Chemical Synthesis of Circular Proteins

James P. Tam* and Clarence T. T. Wong

School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551.

*To whom correspondence should be addressed: James P. Tam, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551 Tel.: (+65) 63162863. Fax: (+65) 67913856. E-mail: jptam@ntu.edu.sg

Keywords: Cyclic peptide; entropic ligation; thioester; cysteine-rich peptides; oxidative folding

Abstract

Circular proteins, once thought to be rare, are now commonly found in plants. Their chemical synthesis, once thought to be difficult, is now readily achievable. The enabling methodology is largely due to the advances in entropic chemical ligation to overcome the entropy barrier in coupling the N- and C-terminal ends of large peptide segments for either intermolecular ligation or intramolecular ligation in end-to-end cyclization. Key elements of an entropic chemical ligation consist of a chemoselective capture step merging the N- and C-termini as a covalently linked O/S-ester intermediate to permit the subsequent step of an intramolecular O/S-N acyl shift to form an amide. Many ligation methods exploit the supernucleophilicity of a thiol side chain at the N-terminus for the capture reaction, which makes cysteine-rich peptides ideal candidates for the entropy-driven macrocyclization. Advances in desulfurization and modification of the thiol-containing amino acids at the ligation sites to other amino acids add extra dimensions to the entropy-driven ligation methods. This review describes recent advances of entropy-driven ligation to prepare circular proteins with or without a cysteinyl side chain.

Introduction

Circular proteins and their smaller versions, cyclic peptides, have a characteristic head-to-tail or end-to-end peptide backbone structure. The absence of both N- and C-termini in these macrocycles confers them resistance to exopeptidase and heat degradation, enhances their conformational stability and maximizes epitope display upon their circular contiguous sequences for interactions with other molecules. These advantages have provided incentives to engineer circular proteins by grafting linear bioactive peptides or epitopes into various structural scaffolds for therapeutic applications (1).

For the purpose of this review, we refer to a cyclic peptide >15 amino acids as a "circular protein or miniprotein". This arbitrary cutoff point has two justifications. Firstly, many cyclic peptides <15 amino acid residues are derived from non-ribosomal synthesis and the substrate specificity of a cyclase enzyme such as tyrocidine thioesterase (Tyc TE) could accommodate precursors 6 to 14 amino acids in length (2). Secondly, circular proteins are gene-encoded and processed by specific enzymes from their linear precursors containing a signal peptide and one or more pro-domains (3). Cyclic peptides and circular proteins are frequently found in bacteria (microcin), fungi (cyclosporin), animal (θ-defensins), and more commonly, in plants (cyclotides) (4-7).

From a synthetic standpoint, chemical synthesis of circular proteins has been a formidable challenge using the traditional enthalpic methods which require partially or globally protected linear precursors and a strong enthalpic activation of the C-terminal residue for the cyclization reaction (Fig 1A). Strong activation of the C-terminal moiety is necessary to overcome the entropy barrier in the coupling reactions often leads to epimerization of the C-terminal amino acid residue and oligomerization to dimers and trimers. Work prior to 1997 showcased the challenges associated with enthalpy-driven cyclization of peptides <15 amino acids. In 1997, we reported the total synthesis of circular proteins of 31 amino acids, cyclotides circulin B (cirB), cyclopyrrolotride
Chemical Synthesis of Circular Proteins

(CPT) and two other cyclotides in the following year (8,9). These studies represent a breakthrough since they were the first reports on a successful chemical synthesis of naturally occurring circular proteins using an entropy-driven ligation chemistry which is conceptually different and operationally much simpler than the conventional enthalpy cyclization method.

Many review articles have been published over the last few years, with a majority articulating the occurrence, chemistry and biological functions of cyclic peptides (10-13). Here, we will focus on contemporary entropy-driven ligation chemistry for the synthesis of circular proteins.

Entropy-driven ligation chemistry.

