was terminated by freezing the samples with liquid nitrogen. To dissociate proteins, 1% SDS was added to lysates, which were then heated and diluted as described above. Samples were immunoprecipitated with anti-p53 rabbit polyclonal antibody (FL393, Santa Cruz), followed by immunoblotting with antiubiquitin mouse monoclonal antibody (anti-Ubi-1, Millipore) to detect ubiquitinated p53.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, one table, and supplemental Experimental Procedures and may be found with this article online at doi:10.1016/j.chembiol.2011.09.013.

ACKNOWLEDGMENTS

We thank Dr. Hiroshi Murakami for the discussion on conducting this work and Dr. Naoki Goshima for providing the cDNAs of E6AP and Smurf2. This work was supported by the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for the Specially Promoted Research (Grant 21000005), Research and Development Projects of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (support to H.S.), Grants-in-Aid for JSPS Fellows (Grant 7734 to Y.Y.), JSPS Grants-in-Aid for Young Scientists (Grant B22710210 to T. Katoh and B22750145 to Y.G.), and Grants-in-Aid for Scientific Research from the Ministry of Health, Labour, and Welfare of Japan (to I.S. and H.S.).

Received: July 13, 2011
Revised: September 20, 2011
Accepted: September 20, 2011
Published: December 22, 2011

REFERENCES

Beaudenon, S., and Huibregtse, J.M. (2008). HPV E6, E6AP and cervical cancer. BMC Biochem. *9* (*Suppl 1*), S4.

Beaudenon, S., Dastur, A., and Huibregtse, J.M. (2005). Expression and assay of HECT domain ligases. Methods Enzymol. *398*, 112–125.

Biron, E., Chatterjee, J., Ovadia, O., Langenegger, D., Brueggen, J., Hoyer, D., Schmid, H.A., Jelinek, R., Gilon, C., Hoffman, A., and Kessler, H. (2008). Improving oral bioavailability of peptides by multiple N-methylation: somatostatin analogues. Angew. Chem. Int. Ed. Engl. *47*, 2595–2599.

Chatterjee, C., Paul, M., Xie, L., and van der Donk, W.A. (2005). Biosynthesis and mode of action of lantibiotics. Chem. Rev. *105*, 633–684.

Chatterjee, J., Gilon, C., Hoffman, A., and Kessler, H. (2008). N-methylation of peptides: a new perspective in medicinal chemistry. Acc. Chem. Res. *41*, 1331–1342.

Doedens, L., Opperer, F., Cai, M., Beck, J.G., Dedek, M., Palmer, E., Hruby, V.J., and Kessler, H. (2010). Multiple N-methylation of MT-II backbone amide bonds leads to melanocortin receptor subtype hMC1R selectivity: pharmacological and conformational studies. J. Am. Chem. Soc. *132*, 8115–8128.

Driggers, E.M., Hale, S.P., Lee, J., and Terrett, N.K. (2008). The exploration of macrocycles for drug discovery—an underexploited structural class. Nat. Rev. Drug Discov. *7*, 608–624.

Eldridge, A.G., and O'Brien, T. (2010). Therapeutic strategies within the ubiquitin proteasome system. Cell Death Differ. *17*, 4–13.

Eletr, Z.M., and Kuhlman, B. (2007). Sequence determinants of E2-E6AP binding affinity and specificity. J. Mol. Biol. *369*, 419–428.

Forster, A.C., Tan, Z., Nalam, M.N., Lin, H., Qu, H., Cornish, V.W., and Blacklow, S.C. (2003). Programming peptidomimetic syntheses by translating genetic codes designed *de novo*. Proc. Natl. Acad. Sci. USA *100*, 6353–6357.

Goto, Y., Ohta, A., Sako, Y., Yamagishi, Y., Murakami, H., and Suga, H. (2008). Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. ACS Chem. Biol. *3*, 120–129.

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

Grünewald, J., and Marahiel, M.A. (2006). Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides. Microbiol. Mol. Biol. Rev. *70*, 121–146.

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. Annu. Rev. Biochem. *67*, 425–479.

Hoeller, D., and Dikic, I. (2009). Targeting the ubiquitin system in cancer therapy. Nature *458*, 438–444.

Hoeller, D., Hecker, C.M., and Dikic, I. (2006). Ubiquitin and ubiquitin-like proteins in cancer pathogenesis. Nat. Rev. Cancer *6*, 776–788.

Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P.M., Huibregtse, J.M., and Pavletich, N.P. (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. Science *286*, 1321–1326.

Huibregtse, J.M., Scheffner, M., and Howley, P.M. (1991). A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. EMBO J. *10*, 4129–4135.

Huibregtse, J.M., Scheffner, M., and Howley, P.M. (1993). Localization of the E6-AP regions that direct human papillomavirus E6 binding, association with p53, and ubiquitination of associated proteins. Mol. Cell. Biol. *13*, 4918–4927.

Huibregtse, J.M., Scheffner, M., Beaudenon, S., and Howley, P.M. (1995). A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. Proc. Natl. Acad. Sci. USA *92*, 2563–2567.

Josephson, K., Hartman, M.C., and Szostak, J.W. (2005). Ribosomal synthesis of unnatural peptides. J. Am. Chem. Soc. *127*, 11727–11735.

