

In vivo biosynthesis of an Ala-scan library based on the cyclic peptide SFTI-1

Jeffrey Austin · Richard H. Kimura ·
Youn-Hi Woo · Julio A. Camarero

Received: 30 June 2009 / Accepted: 30 July 2009 / Published online: 14 August 2009
© Springer-Verlag 2009

Abstract We present the in vivo biosynthesis of wild-type sunflower trypsin inhibitor 1 (SFTI-1) inside *E. coli* cells using an intramolecular native chemical ligation in combination with a modified protein splicing unit. SFTI-1 is a small backbone cyclized polypeptide with a single disulfide bridge. A small library containing multiple Ala mutants was also biosynthesized and its activity was assayed using a trypsin-binding assay. This study clearly demonstrates the exciting possibility of generating large cyclic peptide libraries in live *E. coli* cells, and is a critical first step for developing in vivo screening and directed evolution technologies using the cyclic peptide SFTI-1 as a molecular scaffold.

Keywords Bowman–Birk inhibitor · Trypsin inhibitor · Backbone cyclized peptides · Genetically encoded libraries · Protein splicing

Abbreviations

AMC 7-Amido-4-methyl-coumarin
CBD Chitin-binding domain
Cbz Benzyloxycarbonyl

CCK Cyclic cystine-knot motif
BBi Bowman–Birk inhibitor
EDTA Ethylenediaminetetraacetic acid
EtSH Ethanethiol
GdmCl Guanidinium hydrochloride
GSH Glutathione
HPLC High performance liquid chromatography
LB Luria–Bertani
MAP Methionyl aminopeptidase
MESNA Mercaptoethanesulfonic acid
NHS N-hydroxysuccinimide ester
PAGE Polyacrylamide gel electrophoresis
PMSF Phenylmethylsulfonyl fluoride
SDS Sodium dodecyl sulfate
SFTI-1 Sunflower trypsin inhibitor 1
TFA Trifluoroacetic acid
UV-Vis Ultraviolet-visible

Electronic supplementary material The online version of this article (doi:10.1007/s00726-009-0338-4) contains supplementary material, which is available to authorized users.

J. A. Camarero (✉)
Department of Pharmaceutical Sciences and Pharmacology,
School of Pharmacy, University of Southern California,
1985 Zonal Avenue, PSC 616, Los Angeles, CA 90033, USA
e-mail: camarej@usc.edu; jcamarer@usc.edu

J. Austin · R. H. Kimura · Y.-H. Woo · J. A. Camarero
Lawrence Livermore National Laboratory, Livermore,
CA 94550, USA

Introduction

In recent years, there has been an increased interest in discovering new cyclic peptides, and also in developing these backbone-cyclized peptides to regulate targeted biological activities (Hruby and Al-Obeidi 1990; Rizo and Gierasch 1992). Much of the current efforts are centered upon exploiting cyclic peptide topology as scaffolds for libraries, which are then screened for novel biological activity. To date, several hundreds of cyclic polypeptides (ranging from 14 to 70 residues long) have been discovered in a diverse range of organisms, which includes simple prokaryotes and higher primates (Craik et al. 2003, 2006a, b; Crovella et al. 2005). Backbone-cyclized peptides have