During the 1990s, there was a paradigm shift in the synthesis of large peptides and proteins (14-18). The sea change was driven by the discovery of entropic activation in convergent peptide synthesis using unprotected peptides as building blocks to enable an efficient coupling reaction (ligation) of the N and C termini of two peptide segments (intermolecular ligation) or a single peptide segment (macrocyclization). The conceptual difference between an entropic and an enthalpic ligation is that the entropic ligation is proximity-driven to overcome the entropy barrier, merging the N- and C-termini as a covalent intermediate, usually as an O- or S-ester, to permit an intramolecular, proximity-driven O-N or S-N acyl shift to form an amide (Fig. 1B). As such, it eliminates the risk of the C-terminal epimerization in peptide synthesis, a major side reaction that results in a side product difficult to be removed by chromatographic methods.

Although the principles of entropy-driven reactions are well established in the cyclization of small molecules in organic synthesis, Kemp and his coworkers are pioneers and strong advocates of entropic chemical ligation in the arena of large molecules, including peptide synthesis. In the 1980s, they demonstrated its feasibility by placing both peptide segments on rigid organic templates to facilitate an O-N acyl shift in organic solvents to form an amide bond (14). Kemp's template-based ligation approach was limited to general applications due to the attendant slow rates of the O-N acyl shift reactions mediated by a large 12-member-ring intermediate. In 1994, our laboratory reported an entropy-driven chemical ligation method. Without the use of an organic template, we used an N-terminal Cys-, Ser- or Thr-containing peptide segment with a C-terminal fragment bearing an ester aldehyde (15,16). The ligation was performed in water, without the use of protecting groups or a coupling reagent. The key element of our approach is a chemoselective capture, forming a covalent ester between two peptide fragments to enable an intramolecular O-N acyl shift mediated by a small five-member ring intermediate to form an amide bond, and in our case, a pseudoproline bond. In the same year, Kent's group reported a cysteine-based ligation with another segment bearing a C-terminal thioester and an S-N acyl shift through a five-member ring intermediate to form an amide bond (17). Since cysteine is regenerated at the ligation site, they have called their ligation method as native chemical ligation (NCL). The super-nucleophilic cysteinyl thiol of an unprotected peptide permits a rapid thiol-thioester exchange at a basic pH to form a covalent thioester intermediate and the rate of S-N acyl shift is faster than the corresponding O-N acyl shift. Both factors contribute to the effectiveness of NCL, which our laboratory confirmed shortly after Kent's communication (18). Indeed, the use of a C-terminal thioester in the chemoselective capture step, first reported by Wieland in 1953 (19), is an excellent choice for chemical ligation. The combined work of Kemp's group and the early ligation studies published in the 1994 and 1995 period provided the conceptual framework of the entropic ligation approach for nearly all subsequent chemical ligation methods developed in the past 16 years (14-18). Interestingly, we found that an N-terminal His-containing segment could also be exploited for chemical ligation because of the exceptional reactivity of its nucleophilic imidazole side chain. Under acidic conditions, the imidazole-based capture of another peptide segment containing a C-terminal thioester derivative affords a covalent acyl imidazole intermediate, leading to an N-N acyl shift to form a histidine peptide bond at the ligation site (20).

Two congruent but independent developments accelerated the acceptance of entropy-driven ligation during the 1990s. Xu et al. (21) found...
that the intein-mediated protein splicing process could undergo four acyl shift reactions, three of which are catalyzed by the intein enzyme in the early stages. The final acyl shift involving either an O-N or S-N acyl shift to form the peptide bond between the spliced extein fragments is uncatalyzed and spontaneous. This O-N or S-N acyl shift reaction is also mediated by a five-member ring intermediate, similar to the entropic chemical ligation reactions described in the 1994-5 papers (15-18). In 1998, Muir and his coworkers developed ‘expressed protein ligation’, a semi-biosynthetic approach for protein synthesis to generate a peptide thioester using the intein-mediated splicing mechanism (22). The ability to generate building blocks of the N-terminal Cys-containing and C-terminal thioester-containing segments by either chemical or recombinant methods further facilitates protein synthesis of all shapes and sizes by the entropic chemical ligation approach for various biological studies (23).

