Camarero, J.A. & Muir, T.W. Biosynthesis of a head-to-tail cyclized protein with improved biological activity. J. Am. Chem. Soc. 121, 5597-5598

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · MAY 1999

Impact Factor: 12.11 · DOI: 10.1021/ja990929n

Biosynthesis of a Head-to-Tail Cyclized Protein with Improved Biological Activity

Julio A. Camarero and Tom W. Muir*

The Laboratory of Synthetic Protein Chemistry
The Rockefeller University, 1230 York Ave
New York 10021, New York

Received March 22, 1999

Head-to-tail (backbone) peptide cyclization involves condensing the N- and C-termini of a peptide together and is typically used to rigidify structure and improve *in vivo* stability.[1] Although backbone cyclization is routinely used to modify the properties of small bioactive peptides as part of drug discovery programs, its application to proteins remains largely unexplored.[2] This is primarily because the synthetic methodologies developed for the cyclization of small peptides cannot be easily extended to much larger protein systems. Nonetheless, protein cyclization remains of considerable interest to the protein engineering and protein folding communities[3] since it is expected to provide useful insights into the importance of the C- and N-termini of a protein in folding, structural stability, and, consequently, biological activity. In the present work, we describe a novel biosynthetic process, based on an intramolecular chemical ligation reaction, which allows the cyclization of recombinant polypeptide precursors. This approach has been successfully applied to the generation of a circular version of a Src homology 3 (SH3) domain with improved biological activity over the wild-type protein.

Several strategies for the preparation of circular polypeptides from unprotected linear precursors have been described. In pioneering studies in this area, Creighton and co-workers used a chemical cross-linking approach to prepare a backbone cyclized version of bovine pancreatic trypsin inhibitor.[2a] More recently, chemical[2b,c,4] and enzymatic[2d] intramolecular ligation methods have been developed which allow linear synthetic peptides to be efficiently cyclized under aqueous conditions. However, the requirement for synthetic peptide precursors has limited these chemical/enzymatic cyclization approaches to relatively small polypeptide systems. One potential solution to this size problem would be to use recombinant polypeptides as precursors, thereby allowing large polypeptides and even proteins to be backbone-cyclized.

In principle, it should be possible to generate a circular recombinant protein by using an intramolecular version of the native chemical ligation approach.[5] Native chemical ligation involves the chemoselective reaction that occurs between an N-terminal cysteine residue in one peptide and an α-thioester group within a second peptide, resulting in the formation of a normal peptide bond. Importantly, incorporation of both of these reactive moieties within the same synthetic polypeptide leads to efficient backbone cyclization.[2b,c,4] Considerable progress has been made in generating recombinant polypeptides for use in native chemical ligation reactions: Verdine and co-workers have used a mutagenesis/factor Xa proteolysis strategy to generate reactive recombinant N-terminal cysteine proteins,[6] while we and others have shown that recombinant polypeptide α-thioesters can be prepared by chemically intercepting a naturally occurring protein-splicing reaction.[7] To date, recombinant polypeptide segments have been used only in intermolecular native chemical ligation reactions.[7] Here we describe a straighforward route to backbone-cyclized recombinant polypeptides based on an intramolecular version of native chemical ligation (Figure 1).

As a demonstration of our biosynthetic cyclization strategy, we attempted to cyclize the N-terminal SH3 domain from the c-Crk adaptor protein.[8] This 57-residue protein domain possesses a globular structure[9] composed of five β-strands which position the N- and C-termini in close proximity.[10] A bacterial expression plasmid was prepared in which the gene corresponding to the SH3 domain (residues Y^{136} to Y^{190} of murine c-Crk) was cloned into a commercially available intein expression system.[11] The SH3 domain was modified at the DNA level to append the sequence MIEGRC at the N-terminus and to add an extra Gly residue at the C-terminus. The MIEGRC motif contains a factor Xa proteolysis site and allows the generation of an N-terminal Cys upon *in vitro* proteolysis.[6,7e] Importantly, this motif acts as a cysteine-protecting group, preventing premature ligation reactions.[7e] The addition of the C-terminal Gly residue was carried out to improve the kinetics of cyclization as well as to stabilize the formation of the new loop between the N- and C-termini.[12]

