ChemComm

Cite this: *Chem. Commun.*, 2011, **47**, 9116–9118

www.rsc.org/chemcomm

COMMUNICATION

Synthesis of a cyclic peptide/protein using the NEXT-A reaction followed by cyclization†

Toshimasa Hamamoto, Masahiko Sisido, Takashi Ohtsuki and Masumi Taki*

Received 16th April 2011, Accepted 23rd June 2011
DOI: 10.1039/c1cc12196k

By using the NEXT-A reaction, we introduced a non-natural amino acid at the N-terminus of a peptide/protein that contained a cysteine unit. The side chain of the introduced amino acid spontaneously reacted with the cysteine to afford a cyclic peptide/protein.

Cyclic peptides/proteins are attracting much attention, because they have several advantages compared with linear ones. For example, cyclization effectively improves thermostability[1,2] and/or proteolytic stability.[3] Cyclic peptides usually exhibit higher target-binding ability compared to the corresponding linear ones, because their rigid scaffolds, if they are optimized to the receptor, minimize conformational entropy loss associated with the binding.[1,3] Suga *et al.* recently reported a sophisticated methodology of ribosomal synthesis of various artificial cyclic peptides.[4] Their system, however, cannot afford large amounts of cyclic peptides/proteins, because it requires an expensive cell-free translation system as well as fragile ribozymes. Intein-mediated *in vivo* native chemical ligation (NCL) is an alternative approach to overcome the scale problem of cyclic peptide/protein synthesis.[2] One of the serious limitations of this system is that peptides/proteins can be cyclized only at the N- and C-termini.

To solve these problems, we report here a new methodology to obtain cyclic peptides/proteins by using our chemoenzymatic method[5–7] for N-terminal EXtension of peptides/proteins by using Transferase and Aminoacyl-tRNA synthetase (NEXT-A).[8] By this method we can introduce various non-natural amino acids at the basic N-terminus of peptides or proteins without using the ribosomal system. Advantages of the NEXT-A method over conventional chemical modifications are: (1) the reaction spontaneously occurs and usually ends up within 5–30 minutes. (2) The reaction can be conducted under mild conditions.

Department of Medical and Bioengineering Science, Chemical Biology Area, Division of Chemistry and Biochemistry, Okayama University, 1-1-1, Tsushimanaka, Okayama, 700-0082, Japan. E-mail: taki@biotech.okayama-u.ac.jp; Fax: +81-86-251-8021; Tel: +81-86-251-8218

† Electronic supplementary information (ESI) available: Synthesis of non-natural amino acids, the NEXT-A/Cyclization reaction, solid-phase synthesis of target peptides, tandem mass–mass analysis, quantitative analysis by Edman degradation, and expression of linear peptide-fused GFP. See DOI: 10.1039/c1cc12196k

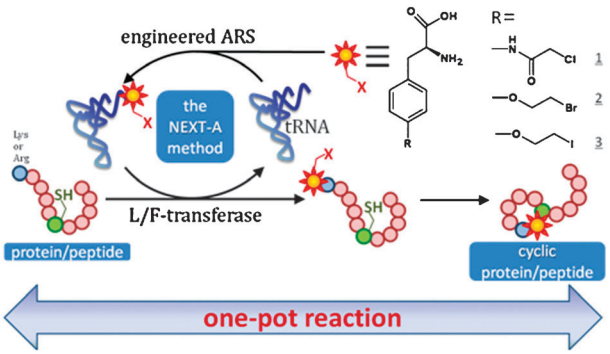

Fig. 1 The NEXT-A/Cyclization reaction. Three individual reaction steps, aminoacylation of tRNA, aminoacyl transfer, and intramolecular cyclization, proceeded as a one-pot domino reaction in a single tube.

(3) Only a single non-natural amino acid can be introduced specifically at the N-terminal Lys or Arg.

By using the NEXT-A reaction, we can introduce a non-natural amino acid possessing a haloalkyl side group. This spontaneously reacts with the thiol group of cysteine on the same peptide chain, to afford a cyclic peptide (Fig. 1). For this purpose, we newly synthesized phenylalanine or tyrosine derivatives (**1**–**3** in Fig. 1) which contain a haloalkyl group. The aromatic ring is preferable for efficient recognition by both enzymes, L/F-transferase and engineered ARS. In contrast to the hydrophobic nature of **2** and **3**, hydrophilicity of *p*-(chloroacetylamino)-L-phenylalanine (caaPhe; **1**) was unexpectedly high so that it could be soluble in water at greater than 20 mM. As shown in Fig. 2, all of these amino acids were quantitatively introduced at the N-terminus of a model peptide (UTIF; RGP**C**RAFI) by the NEXT-A reaction. Unfavourable side-reactions, such as peptide dimer formation, did not occur even after a prolonged reaction. Among the non-natural amino acids, caaPhe (**1**) gave the highest reactivity. Once **1** was introduced, the SH group of cysteine on the model peptide spontaneously attacked the α-carbon of the haloalkyl group,[4] to afford a cyclic peptide (Fig. 2B). In contrast, about 50% of the linear peptide attached to an *O*-haloethyl-L-tyrosine (**2** or **3**) was not cyclized even after the reaction mixture was incubated for 1 h at 37 °C (Fig. 2C and D). In these cases, quantitative cyclization was achieved only after prolonged incubation time (5 h) (Fig. S8, ESI†). It must be noted that all the above reactions exclusively gave cyclic