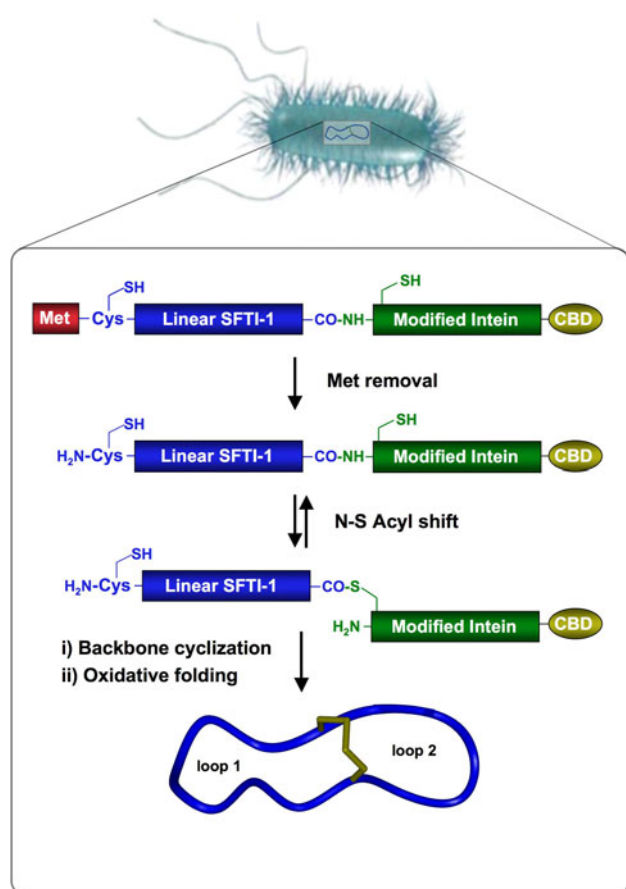
the N- and C-terminal ends joined together through a natural peptide bond. By eliminating the usually flexible terminal ends, these peptides have additional protection from proteolytic enzymes, which makes them less susceptible to attack by exopeptidases, thus increasing their *in vivo* lifetime (Trabi and Craik 2002). Some cyclic peptides are also noted for their exceptional stability and ability to withstand chemically harsh environments such as the gastrointestinal tract, and extreme physical environments such as boiling water. Moreover, peptide cyclization may help to stabilize structure in a manner analogous to a disulfide bond constraint; these rigid molecules then gain a thermodynamic or entropic advantage upon binding its target. Together, these characteristics suggest that cyclic peptides are ideal candidates to use as molecular scaffolds for peptide libraries, which can be screened for desired biological activity (Hilpert et al. 2000; Craik et al. 2002, 2006a, b).

In a recent study, we described a method to biosynthesize cyclic peptides using intramolecular native chemical ligation (NCL) in combination with a modified protein-

splicing unit or intein (Kimura et al. 2006; Camarero et al. 2007). This process requires the presence, within the same polypeptide sequence, of a N-terminal Cys residue and a C-terminal α -thioester function (Camarero and Muir 1999a, b; Camarero et al. 2001). Recombinant polypeptides containing N-terminal Cys residues can easily be produced using endogenous proteases such as methionyl aminopeptidases (MAP) (Camarero et al. 2001). Recombinant proteins containing α -thioesters can be obtained by using engineered inteins (Evans et al. 1998; Muir et al. 1998; Severinov and Muir 1998; Camarero and Muir 1999a, b). We recently used this approach to biosynthesize several cyclotides in *E. coli* (Kimura et al. 2006; Camarero et al. 2007).

Cyclotides are small globular microproteins with a unique head-to-tail cyclized backbone, which are stabilized by three disulfide bonds (Craik et al. 1999). The number and positions of cysteine residues are conserved throughout the family, forming the cyclic cystine-knot motif (CCK) (Craik et al. 1999). This CCK framework gives the cyclotides exceptional resistance to thermal and chemical denaturation and enzymatic degradation.

Based on these results we also decided to test if this technology could be also used to produce smaller cyclic peptides. For this purpose we decided to use sunflower trypsin inhibitor 1 (SFTI-1), a 14 amino acid backbone-cyclized peptide containing a single disulfide bond (Scheme 1). SFTI-1 is a naturally occurring protease inhibitor found in the seeds of sunflower (*Helianthus annuus*) (Luckett et al. 1999). SFTI-1 belongs to the Bowman-Birk inhibitor (BBI) family whose members, found in many plants, are potent serine protease inhibitors (Korsinczky et al. 2004). SFTI-1 is the smallest and most potent BBI, whose K_i has been reported to be in the low nanomolar range (Luckett et al. 1999). Structural analysis



Scheme 1 Biosynthetic approach for the production of cyclic peptide SFTI-1 inside living *E. coli* cells. Backbone cyclization of the linear precursor is mediated by a modified protein splicing unit or intein. The cyclized peptide then folds spontaneously in the bacterial cytoplasm (CBD stands for chitin-binding domain)