The recombinant fusion protein (**1**) was expressed in *Escherichia coli* to give, after affinity purification, a major component on SDS-PAGE with the correct apparent molecular weight (Figure 2A). A secondary band representing less than 10% of total protein was also detected. This was assigned as the intein-CBD protein fragment and is presumably due to some premature cleavage of the fusion protein (**1**). As illustrated in Figure 1, cyclization of the recombinant SH3 domain was achieved by simply treating the fusion protein (**1**) adsorbed on chitin beads with factor Xa protease at pH 7.2 for 10 h. This proteolysis step afforded N-terminal Cys fusion protein (**2**) which spontaneously reacted in an intramolecular fashion to yield the corresponding circular SH3 domain (**3**) with concomitant cleavage from the beads. As shown in Figure 2A and B, the cyclization process was extremely clean, allowing the circular SH3 domain (**3**) to be readily purified

* Corresponding author: Telephone: 212-327-7368. Fax: 212-327-7358. E-mail: muirt@rockvax.rockefeller.edu.

(1) For recent reviews, see: (a) Hruby, V. J.; Al-Obeidi, F. *J. Biochem.* **1990**, *268*, 249−262. (b) Rizo, J.; Gierasch, L. M. *Annu. Rev. Biochem.* **1992**, *61*, 387−418.

(2) (a) Goldenburg, D. P.; Creighton, T. E. *J. Mol. Biol.* **1983**, *165*, 407−413. (b) Camarero, J. A.; Pavel, J.; Muir, T. W. *Angew. Chem., Int. Ed.* **1998**, *37*, 347−349. (c) Tam, J. P.; Lu, Y. A. *Prot. Sci.* **1998**, *7*, 1583−1592. (d) Jackson, D. Y.; Burnier, J. P.; Wells, J. A. *J. Am. Chem. Soc.* **1995**, *117*, 819−820.

(3) Disulfide linkages have typically been used as a structural constraint: (a) Sauer, R. T.; Hehir, K.; Stearman, R. S.; Weiss, M. A.; Jeitler-Nilsson, A.; Suchanek, E. G.; Pabo, C. O. *Biochemistry* **1986**, *25*, 5992−5998. (b) Matsumura, M.; Signor, G.; Matthews, B. W. *Nature* **1989**, *342*, 291−293. (c) Li, B.; Tom, J. Y. K.; Oare, D.; Yen, R.; Fairbrother, W. J.; Wells, J. A.; Cunningham, B. C. *Science* **1995**, *270*, 1657−1660. (d) Otzen, D. E., Fersht, A. R. *Biochemistry* **1998**, *37*, 8139−8140.

(4) (a) Camarero, J. A.; Muir, T. W. *Chem. Commun.* **1997**, *1997*, 1369−1370. (b) Zhang, L.; Tam, J. P. *J. Am. Chem. Soc.* **1997**, *119*, 2363−2370.

(5) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776−779.

(6) Erlandson, D. A.; Chytil, M.; Verdine, G. L. *Chem. Biol.* **1996**, *3*, 981−991.

(7) (a) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705−6710. (b) Severinov, K.; Muir, T. W. *J. Biol. Chem.* **1998**, *273*, 16205−16209. (c) Xu, R.; Ayers, B.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 388−393. (d) Evans, T. C.; Benner, J.; Xu, M.-Q. *Protein Sci.* **1998**, *7*, 2256−2264. (e) Cotton, G. J.; Ayers, B.; Xu, R.; Muir, T. W. *J. Am. Chem. Soc.* **1999**, *121*, 1100−1101. (f) Evans, T. C.; Benner, J.; Xu, M.-Q. *J. Biol. Chem.* **1999**, *274*, 3923−3926. (g) Welker, E.; Scheraga, H. A. *Biochem. Biophys. Res. Commun.* **1999**, *254*, 147−151.

(8) Knudsen, B. S.; Feller, S. M.; Hanafusa, H. *J. Biol. Chem.* **1994**, *269*, 32781−32787.

(9) Wu, X.; Knudsen, B.; Feller, S. M.; Zheng, J.; Sali, A.; Cowburn, D.; Hanafusa, H.; Kuriyan, J. *Structure* **1995**, *3*, 215−226.