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

Kohli, R.M., Walsh, C.T., and Burkart, M.D. (2002). Biomimetic synthesis and optimization of cyclic peptide antibiotics. Nature *418*, 658–661.

Kumar, S., Talis, A.L., and Howley, P.M. (1999). Identification of HHR23A as a substrate for E6-associated protein-mediated ubiquitination. J. Biol. Chem. *274*, 18785–18792.

Kung, H.F., Redfield, B., Treadwell, B.V., Eskin, B., Spears, C., and Weissbach, H. (1977). DNA-directed in vitro synthesis of beta-galactosidase: studies with purified factors. J. Biol. Chem. *252*, 6889–6894.

Layfield, R., Lowe, J., and Bedford, L. (2005). The ubiquitin-proteasome system and neurodegenerative disorders. Essays Biochem. *41*, 157–171.

Li, P., and Roller, P.P. (2002). Cyclization strategies in peptide derived drug design. Curr. Top. Med. Chem. *2*, 325–341.

Liu, Y., Cherry, J.J., Dineen, J.V., Androphy, E.J., and Baleja, J.D. (2009). Determinants of stability for the E6 protein of papillomavirus type 16. J. Mol. Biol. *386*, 1123–1137.

Louria-Hayon, I., Alsheich-Bartok, O., Levav-Cohen, Y., Silberman, I., Berger, M., Grossman, T., Matentzoglu, K., Jiang, Y.H., Muller, S., Scheffner, M., et al. (2009). E6AP promotes the degradation of the PML tumor suppressor. Cell Death Differ. *16*, 1156–1166.

McIntosh, J.A., Donia, M.S., and Schmidt, E.W. (2009). Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. Nat. Prod. Rep. *26*, 537–559.

Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006). A highly flexible tRNA acylation method for non-natural polypeptide synthesis. Nat. Methods *3*, 357–359.

Nakagawa, S., and Huibregtse, J.M. (2000). Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. Mol. Cell. Biol. *20*, 8244–8253.

Nalepa, G., Rolfe, M., and Harper, J.W. (2006). Drug discovery in the ubiquitin-proteasome system. Nat. Rev. Drug Discov. *5*, 596–613.

Nasu, J., Murakami, K., Miyagawa, S., Yamashita, R., Ichimura, T., Wakita, T., Hotta, H., Miyamura, T., Suzuki, T., Satoh, T., and Shoji, I. (2010). E6AP ubiquitin ligase mediates ubiquitin-dependent degradation of peroxiredoxin 1. J. Cell. Biochem. *111*, 676–685.

Nemoto, N., Miyamoto-Sato, E., Husimi, Y., and Yanagawa, H. (1997). In vitro virus: bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome *in vitro*. FEBS Lett. *414*, 405–408.

Nestor, J.J., Jr. (2009). The medicinal chemistry of peptides. Curr. Med. Chem. *16*, 4399–4418.

Ogunjimi, A.A., Briant, D.J., Pece-Barbara, N., Le Roy, C., Di Guglielmo, G.M., Kavsak, P., Rasmussen, R.K., Seet, B.T., Sicheri, F., and Wrana, J.L. (2005). Regulation of Smurf2 ubiquitin ligase activity by anchoring the E2 to the HECT domain. Mol. Cell *19*, 297–308.

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

Ohta, A., Yamagishi, Y., and Suga, H. (2008). Synthesis of biopolymers using genetic code reprogramming. Curr. Opin. Chem. Biol. *12*, 159–167.

Ohuchi, M., Murakami, H., and Suga, H. (2007). The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. Curr. Opin. Chem. Biol. *11*, 537–542.

Oman, T.J., and van der Donk, W.A. (2010). Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat. Chem. Biol. *6*, 9–18.

Pickart, C.M. (2001). Mechanisms underlying ubiquitination. Annu. Rev. Biochem. *70*, 503–533.

Roberts, R.W., and Szostak, J.W. (1997). RNA-peptide fusions for the *in vitro* selection of peptides and proteins. Proc. Natl. Acad. Sci. USA *94*, 12297–12302.

Rubinsztein, D.C. (2006). The roles of intracellular protein-degradation pathways in neurodegeneration. Nature *443*, 780–786.

Sagan, S., Karoyan, P., Lequin, O., Chassaing, G., and Lavielle, S. (2004). N- and Calpha-methylation in biologically active peptides: synthesis, structural and functional aspects. Curr. Med. Chem. *11*, 2799–2822.

Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J., and Howley, P.M. (1990). The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell *63*, 1129–1136.

Scheffner, M., Huibregtse, J.M., Vierstra, R.D., and Howley, P.M. (1993). The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell *75*, 495–505.

Scheffner, M., Nuber, U., and Huibregtse, J.M. (1995). Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. Nature *373*, 81–83.

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

Shirakura, M., Murakami, K., Ichimura, T., Suzuki, R., Shimoji, T., Fukuda, K., Abe, K., Sato, S., Fukasawa, M., Yamakawa, Y., et al. (2007). E6AP ubiquitin ligase mediates ubiquitylation and degradation of hepatitis C virus core protein. J. Virol. *81*, 1174–1185.

Werness, B.A., Levine, A.J., and Howley, P.M. (1990). Association of human papillomavirus types 16 and 18 E6 proteins with p53. Science *248*, 76–79.