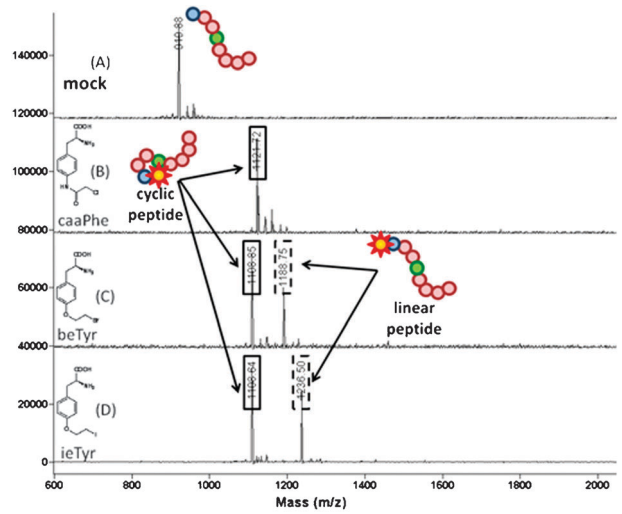

Fig. 2 MALDI-TOF mass spectra of a model peptide (RGPCRAFI) observed after the NEXT-A cyclization reaction with (A) no amino acid, (B) **1**, (C) **2**, and (D) **3**.

peptides even in the presence of 2 mM dithiothreitol (DTT; HS-R-SH) contained in the reaction buffer.

Next we tried to cyclize various peptides by using caaPhe (**1**) *via* the NEXT-A reaction followed by spontaneous cyclization (NEXT-A/Cyclization). The amino acid sequences of the peptides were RPC-acdAla, RVC-acdAla, and RGDCGGSD-YKDDDDK-badAla.[9] As shown in Fig. 3, MALDI-TOF-MS analysis of the reaction products showed that all the above

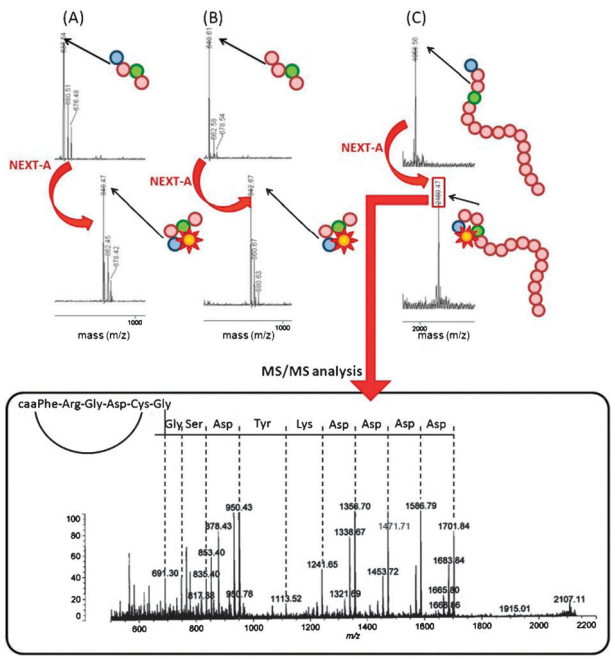

Fig. 3 MALDI-TOF mass spectra of (A) RPC-acdAla, (B) RVC-acdAla, and (C) RGDCGGSDYKDDDDK-badAla observed after the NEXT-A cyclization reaction. The signals of lower columns were obtained when **1** was present, whereas those of upper columns were obtained when **1** was absent. The inset of (C) represents tandem mass spectrum of the same peptide. Peaks corresponding to a set of assigned sequential fragment ions, generated by cleavage of amide bonds, are shown.

linear peptides spontaneously cyclized when **1** was introduced at the N-terminus. These data indicate that the cyclization efficiency is independent of the peptide sequences. Tandem mass spectrometry also supported the expected cyclic structures. At this stage, we did not find any peaks of caaPhe-peptide thioethers which would be formed through an intermolecular reaction between caaPhe and cysteine before the NEXT-A reaction. This means that the NEXT-A reaction is much faster than intermolecular thioether formation.