(10) Proximal N- and C-termini, an apparent prerequisite for head-to-tail cyclization, are a surprisingly common feature in protein folds, particularly in single-domain systems: Thornton, J. M.; Sibanda, B. L. *J. Mol. Biol.* **1983**, *167*, 443−460.

(11) This vector allows the generation of a fusion protein in which the polypeptide of interest is linked via its C-terminus to a genetically modified protein-splicing domain (termed an intein): Chong, S.; Mersha, F. B.; Comb, D. G.; Scott, M. E.; Landry, D.; Vence, L. M.; Perler, F. B.; Benner, J.; Kucera, R. B.; Hirnonen, C. A.; Pelletier, J. J.; Paulus, H.; Xu, M.-Q.; *Gene* **1997**, *192*, 271−281.

(12) (a) Iwakura, M.; Nakamura, T. *Protein Eng.* **1998**, *11*, 707−713. (b) Martinez, J. C.; Viguera, A. R.; Berisio, R.; Wilmanns, M.; Mateo, P. L.; Filiminov, V. V.; Serrano, L. *Biochemistry* **1999**, *38*, 549−559.

10.1021/ja990929n CCC: $18.00
Published on Web 05/29/1999

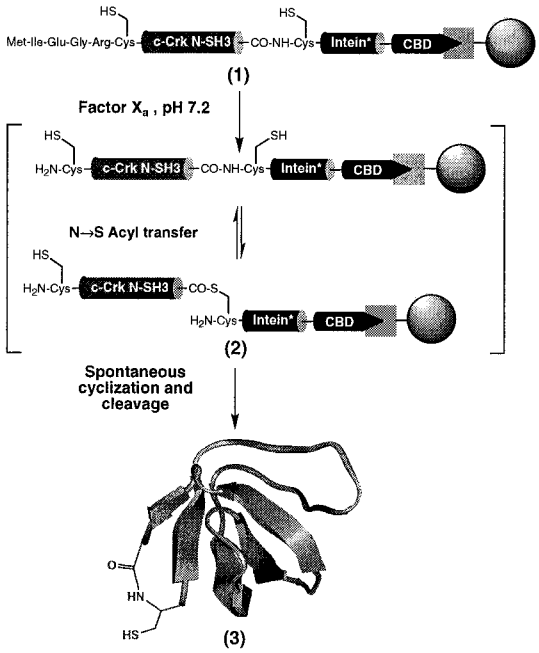

Figure 1. Biosynthetic protein cyclization strategy. The first step involves the cloning and expression of the recombinant fusion protein (**1**) (CBD refers to a chitin binding domain). Importantly, the presence of an engineered intein domain in (**1**) means the amide bond connecting the SH3 domain and the intein is in equilibrium with a thioester (i.e., the intein promotes an N → S acyl transfer). Following affinity purification of (**1**) on chitin beads, the leader sequence MIEGR is removed by treatment with Factor Xa protease to give intermediate protein (**2**). The presence of both an N-terminal Cys and a thioester within (**2**) results in a spontaneous intramolecular native chemical ligation reaction (and simultaneous cleavage off the chitin bead) to give the circular c-Crk N-SH3 domain (**3**).