Zhou, P., Lugovskoy, A.A., and Wagner, G.A. (2001). A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. J. Biomol. NMR *20*, 11–14.

View Online / Journal Homepage / Table of Contents for this issue
Dynamic Article Links

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

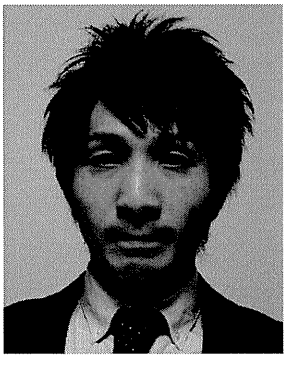
A wealth of knowledge has been accumulated on ribosomal synthesis of macrocyclic peptides in the past decade. In nature, backbone cyclization of the translated linear peptides is generally catalyzed by specific enzymes, giving them peptidase resistance, thermodynamic stability and various other physiological 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*

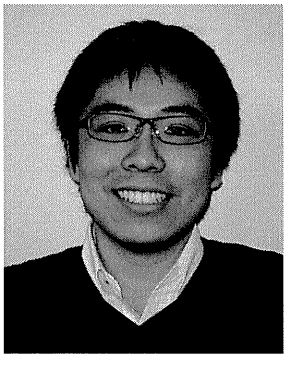
Introduction

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.

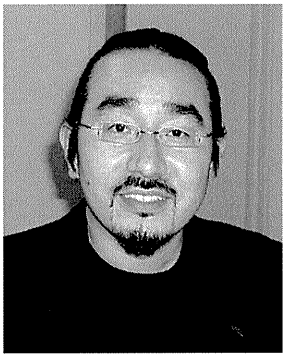
While nature uses diverse kinds of chemical strategies for macrocyclization, here we will focus our discussions on "backbone macrocyclization". The backbone macrocyclization seen in naturally occurring peptides is generally employed by specific enzymes that modify the precursor linear peptides to ligate 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.

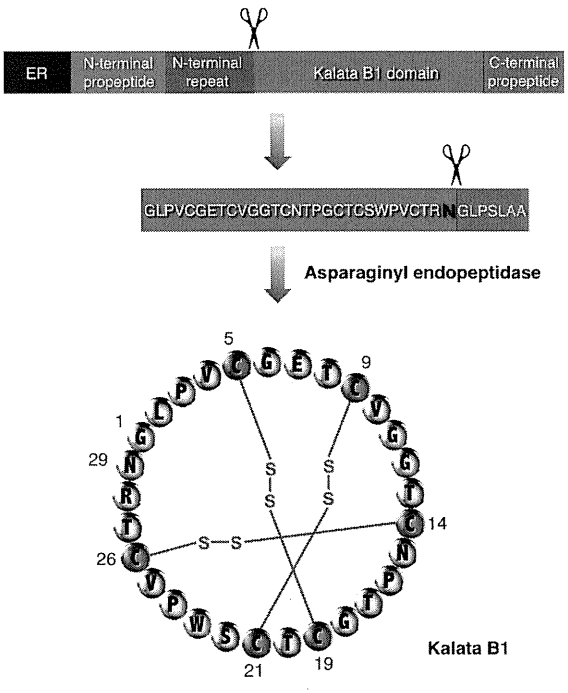


Fig. 1 A proposed mechanism for the cyclization of Kalata B1. The N-terminal flanking sequences including ER signal, N-terminal propeptide and N-terminal repeat are first excised from the precursor by as-yet unidentified enzyme(s). Then, the cleavage of the C-terminal propeptide and macrocyclization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

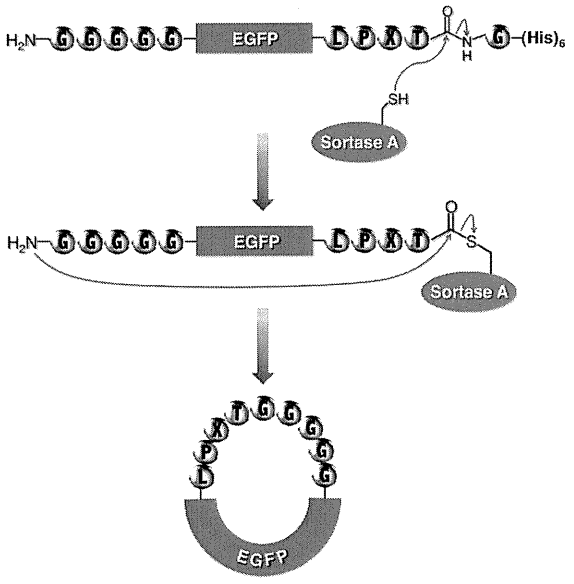


Fig. 2 Macrocyclization of polypeptides (proteins) by utilizing the Sortase-mediated ligation method. EGFP having a N-terminal oligoglycine tag and C-terminal LPXTG tag can be cyclized by a single transpeptidase Sortase A (SrtA).