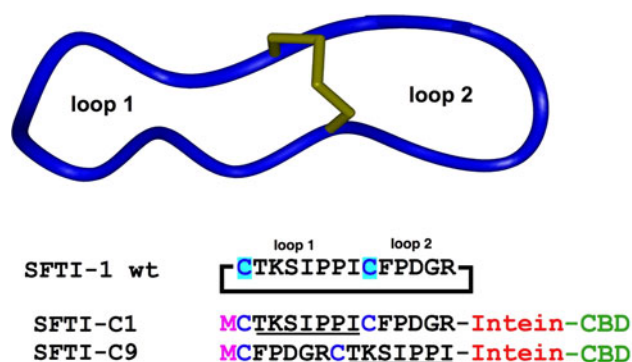


Fig. 1 Primary and tertiary structure of the cyclic peptide SFTI-1. Sequences of the linear precursors used for the *in vitro* and *in vivo* production of SFTI-1. The precursors contain an N-terminal Met residue and are fused in frame to the corresponding modified intein (See VMA or Mxe Gyase). The sequence corresponding to loop 1 is *underlined* as a reference (CBD stands for chitin-binding domain)

of the natural product reveals a well-defined double β -hairpin loop linked by two short antiparallel β -strands (Luckett et al. 1999; Korsinczky et al. 2001, 2005) (Fig. 1). Previous studies indicate that SFTI-1 readily adopts its simple native fold in aqueous solution without need for an oxidizing redox buffer system (Marx et al. 2003). Furthermore, by virtue of SFTI-1's ability to bind to and inhibit trypsin, it has a built in affinity handle, which may be used to selectively pull down SFTI-1 from a complex cytoplasmic mixture. Together, these characteristics suggest that this natural product is the ideal candidate upon which to develop and graft novel bioactivities important for biosecurity, human health or agriculture.

Materials and methods

General materials and methods

Analytical high performance liquid chromatography (HPLC) was performed on a HP1100 series instrument with 220 and 280 nm detection using a Vydac C18 column (5 μ m, 4.6 mm $\times$ 150 mm) at a flow rate of 1 mL/min. Semipreparative HPLC was performed on a Waters Delta Prep system fitted with a Waters 2487 Ultraviolet-Visible (UV-Vis) detector using a Vydac C18 column (15–20 μ m, 10 mm $\times$ 250 mm) at a flow rate of 5 mL/min. All runs used linear gradients of 0.1% aqueous trifluoroacetic acid (TFA, solvent A) vs. 0.1% TFA, 90% acetonitrile in H₂O (solvent B). UV-Vis spectroscopy was carried out on an Agilent 8453 diode array spectrophotometer. Electrospray mass spectrometry (ES-MS) analysis was routinely applied to all cyclized peptides. ES-MS was performed on a Sciex API-150EX single quadrupole electrospray mass spectrometer, MS/MS was performed on an Applied Biosystems API-3000 triple quadrupole mass spectrometer. Calculated masses were obtained by using ProMac v1.5.3. Protein samples were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Samples were run on 4–20% Tris–Glycine Gels (Invitrogen, Carlsbad, CA, USA). The gels were then stained with Pierce (Rockford, IL, USA) Gelcode Blue, photographed/digitized using a Kodak (Rochester, NY, USA) EDAS 290, and quantified using NIH Image-J software (<http://rsb.info.nih.gov/ij/>). Amino acid analysis was performed by the Molecular Structure Facility at University of California Davis on a Hitachi L-8800 amino acid analyzer. DNA sequencing was performed by Davis Sequencing (Davis, CA, USA) using an ABI 3730 DNA sequencer, and the sequence data was analyzed with DNASTar (Madison, WI, USA) Lasergene v5.5.2. All chemicals were obtained from Sigma-Aldrich (Milwaukee, WI, USA) unless otherwise indicated.