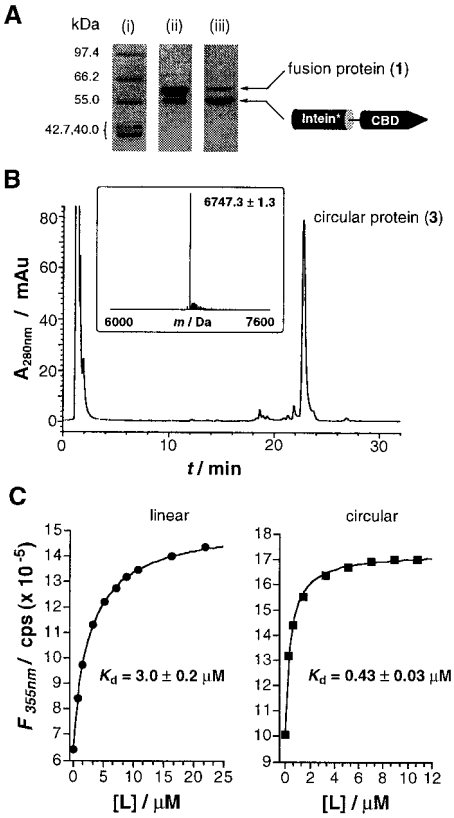

Figure 2. (A) SDS-PAGE analysis of the recombinant fusion proteins: (i) protein markers, (ii) fusion protein (**1**) after affinity purification, and (iii) residual protein on chitin beads after being treated with Factor Xa protease. (B) Analytical reverse-phase HPLC of the crude reaction supernatant after 10 h. The major peak corresponds to the expected circular protein (**3**). Inset: ESMS of purified (**3**). (C) Change in fluorescence emission intensity of the linear and circular forms of the c-Crk N-SH3 domain upon addition of the proline-rich ligand H-PPPALPPKKRXYX-NH_2 (L), where X stands for 6-aminohexanoic acid. Excitation and emission were at 298 and 355 nm, respectively.

by reverse phase HPLC.[13] Interestingly, this intramolecular process did not require the presence of a thiol cofactor (absolutely necessary to facilitate the intermolecular ligation reaction[7]). This may be because the cyclization reaction was assisted by the folded state of the SH3 domain, which leads to an extremely high local concentration of the two reactive groups.[2b]

The ligand-binding properties of the linear[14] and cyclic Crk SH3 domains were studied by using a fluorescence-based binding assay (Figure 2C).[15] As predicted based on the native structure of the SH3 domain,[9] cyclization did not have a detrimental effect on ligand-binding activity; on the contrary, the circular SH3 domain had a ~7-fold higher affinity for the ligand than the linear SH3 domain ($K_{d[circular]} = 0.43 \pm 0.03$ μM and $K_{d[linear]} = 3.0 \pm 0.2$ μM). Preliminary spectroscopic analysis of the circular vs linear proteins provides some clues as to the origin of the increase in biological activity.[16] The Trp fluorescence emission spectrum of the circular SH3 domain is blue-shifted compared to the linear protein ($\lambda_{max(circular)} = 348$ nm and $\lambda_{max(linear)} = 355$ nm) and is ~150% more intense, suggesting a net increase in the hydrophobic environment of the Trp residues. Furthermore, structural analysis by circular dichroism spectroscopy indicates that there are some subtle differences in the secondary structure content between the linear and circular versions of the protein, with the former apparently having less β-sheet than the latter. Taken together, these results are consistent with the idea that cyclization results in a less flexible tertiary structure and that the increase in ligand-binding affinity is thus largely an entropic effect. Future structural and thermodynamic studies will explore this hypothesis further.

In summary, a novel biosynthetic approach for preparing head-to-tail cyclized recombinant polypeptides has been developed. We believe this technology will be a generally useful tool in peptide and protein engineering and will complement the existing approaches for constraining polypeptide structures.[3] Equally, new opportunities can be envisioned in protein sequence permutation[17] by combining the chemoselective cyclization of recombinant proteins with subsequent main-chain cleavages.

Acknowledgment. We thank Francine Perler, Graham Cotton, and Professor David Cowburn for insightful discussions. The research was supported by the Rockefeller University, a Pew Scholarship in the Biomedical Sciences (T.W.M.), and the National Institutes of Health (GM55843, T.W.M).

Supporting Information Available: Full experimental details including characterization of all peptide and protein products (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA990929N

(13) Circular protein (**3**) was characterized by ESMS [observed mass = 6747.3 ± 1.3 Da, expected (average isotopic composition) = 6747.6 Da] and tryptic digestion which afforded a peptide map consistent with a head-to-tail circular structure. Approximately 0.5 mg of purified (**3**) was obtained from 1 L of bacterial growth.

(14) The linear c-Crk N-terminal SH3 domain was obtained using standard recombinant expression techniques.

(15) c-Crk N-SH3 binds to the sequence, PPPALPPKKR, derived from the nucleotide exchange factor, C3G: see ref 8.

(16) An increase in binding affinity was also recently observed for a circular version of a WW protein domain: see ref 2b.

(17) (a) Graf, R.; Schachman, H. K. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11591−11596. (b) Hennecke, J.; Sebbel, P.; Glockshuber, R. *J. Mol. Biol.* **1999**, *286*, 1197−1215.