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

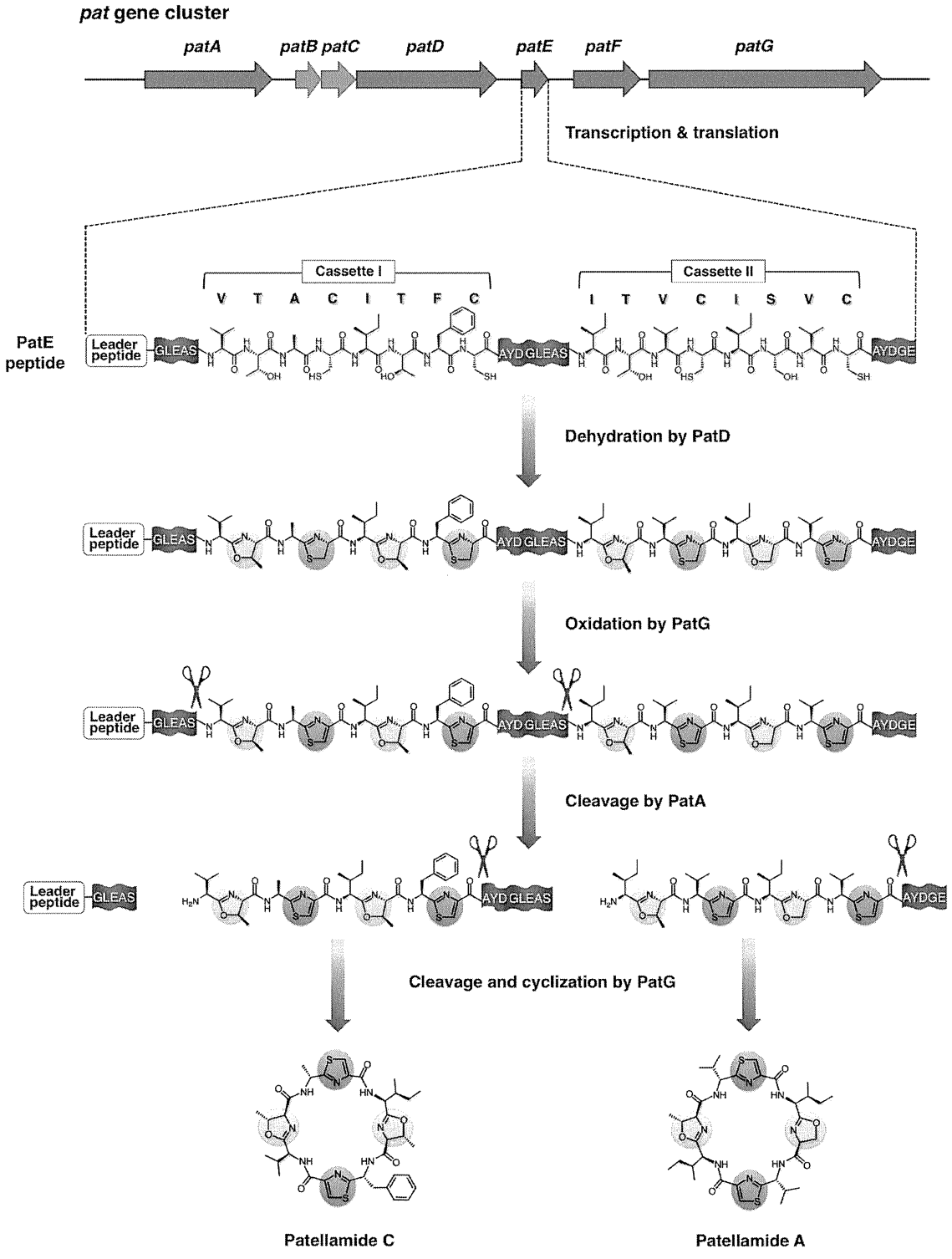


Fig. 3 A model for the biosynthesis of patellamides C and A. Genes indicated by the red arrows are essential for patellamide biosynthesis, whereas the blue arrows indicate unnecessary genes. The precursor peptide PatE contains two core peptide regions (cassettes I and II), both of which have flanking 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]