Construction of expression plasmids

Plasmids expressing the linear SFTI-1 precursors were constructed using either the pTYB1 or pTXB1 expression plasmids (New England Biolabs), which contain an engineered *Sce* VMA or *Mxe* Gyrase intein, respectively, and a chitin-binding domain (CBD). Oligonucleotides coding for the SFTI-1 wild type and mutant sequences (Table 1) were synthesized, phosphorylated and PAGE purified by IDT DNA (Coralville, IA, USA). Complementary strands were annealed in 0.3 M NaCl and the resulting double stranded DNA (dsDNA) was purified using Qiagen's (Valencia, CA, USA) miniprep column and buffer PN. pTXB1 or pTYB1 plasmids was double digested with *Nde*I and *Sap*I (NEB). The linearized vectors and the SFTI-1 dsDNA fragments were ligated at 15°C overnight using T4 DNA Ligase (New England Biolabs). The ligated plasmids were transformed into DH5 α cells (Invitrogen) and plated on Luria-Bertani (LB)-agar containing ampicillin. Positive colonies were grown in 5 mL LB containing ampicillin at 37°C overnight and the corresponding plasmids purified using a Miniprep Kit (Qiagen). Plasmids were initially screened by *Eco*RI digestion, as this restriction site is removed during cloning. Preliminary positives were expressed (see below) and fully characterized by ES-MS.

Expression and purification of recombinant proteins

E. coli BL21(DE3) or Origami(DE3) cells (Novagen, San Diego, CA, USA) were transformed with the SFTI-1 plasmids (see above). Expression was carried out in LB medium (1–2 L) containing ampicillin at room temperature or 30°C for 2 h to overnight. Briefly, 5 mL of an overnight starter culture derived from either a single clone or single plate (Ala-scan library) were used to inoculate 1 L of LB media. Cells were grown to an OD at 600 nm of $\approx$ 0.5 at 37°C, and expression was induced by the addition of isopropyl- β -D-thiogalactopyranoside to a final concentration of 0.5 mM at the temperatures and times mentioned above and in the manuscript. The cells were then harvested by centrifugation. For fusion protein purification, the cells were resuspended in 30 mL of lysis buffer (0.1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM phenylmethylsulfonyl fluoride, 50 mM sodium phosphate, 250 mM NaCl buffer at pH 7.2 containing 5% glycerol) and lysed by sonication. The lysate was clarified by centrifugation at 15,000 rpm in a Sorval SS-34 rotor for 30 min. The clarified supernatant was incubated with chitin-beads (2 mL beads/L cells, New England Biolabs), previously equilibrated with column buffer (0.1 mM EDTA, 50 mM sodium phosphate, 250 mM NaCl buffer at pH 7.2) at 4°C for 1 h with gentle rocking. The beads were extensively washed with 100 bead-volumes of column buffer containing 0.1%

Table 1 Forward (p5) and reverse (p3) 5'-phosphorylated oligonucleotides used to clone the different SFTI-based peptides