**Chemical ligation and macrocyclization based on entropic ligation.** With the development of entropy-driven ligation, the synthesis of circular proteins became approachable in the late 1990s. This was timely as many cysteine-rich peptides consisting of 30 to 60 amino acids had been identified from natural sources and their pharmacological evaluations were limited by suitable methods of chemical synthesis (24,25). Their characteristic signature of a multiple-cysteine structure which is a liability in conventional chemical synthesis becomes an asset in entropic chemical ligation methods. An example is the family of cyclotides which contain 28-37 residues, six cysteines, and a circular peptide backbone. For cyclotides, there are six possible X-Cys ligation sites, and the choice is often governed by choosing the least hindered amino acid to serve as the C-terminal thioester. Thus, the order preference of X is Gly>Ala>Leu, Phe, Ser>Thr, Val, Ile>>Pro. The C-terminal thioesters of Asp and Glu are unstable and seldom used for ligation reactions.

The thiol-thioester exchange reaction in the capture reaction of a thiol-based ligation also plays an essential role in facilitating the entropy-driven macrocyclization of cysteine-rich peptides (Fig 1C). We found that the internal cysteines enhance the cyclization rates of cyclotides, kalata B1, cyclopyrrolidide, circulin A and B, which contain six cysteines (26). The cyclization rate was fast under denaturing condition of 6M GnHCl. In the enthalpic activation, the cyclization is performed at high dilution to prevent oligomerization. In contrast, we observed that there was no dimer or oligomer formation even at peptide concentration up to 0.5 mM. We proposed that the rate enhancement and lack of oligomerization of the Cys-rich cyclotides could be attributed to the thia-zip cyclization mechanism (26,27). Thia-zip cyclization involves two reactions: the reversible thiol-thioester exchanges through intramolecular transthioesterifications as thiolactones and the irreversible S-N acyl migration of the amino-terminal thiolactone to the lactam (amide). The facile and reversible intramolecular transthioesterifications lead to the ring expansion as discrete thiolactone intermediates, pulling the two ends successively into close proximity as the end-to-end, N-terminal thiolactone to permit an S-N acyl shift via a five-member ring to form an end-to-end circular protein. As a result, the thia zip-assisted cyclization, through small discrete intermediates, is entropy-driven and more efficient than the corresponding one-step, end-to-end cyclization. Along the same line of reasoning, we demonstrated the use of silver ions in assisting cyclization between the N-terminal Ser/Thr/Asn/Gly and a C-terminal thioester. In this approach, silver ion captures and brings two ends to close proximity for an acyl transfer to occur. Furthermore, the silver ion forms a complex with the thiol group on the thioester, making it a better leaving group in the acyl transfer reaction (Table 1, Entry 14) (28).

Given the advances of chemical ligation, the synthesis of circular proteins, both designed and naturally occurring, became routinely achievable. Naturally-occurring circular proteins including cycloviolacin O2, hedyotide B1, and rhesus θ-defensin-1 (RTD-1) have been successfully prepared (6,29,30). Several groups including Craik and Camarero have extensively applied the cysteine-based ligation for preparing cyclotides and their analogs for biochemical and biophysical studies (31-33). For designed circular proteins, Yu et al. employed cysteine ligation to introduce an end-to-end cyclic backbone to an α-defensin (34). The circularized defensin leads to improved stability against exopeptidases, increased conformational stability and enhanced tolerance to high salt sensitivity in antimicrobial assays.
Cysteine-free ligation and macrocyclization.
The occurrence of cysteine (1.7%) in proteins is lower than the other amino acids and its absence is common in naturally-occurring cyclic peptides. Furthermore, there is also a need to increase the flexibility of entropic ligation methods that use non-cysteinyl N-terminal amino acids. Over the past 16 years, a suite of ligation methods has emerged to meet these needs, and they are summarized in Table 1. Many employed a combination strategy of entropic ligation and a follow-up chemical modification step. They include the use of N-terminal cysteine mimetics (Table 1, Entries 4-8,12) or thiol-auxiliary groups on or near the N-terminal amino acids (Entries 9,11). These methods generally adhere to the principle of the thiol-based ligation but use various forms of N-terminal thiol esters to facilitate the capture reactions by thiol-thioester exchange reactions with C-terminal esters and then an S-N acyl shift to form an amide. The resulting thiolate side chain or auxiliary group at the ligation site is followed by an additional step, either a chemical modification or removal to give rise to a different amino acid (Table 1).