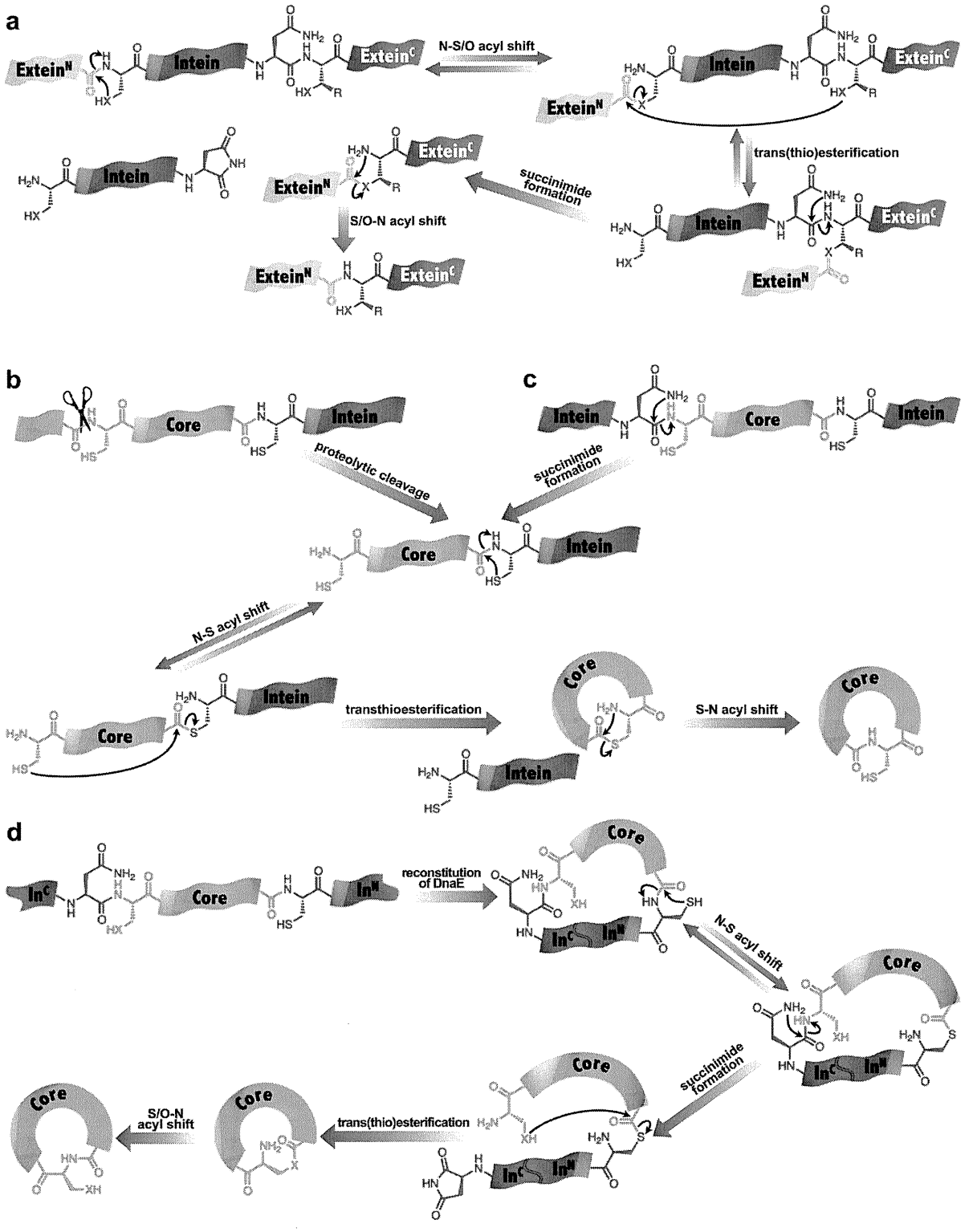


Fig. 4 Mechanism of intein-mediated protein splicing and its applications for the synthesis of backbone cyclic peptides. (a) Protein splicing by inteins. X and R represent S/O and H/Me, respectively, in the side chain of serine, threonine, or cysteine. (b, c) Intein-based macrocyclization of peptides *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