We also succeeded in N-terminal specific cyclization of peptide-fused green fluorescent protein (KNFFWKTFTSCD-YKDDDDK-GFP). The peptide sequence (underlined), derived from somatostatin, is known to be flexible and take random conformations.[10] The peptide-fused protein was synthesized by using *E. coli* mediated over-expression, followed by site-specific protease (usp2-cc) treatment.[8] After the NEXT-A/Cyclization was performed on the peptide-fused protein in the presence of caaPhe (**1**), the reaction product was identified by MALDI-TOF-MS (Fig. 4). The average mass of the original fusion protein [(m/z) was found to be 29 045, calculated for $[M + H]^+$: 29 050] shifted by +202. This shift corresponds to the mass of a single caaPhe unit subtracted by the Cl^- (202 Da). The N-terminal sequence of the cyclized protein was confirmed and quantitatively analyzed by Edman degradation on an amino-acid sequencer.[5–8] Efficiency of the caaPhe attachment was 75 ± 5%, and the rest of the N-terminus of the protein remained unreacted (Fig. S13, ESI†). The efficiency was not improved under several different NEXT-A reaction conditions, suggesting that some portion of the N-terminal region might be hidden in the inner core of GFP and unable to undergo the NEXT-A reaction.[5]

The protein product was site-specifically digested by enterokinase and analyzed by MS for further structural identification (Fig. 5). After the digestion, the peptide-fused GFP was split into two fragments, which were identified as the cyclic peptide and GFP, respectively. As we expected, caaPhe at the N-terminus exclusively reacted with the SH group at the C-terminus of the linear peptide, and not with two free SH groups on GFP. The SH groups on GFP are located at the inner core and outside but far from the N-terminus, respectively, and could not be reacted with the reactive haloalkyl group at the N-terminus.

GFP is known to form an intermolecular non-covalent dimer in aqueous buffer solution.[11] Nevertheless, we could hardly see intermolecular thioether formation between the two

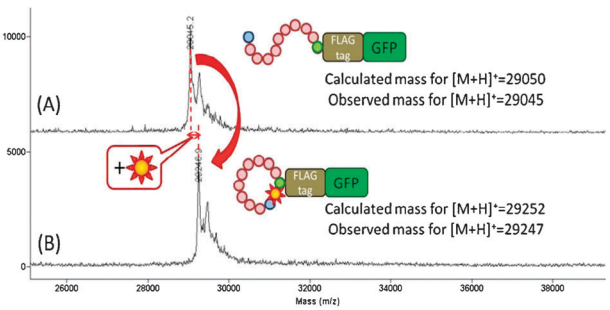

Fig. 4 MALDI-TOF mass spectra of peptide-fused green fluorescent protein (GFP) observed after the NEXT-A cyclization reaction with **1**. The signal of (A) upper column was obtained when **1** was absent, whereas that of (B) lower column was obtained when **1** was present.

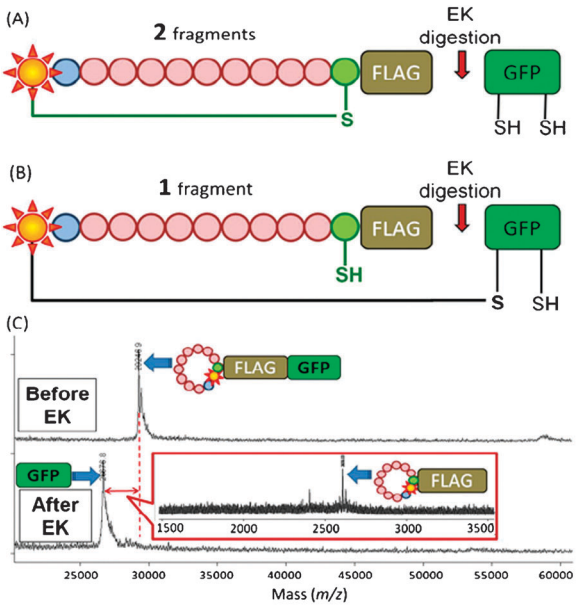

Fig. 5 Identification of cyclic peptide-GFP. The protein was site-specifically digested by enterokinase (EK) and analyzed by MS. (A) Case where caaPhe is reacted with the SH group at the C-terminus of the linear peptide. The peptide-fused protein should split into two fragments after the site-specific cleavage. (B) Case where caaPhe is reacted with the SH group on GFP. The peptide-fused protein should remain as one fragment after the site-specific cleavage. (C) MALDI-TOF mass spectrum of cyclic peptide-GFP observed after the NEXT-A cyclization reaction followed by EK treatment. The average masses of the protein (m/z) before and after the EK cleavage were found to be 29 247 (calculated for $[M + H]^+$: 29 252) and 26 677 (calculated for $[M + H]^+$: 26 665), respectively. This shift corresponds to the mass of the digested cyclic peptide fragment; the inset represents MS spectrum of the digested cyclic peptide fragment [(m/z) was found to be 2606.03, calculated for $[M + H]^+$: 2605.14].