An example of combining thiol-based ligation and chemical modification is methionine ligation. It employs homocysteine (Hcy) to replace Cys as the N-terminal amino acid (35). The S-alkylation of Hcy by a methylating agent at the ligation site smoothly converts it to methionine (Table 1, Entry 3). The combination of using an N-terminal Cys or Hcy for ligation followed by an alklylation reaction has been extended to pseudo-lysine, and pseudo Asp and Glu acids (Table 1, Entry 10) (36,37).

A promising approach is the combination thiol-based chemical ligation and desulfurization of the resulting thiol. Yan and Dawson exploited the cysteine-based ligation in the synthesis of a cyclic peptide and then converted the resulting cysteiny1 residue to alanine through a metal-based desulfurization by Raney nickel or Pd/Al2O3 (38). Metal-based desulfurization of small proteins may not be practical as it often requires a large excess of metal which produces side reactions and irreversible peptide loss due to aggregation and absorption onto the metal surfaces. To address these limitations, Wan and Danishefsky developed a nonmetal desulfurization of cysteine via reaction with a free radical using tris(2-carboxethyl)phosphine (TCEP), a thiol and a radical initiator 2,2'-azobis(2-ethylazobisethyl) peroxide (39). Desulfurization was extended to N-terminal β-mercapto-phenylalanine, β-mercaptovaline (penicillamine) and γ-mercaptoalnine, and γ-mercaptovaline to afford Phe, Val and Lys respectively (Table 1, Entries 6-8,12) (36,40-42). A limitation of this approach and the N-terminal auxiliary approach is that few such building blocks (except for Cys to Ala and penicillamine to Val) are commercially available and thus may require considerable synthetic expertise for their preparation.

Chemo-enzymatic macrocyclization.
The use of enzymes for peptide coupling reactions through reverse proteolysis was described by Bergmann and Fruton over sixty years ago (43). Over the past few decades, enzymatic coupling of peptide segments has been explored as an alternative to enthalpic ligation (44,45). Jackson et al. reported the use of subtiligase in peptide cyclization in 1995 (Table 1, Entry 16). The linear peptide precursor was synthesized as a C-terminal glycyl-phenylalanylamide ester for subtiligase recognition that results in end-to-end backbone cyclization (46). Another chemo-enzymatic method was reported by Tan et al. who exploited subtiligase to catalyze the hydrolysis or the aminolysis of a peptide glycolate ester substrate as a hydrothiolase in the synthesis of peptide thioacids for peptide ligation (47). For the circular trypsin inhibitor family, trypsin can play the role as a cyclase to cyclize a linear synthetic precursor to a circular protein such as the sunflower trypsin inhibitor (48) and the cyclic trypsin inhibitor MCoTI-I (49).

Oxidative folding of cysteine-rich peptides.
Since many circular proteins are cysteine-rich, the oxidative folding step in forming the correct connectivity of disulfide bonds is another key step in their synthesis. However, forming the native disulfide bonds under oxidative folding conditions remains empirical with unpredictable outcomes. Therefore, the chemoselective method was introduced in the synthesis of cyclotides in 1997 (8). With a pair of HF-stable cysteine protecting group such as the acetamidomethyl (Acm) group, it greatly reduces the number of
disulfide isomers formation. This method had been used in the synthesis of many cysteine-rich peptides such as engineered protegrins and defensins. The other type of chemoselective method was reported by Alewood and his coworkers in their work on conotoxins, by replacing a pair of cysteines by selenocysteine. As selenium is much more reactive than sulfur, the diselenide bond forms almost spontaneously to confer its selectivity over cystines (50).