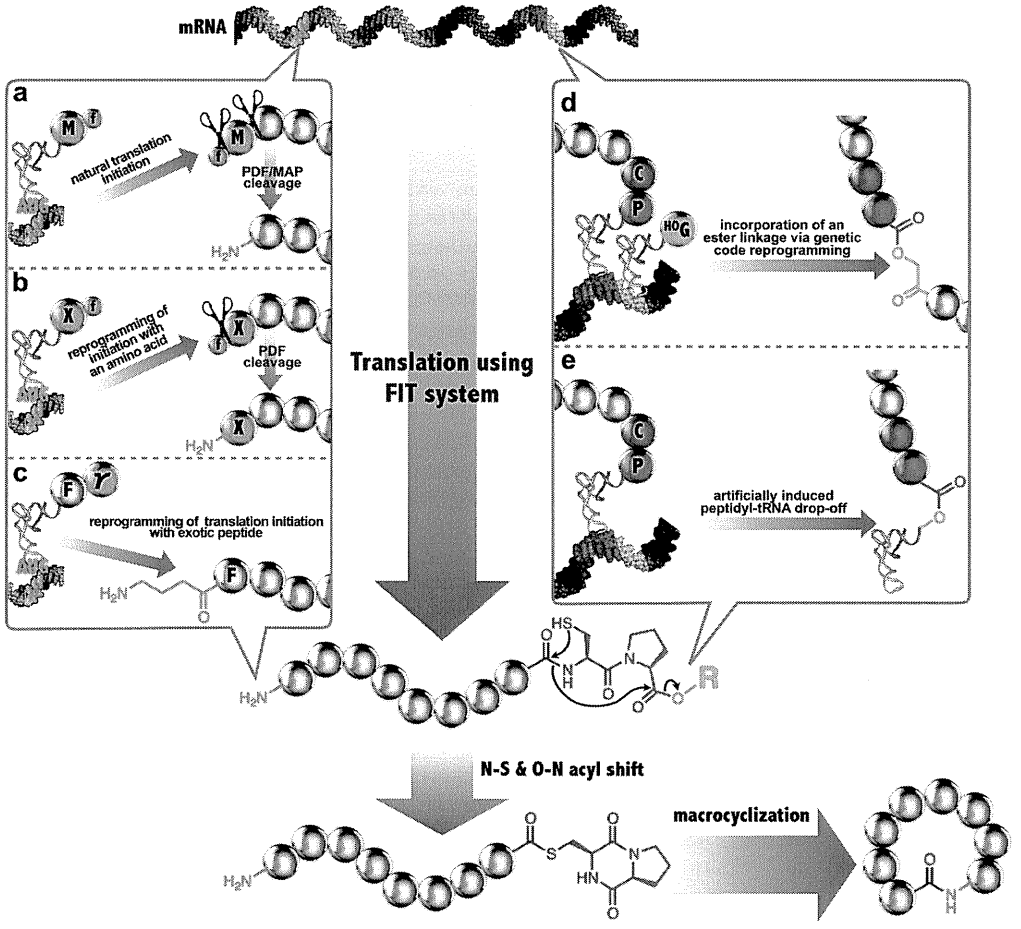


Fig. 5 Ribosomal synthesis of backbone cyclic peptides using the FIT system in which the genetic code is artificially reprogrammed. (a, b, c) Methods for the preparation of linear peptides bearing the N-terminal free amino group. The N-terminal free amino group is produced in the expressed peptides by (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

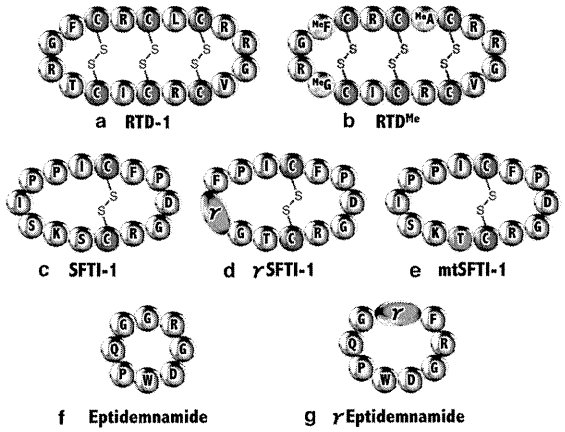


Fig. 6 Examples of backbone cyclic peptides expressed in the FIT system. (a) Wild type RTD-1 and (b) RTD-1 containing N-methyl amino acids (RTD^{Me}). (c) Wild type SFTI-1, (d) SFTI-1 containing aγ-amino acid(γSFTI-1) and (e) a mutant SFTI-1 involving a single mutation of position 2 from threonine to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