GFP molecules. It is emphasized that the intramolecular cyclization exclusively takes precedence over the intermolecular thioether-bond formation. To the best of our knowledge, this is the first demonstration of chemoenzymatic intramolecular protein cyclization, unlike using native chemical ligation (NCL)[12] or other side-chain chemical ligation techniques.[13–15]

In conclusion, the NEXT-A/Cyclization proceeded in a single test tube as a one-pot three-step domino reaction to afford N-terminal cyclic proteins or cyclic peptides *via* thioether linkage. The intramolecular cyclization occurred exclusively, instead of intermolecular thioether formation, even though the molar concentration of the protein substrate was moderately high.[4] The NEXT-A/Cyclization proceeded under mild conditions within a very short time, so that the resulting cyclic peptide-fused GFP retained its native fluorescence activity. In principle, milligram to gram scale cyclization of a protein is now possible by using the NEXT-A reaction.[8]

Unlike existing cyclization methods for recombinant proteins,[2,12] we can build linkages between the N-terminus and any spatially compatible positions along a protein chain. The NEXT-A/Cyclization will be applicable to more complicated living systems such as peptide-fused viral proteins to construct artificial cyclic-peptide libraries.[16] Establishment of an artificial phage display system in combination with the NEXT-A cyclization is currently underway in our laboratory.

This work was supported by development projects of the Industrial Technology Research Grant Program in 2009 from New Energy and Industrial Technology Development Organization (NEDO) of Japan, Japan Society for the Promotion of Science (JSPS) for the Promotion of Science Grant-in-Aid for Young Scientists, and Otsuka Pharmaceutical Award in Synthetic Organic Chemistry. We also thank Dr Laura Nelson for careful reading of this manuscript.

Notes and references

1 E. M. Driggers, S. P. Hale, J. Lee and N. K. Terrett, *Nat. Rev. Drug Discovery*, 2008, **7**, 608–624.
2 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.
3 R. E. Kleiner, C. E. Dumelin, G. C. Tiu, K. Sakurai and D. R. Liu, *J. Am. Chem. Soc.*, 2010, **132**, 11779–11791.
4 Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami and H. Suga, *ACS Chem. Biol.*, 2008, **3**, 120–129.
5 M. Taki, A. Kuno, S. Matoba, Y. Kobayashi, J. Futami, H. Murakami, H. Suga, K. Taira, T. Hasegawa and M. Sisido, *ChemBioChem*, 2006, **7**, 1676–1679.
6 M. Taki and M. Sisido, *Biopolymers*, 2007, **88**, 263–271.
7 M. Taki, H. Kuroiwa and M. Sisido, *ChemBioChem*, 2008, **9**, 719–722.
8 K. Ebisu, H. Tateno, H. Kuroiwa, K. Kawakami, M. Ikeuchi, J. Hirabayashi, M. Sisido and M. Taki, *ChemBioChem*, 2009, **10**, 2460–2464.
9 M. Taki, Y. Yamazaki, Y. Suzuki and M. Sisido, *Chem. Lett.*, 2010, 818–819.
10 B. Hernandez, C. Carelli, Y. M. Coic, J. De Coninck and M. Ghomi, *J. Phys. Chem. B*, 2009, **113**, 12796–12803.
11 R. Y. Tsien, *Annu. Rev. Biochem.*, 1998, **67**, 509–544.
12 J. A. Camarero, J. Pavel and T. W. Muir, *Angew. Chem., Int. Ed.*, 1998, **37**, 347–349.
13 J. W. Chin, H. Neumann, K. H. Wang, L. Davis and M. Garcia-Alai, *Nature*, 2010, **464**, 441–444.
14 B. Schierling, A. J. Noel, W. Wende, L. T. Hien, E. Volkov, E. Kubareva, T. Oretskaya, M. Kokkinidis, A. Rompp, B. Spengler and A. Pingoud, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 1361–1366.
15 M. E. B. Smith, F. F. Schumacher, C. P. Ryan, L. M. Tedaldi, D. Papaioannou, G. Waksman, S. Caddick and J. R. Baker, *J. Am. Chem. Soc.*, 2010, **132**, 1960–1965.
16 C. Heinis, T. Rutherford, S. Freund and G. Winter, *Nat. Chem. Biol.*, 2009, **5**, 502–507.