Beside chemoselective methods, global oxidative folding has also been employed in folding of cysteine-rich peptides. This reaction involves concurrent reduction (S-S breakage) and oxidation (S-S formation) in the presence of both reducing and oxidizing agent, forming a redox pair that can re-shuffle and rearrange the disulfide pairs into its native disulfide bonds, and presumably one of the stable conformational states. The reaction is mediated by the thiol/thiol exchange reaction that takes place in the basic pH in order to generate active thiols for the S-S exchange reactions. This method involves fewer synthetic steps than the chemoselective method.

The global oxidative folding of cysteine-rich peptides generally requires high dilution of peptides in a suitable combination of solvents to prevent aggregation and precipitation. It also requires maintaining a proper redox potential such as GSH and GSSG in a basic buffer to minimize the accumulation of dead-end products (51-53). Interestingly, the use of a combination of aqueous and hydrophobic organic solvent appears to work well for the oxidative folding of cyclotides which display an unusual hydrophobic side-chain arrangement due to the bulky interior disulfide core. A hydrophobic organic solvent may prevent aggregation during the folding process as the hydrophobic side-chains are being externalized to the solvent surface in forming the cystine knot. Such a rationale was used successfully in the oxidative folding of the Möbius cyclotide kalata B1 by Daly et al. (54) who used an aqueous hydrophobic solvent combination containing 50% 2-propanol. Under such a condition, they found that much higher yields of cyclic oxidized cyclotides were achieved by allowing the backbone cyclization to proceed prior to the folding as compared with the reverse order of folding first and then followed by oxidation. In general, this order of synthetic operation, cyclized first and then oxidation, has been employed for the synthesis of circular proteins irrespective of the disulfide forming strategy. However, the use of the solvent combination containing 50% 2-propanol in the oxidative folding of the bracelet family of cyclotides was found to be unsuccessful. For the successful oxidative folding of a bracelet cyclotide, cycloviolacin O2, Aboye et al. (29) reported the use of a high percentage of DMSO and detergent Brij 35 in a solvent combination to fold it with a reasonable yield. In a systematic study, various parameters for the oxidative folding of a bracelet cyclotide were examined and the key parameter was found to be the concentration of an organic co-solvent. It appears that the more hydrophobic cyclotides may require a strong hydrophobic solvent combination in order to fold the circular protein successfully. One of the most hydrophobic bracelet cyclotides, hedyotide B1, requires a combination of solvents containing 70% 2-propanol for its optimal folding (30).

Conclusions
The development of entropic ligation and macrocyclization protocols based on the thiol capture of an N-terminal cysteine enables the chemical synthesis of many circular proteins in recent years. The combination of ligation and chemical modifications would likely extend entropic coupling methods to include most of the N-terminal aliphatic amino acids as ligation sites in the macrocyclization of circular proteins. Thus, the availability of a suite of ligation methods would permit the chemical synthesis of circular proteins of any size for biochemical, biophysical and pharmacological studies in laboratory and for evaluation of biologics as potential drug candidates.

Acknowledgments—We thank Ms. Xinya He Mu and Yibo Qiu for their assistance in preparing this manuscript. This research was supported in part by Biomedical Research Council (BMRC 09/1/22/19/612) of A*STAR and Academic Research Fund (ARC21/08) of Ministry of Education in Singapore.
References

Chemical Synthesis of Circular Proteins

51. Gunasekera, S., Daly, N. L., Clark, R. J., and Craik, D. J. (2009) *Antioxid. Redox Signal.* **11**, 971-980
FIGURE LEGENDS