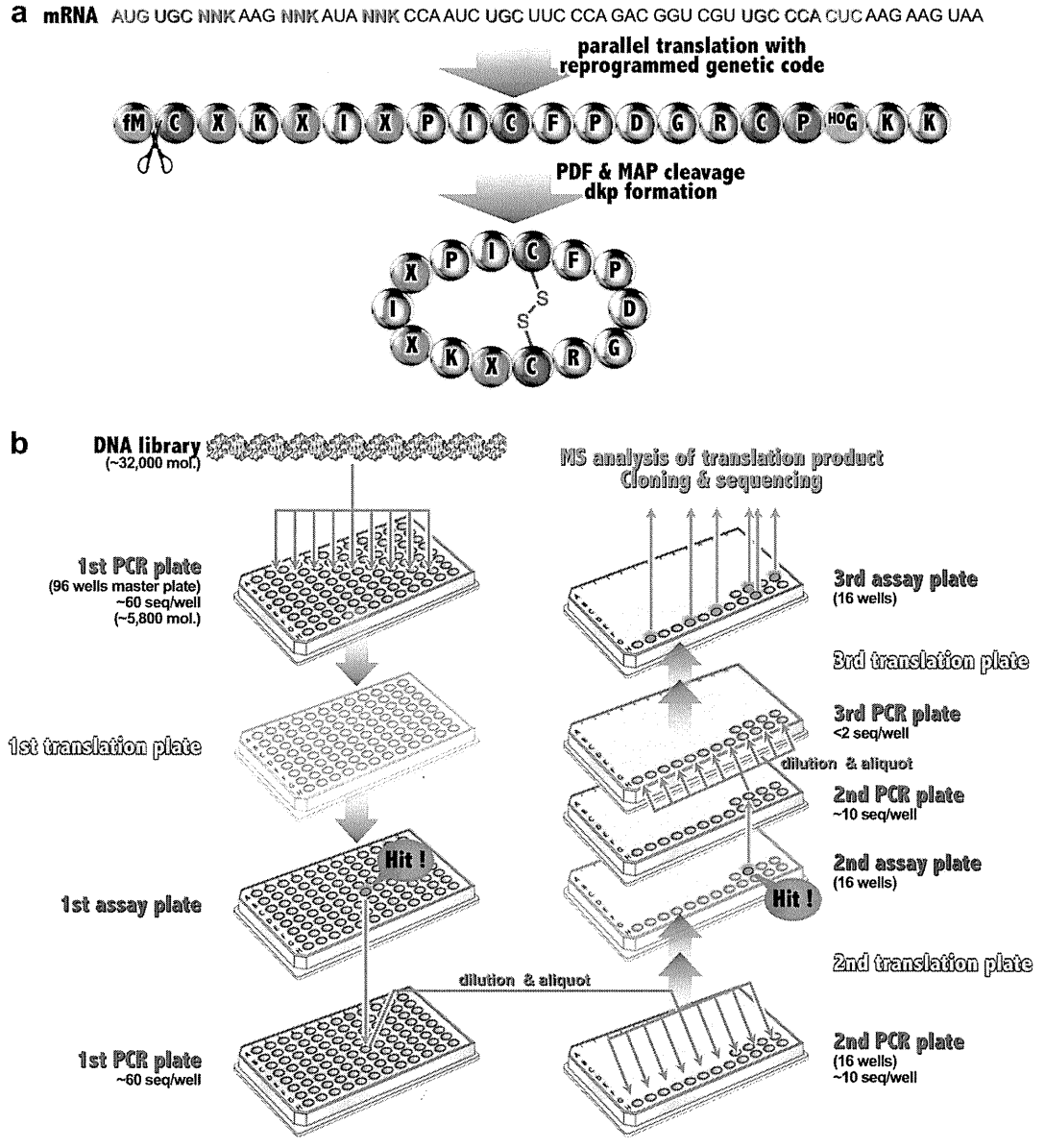


Fig. 7 Construction of a backbone cyclic peptide library in the FIT system and rapid screening of functional species. (a) Sequences of the mRNA library expressing the precursors of the SFTI-1 mutant library. ^{HO}G is assigned to the CUC codon by genetic code reprogramming. X denotes random amino acid 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.

ACCOUNTS
of chemical research

Flexizymes: Their Evolutionary History and the Origin of Catalytic Function

JUMPEI MORIMOTO,[†,‡] YUUKI HAYASHI,[†,§]
KAZUHIRO IWASAKI,[†,‡] AND HIROAKI SUGA*[,†,‡,§]

[†]Department of Chemistry, School of Science, [‡]Department of Chemistry and Biotechnology, School of Engineering, and [§]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan

RECEIVED ON MARCH 28, 2011

CONSPECTUS

Transfer RNA (tRNA) is an essential component of the cell's translation apparatus. These RNA strands contain the anticodon for a given amino acid, and when "charged" with that amino acid are termed aminoacyl-tRNA. Aminoacylation, which occurs exclusively at one of the 3′-terminal hydroxyl groups of tRNA, is catalyzed by a family of enzymes called aminoacyl-tRNA synthetases (ARSs). In a primitive translation system, before the advent of sophisticated protein-based enzymes, this chemical event could conceivably have been catalyzed solely by RNA enzymes. Given the evolutionary implications, our group attempted in vitro selection of artificial ARS-like ribozymes, successfully uncovering a functional ribozyme (r24) from an RNA pool of random sequences attached to the 5′-leader region of tRNA. This ribozyme preferentially charges aromatic amino acids (such as phenylalanine) activated with cyanomethyl ester (CME) onto specific kinds of tRNA.

During the course of our studies, we became interested in developing a versatile, rather than a specific, aminoacylation catalyst. Such a ribozyme could facilitate the preparation of intentionally misacylated tRNAs and thus serve a convenient tool for manipulating the genetic code. On the basis of biochemical studies of r24, we constructed a truncated version of r24 (r24mini) that was 57 nucleotides long. This r24mini was then further shortened to 45 nucleotides. This ribozyme could charge various tRNAs through very simple three-base-pair interactions between the ribozyme's 3′-end and the tRNA's 3′-end. We termed this ribozyme a "flexizyme" (Fx3 for this particular construct) owing to its flexibility in addressing tRNAs.

To devise an even more flexible tool for tRNA acylation, we attempted to eliminate the amino acid specificity from Fx3. This attempt yielded an Fx3 variant, termed dFx, which accepts amino acid substrates having 3,5-dinitrobenzyl ester instead of CME as a leaving group. Similar selection attempts with the original phenylalanine-CME and a substrate activated by (2-aminoethyl)amidocarboxybenzyl thioester yielded the variants eFx and aFx (e and a denote enhanced and amino, respectively). In this Account, we describe the history and development of these flexizymes and their appropriate substrates, which provide a versatile and easy-to-use tRNA acylation system. Their use permits the synthesis of a wide array of acyl-tRNAs charged with artificial amino and hydroxy acids.

In parallel to these efforts, we initiated a crystallization study of Fx3 covalently conjugated to a microhelix RNA, which is an analogue of tRNA. The X-ray crystal structure, solved as a co-complex with phenylalanine ethyl ester and U1A-binding protein, revealed the structural basis of this enzyme. Most importantly, many biochemical observations were consistent with the crystal structure. Along with the predicted three regular-helix regions, however, the flexizyme has a unique irregular helix that was unexpected. This irregular helix constitutes a recognition pocket for the aromatic ring of the amino acid side chain and precisely brings the carbonyl group to the 3′-hydroxyl group of the tRNA 3′-end. This study has clearly defined the molecular interactions between Fx3, tRNA, and the amino acid substrate, revealing the fundamental basis of this unique catalytic system.

History of Aminoacylation Ribozymes

In the "modern world" translation system, aminoacylation of the 3′-terminus of tRNA is catalyzed by a family of protein enzymes, aminoacyl-tRNA synthetases (ARSs). On the other hand, the present evidence from the crystal structural as well as biochemical studies of ribosome have revealed that

Published on the Web 06/28/2011 www.pubs.acs.org/accounts
10.1021/ar2000953

ribosome's catalytic center consists of only RNA, that is, ribosome is a ribozyme. This suggests that a primitive translation catalytic system, including not only ribosome but also ARSs, could have consisted of entirely RNA molecules. However, naturally occurring ARS ribozymes are yet unknown and thus the above hypothesis is not fully supported by the available knowledge from nature. Because sophisticated protein enzymes such as ARSs could not be evolved before the advent of the translation system, it is critical to see if RNA molecules are able to catalyze aminoacylation.

In vitro selection (or SELEX) is a powerful technique that aims at isolating functional RNAs from a pool of random sequences of RNA. In fact, in the last nearly two decades, we have witnessed the discovery of many artificial ribozymes capable of catalyzing various chemical reactions in vitro even though such naturally occurring ribozymes are unknown in the modern life. Among them, a few research teams have successfully isolated ribozymes capable of charging certain amino acids onto RNA. One of the earlier pioneer works was reported by Yarus et al. where they isolated artificial ribozymes that catalyze self-aminoacylation of their own CCG 3′-terminal 2′/3′-OH with Phe-AMP or Tyr-AMP[1,2] as an aminoacylation donor. In 1998, Famulok and Jenne also reported an artificial rybozyme that catalyzes self-aminoacylation with *N*-biotin-Phe-AMP, but the site of aminoacylation is not 3′-terminal 2′/3′-OH but an internal 2′-OH.[3] Both classes of ribozymes were unable to aminoacylate onto the 3′-end of tRNA or tRNA-like CCA, and therefore, a major question remained unanswered whether ribozymes are able to charge amino acids onto the specific site of 3′-terminal CCA end of tRNA or tRNA molecules.

Meanwhile, Szostak and Lohse reported an acyl-transferase ribozyme (ATRib) capable of transferring *N*-biotin-Phe (biotin is a selectable tag using streptavidin resin) from the 3′-end of a short RNA to its own 5′-hydroxyl group.[4] Because this ribozyme was originally aimed at mimicking the function of ribosome as a peptidyl transfer catalyst, the acyl-donor RNA was designed to be 5′-CAACCA-3′ as a mimic of tRNA's 3′-end and the 3′-end of ATRib has an internal guide sequence complementary to this RNA sequence. Using the microreversibility of acyl-transfer reaction, Suga et al. turned ATRib into two new ribozymes that are able to charge amino acids onto the 3′-terminus of tRNAs. A ribozyme, reported in 2000 and referred to as AD02,[5] has a 70 nucleotide (nt) accessory domain at the 3′-terminus of ATRib and catalyzes two steps of reactions; in the first step, AD02's accessory domain catalyzes self-aminoacylation of *N*-biotin-Gln assisted by cyanomethyl ester (CME) onto the 5′-OH group, and in the second step the 5′-acyl group is transferred to the 3′-end of tRNA upon binding to the ATRib internal guide sequence. Although the catalytic ability of AD02 is very modest where only 4% of tRNA is aminoacylated, this ribozyme showed remarkable specificity toward Gln side chain over Met (0.047), Leu (0.004), Phe (<0.001), and Val (<0.001) (the values in parenthese are relative specificity constants).[6] This work represents the first example of a tRNA aminoacylation catalyst consisting of only a RNA scaffold. The second ribozyme, reported in 2002 and referred to as BC28,[7] also has an accessory domain at the 3′-terminus of ATRib and a loop, referred to as anti-anticodon (AC) loop, embedded in this domain, and recognizes a specific AC loop sequence of tRNA by forming six base pairs. BC28 accepts various kinds of amino acids from the 3′-end of 5′-AACCA-3′ (note that one complementary base less than the original donor substrate) to its own 5′-terminal OH group such as ATRib, and also transfers back to the 3′-end of a specific tRNA designated by the interaction between the anti-AC loop and tRNA's AC loop, yielding up to 17% aminoacyl-tRNA. Interestingly, a base mutation(s) in the anti-AC loop is able to reprogram the specificity of BC28 that aminoacylates a desired tRNA with a complete match of six base pairs over other tRNAs containing a mispair(s). Although these ribozymes consisting of the ATRib scaffold endorse some of unique characteristic functions as ARSs, their modest efficiencies of tRNA aminoacylation due to their complex mechanisms involving the equilibrium shift of acyl-transfer chemistry dismissed further development of these ribozymes.

In Vitro Evolution of Flexizyme

In prokaryotes, precursor tRNAs, in which 5′-leader accessory sequences are attached to the 5′-end of the body of tRNA, are transcribed and processed by ribonucleoprotein enzymes, RNase P, to yield mature tRNAs onto which amino acids are charged by cognate ARSs (Figure 1A). The catalytic component of RNase P is known to be RNA (referred to as M1 RNA), representing one of the molecular fossils of the RNA world. One can hypothesize that even though the 5′-leader sequences have no function in the "modern world", they could have catalytic ability for self-aminoacylation of the 3′-terminus of tRNA sequences and later be removed by a M1 RNA-like ribozyme in the "RNA world" (Figure 1B). This hypothesis is also able to give a simple explanation on how such 5′-leader ribozymes could specify cognate tRNAs, that is, they are covalently linked to cognate tRNAs for self-aminoacylation, and thus the specificity is set in the respective precursor tRNAs. We thus decided to set an experiment