FIGURE 1. Chemical ligation and cyclization schemes. (A) Enthalpy-driven cyclization based on C-terminal activation (OA=active ester). (B) Entropy-driven ligation via S-N acyl shift between C-terminal thioester and N-terminal Cys. (C) Thia-zip cyclization mechanism for forming a circular protein through stepwise ring expansion of thiolactone intermediates to an end-to-end thiolactone to enable a lactam formation via a S-N acyl shift. X and Y represent a pair of nucleophile and electrophile. Z is the bond of the capture intermediate. R represents side chains.
Table 1. Summary of entropic ligation methods

<table>
<thead>
<tr>
<th>Site-based ligation methods</th>
<th>Starting materials</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-terminal</td>
<td>C-terminal</td>
<td></td>
</tr>
<tr>
<td><strong>Site-based ligation methods</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Pseudoproline(^1)</td>
<td>Cys/Ser/Thr</td>
<td>Ester aldehyde</td>
</tr>
<tr>
<td>2. Cysteine(^2)</td>
<td>Cys</td>
<td>Thioester</td>
</tr>
<tr>
<td>3. Histidine</td>
<td>His</td>
<td>Perthioester</td>
</tr>
<tr>
<td>4. Methionine(^3)</td>
<td>Homocysteine (Hcy)</td>
<td>Thioester</td>
</tr>
<tr>
<td>5. Alanine(^4)</td>
<td>Cys</td>
<td>Thioester</td>
</tr>
<tr>
<td>6. Pheylalanine(^4)</td>
<td>β-Mercaptophenylalanine</td>
<td>Thioester</td>
</tr>
<tr>
<td>7. Valine(^4)</td>
<td>Penicillamine</td>
<td>Thioester</td>
</tr>
<tr>
<td>8. Valine(^4)</td>
<td>γ-mercaptovaline</td>
<td>Thioester</td>
</tr>
<tr>
<td>9. Serine/Aspartic acid(^5)</td>
<td>Thiosugar-Ser/Asp</td>
<td>Thioester</td>
</tr>
<tr>
<td>10. Pseudolysine(^3)</td>
<td>Cys</td>
<td>Thioester</td>
</tr>
<tr>
<td>11. Glycine(^6)</td>
<td>N-Auxillary-Gly</td>
<td>Thioester</td>
</tr>
<tr>
<td>12. Lysine(^4)</td>
<td>γ-Mercaptopolysine</td>
<td>Thioester</td>
</tr>
<tr>
<td>13. Serine/Threonine(^5)</td>
<td>Ser/Thr</td>
<td>Ester salicylaldehyde</td>
</tr>
<tr>
<td><strong>Other ligation methods</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Ag(^+)-assisted (Ser/Thr)</td>
<td>Ser/Thr/Asn/Gly</td>
<td>Thioester</td>
</tr>
<tr>
<td>15. Staudinger</td>
<td>Azido amino acids</td>
<td>Phosphanyl thioester</td>
</tr>
<tr>
<td>16. Subtiligase</td>
<td>Met/Ala/Lys/Leu/Arg</td>
<td>Glycolated phenylester</td>
</tr>
</tbody>
</table>

1. Thiazolidine/oxazolidine ligation without (Entry 1) or with (Entry 13) a subsequent chemical manipulation. 2. Native Chemical Ligation 3. S-Alkylation after ligation 4. Desulfurization after ligation. 5. Removal of thiosugar (Entry 9), O-auxiliary (Entry 13) and N-auxiliary group (Entry 11) after ligation.
FIGURE 1. Chemical ligation and cyclization schemes. (A) Enthalpy-driven cyclization based on C-terminal activation (OA=active ester). (B) Entropy-driven ligation via S-N acyl shift between C-terminal thioester and N-terminal Cys. (C) Thia-zip cyclization mechanism for forming a circular protein through stepwise ring expansion of thiolactone intermediates to an end-to-end thiolactone to enable a lactam formation via a S-N acyl shift. X and Y represent a pair of nucleophile and electrophile. Z is the bond of the capture intermediate. R represents side chains.