SFTI-C1	p5	TATGTGCACCAAATCTATCCCGCCGATCTGCTTCCCGGACGGTCGT
SFTI-C1	p3	GCAACGACCGTCCGGGAAGCAGATCGGCGGGATAGATTTGGTGCACA
T2A	p5	TATGTGCGCTAAATCTATCCCGCCGATCTGCTTCCCGGACGGTCGT
T2A	p3	GCAACGACCGTCCGGGAAGCAGATCGGCGGGATAGATTTAGCGCACA
K3A	p5	TATGTGCACCGTTCTATCCCGCCGATCTGCTTCCCGGACGGTCGT
K3A	p3	GCAACGACCGTCCGGGAAGCAGATCGGCGGGATAGAAGCGGTGCACA
S4A	p5	TATGTGCACCAAAGCTATCCCGCCGATCTGCTTCCCGGACGGTCGT
S4A	p3	GCAACGACCGTCCGGGAAGCAGATCGGCGGGATAGCTTTGGTGCACA
I5A	p5	TATGTGCACCAAATCTGCTCCGCGGATCTGCTTCCCGGACGGTCGT
I5A	p3	GCAACGACCGTCCGGGAAGCAGATCGGCGGAGCAGATTTGGTGCACA
P6A	p5	TATGTGCACCAAATCTATCGCTCCGATCTGCTTCCCGGACGGTCGT
P6A	p3	GCAACGACCGTCCGGGAAGCAGATCGGAGCGATAGATTTGGTGCACA
P7A	p5	TATGTGCACCAAATCTATCCCGGCTATCTGCTTCCCGGACGGTCGT
P7A	p3	GCAACGACCGTCCGGGAAGCAGATAGCCGGGATAGATTTGGTGCACA
I8A	p5	TATGTGCACCAAATCTATCCCGCCGGCTTCTGCTTCCCGGACGGTCGT
I8A	p3	GCAACGACCGTCCGGGAAGCAAGCCGGCGGGATAGATTTGGTGCACA
C9A	p5	TATGTGCACCAAATCTATCCCGCCGATCGCTTCCCGGACGGTCGT
C9A	p3	GCAACGACCGTCCGGGAAGCGATCGGCGGGATAGATTTGGTGCACA
F10A	p5	TATGTGCACCAAATCTATCCCGCCGATCTGCGCTCCCGGACGGTCGT
F10A	p3	GCAACGACCGTCCGGAGCGCAGATCGGCGGGATAGATTTGGTGCACA
I11A	p5	TATGTGCACCAAATCTATCCCGCCGATCTGCTTCGCTGACGGTCGT
I11A	p3	GCAACGACCGTCAGCGAAGCAGATCGGCGGGATAGATTTGGTGCACA
D12A	p5	TATGTGCACCAAATCTATCCCGCCGATCTGCTTCCCGGCTGGTCGT
D12A	p3	GCAACGACCGTCCGGGAAGCAGATCGGCGGGATAGATTTGGTGCACA
G13A	p5	TATGTGCACCAAATCTATCCCGCCGATCTGCTTCCCGGACGCTCGT
G13A	p3	GCAACGAGCGTCCGGGAAGCAGATCGGCGGGATAGATTTGGTGCACA
R14A	p5	TATGTGCACCAAATCTATCCCGCCGATCTGCTTCCCGGACGGTGCT
R14A	p3	GCAAGCACCGTCCGGGAAGCAGATCGGCGGGATAGATTTGGTGCACA
SFTI-C9	p5	TATGTGCACCAAATCTATCCCGCCGATCGCTTCCCGGACGGTCGT
SFTI-C9	p3	GCAACGACCGTCCGGGAAGCGATCGGCGGGATAGATTTGGTGCACA

Triton X100 and then rinsed and equilibrated with 100 bead-volumes of column buffer. In vivo cleavage was quantified by SDS-PAGE analysis of the purified fusion proteins using the NIH Image-J software package.

Concomitant cleavage, cyclization and folding of SFTI-1 with EtSH, MESNA and GSH

Chitin beads containing the different purified SFTI-1-In-tein-CBD fusion proteins were cleaved with 1% ethanethiol (EtSH), 50 mM mercapto-ethanesulfonate (MESNA) or 50 mM glutathione (GSH) in degassed column buffer. The cyclization/folding reactions were kept for up to 4 days at 25°C with gentle rocking. For small scale reactions, aliquots were taken each day (when necessary) and analyzed by HPLC. The reduced and oxidized circular SFTI-1 product was analyzed by ES-MS (Table 2). Larger-scale (2 L culture) purification of the SFTI-1 wild type, T2A, K3A linear precursors were also performed. The

supernatant of the cyclization reaction was separated by filtration and the beads were washed with additional column buffer (1 column volume per each milliliter of beads). The supernatant and washes were pooled, and the oxidized-circular peptide was purified by semipreparative HPLC using a linear gradient of 15–45% solvent B over 30 min. The isolated yield for purified the SFTI-1 peptides was around 300 µg/L.

Purification of SFTI-1 using trypsin-sepharose beads

Preparation of trypsin-sepharose beads: N-hydroxy-succinimide ester-activated Sepharose was washed with 15 volumes of ice-cold 1 mM HCl. Each volume of beads was incubated with an equal volume of coupling buffer (250 mM NaCl, 200 mM sodium phosphate buffer at pH 6.0) containing 2 mg of Porcine Pancreas Trypsin type IX-S (14,000 U/mg) for 3 h with gentle rocking at room temperature. The beads were then rinsed with 10 volumes

Table 2 Sequences and molecular weights found for the different SFTI-based peptides used in this work

Name	Sequence	Expected mass (Da) (average)		Observed mass ^b (Da)
		Circular ^a	Linear ^a	
SFTI-1 wt	CTKSIPPICFPDGR	1513.81	1531.81	1512.5 ± 0.5
T2A	CAKSIPPICFPDGR	1483.79	1501.79	1483.7 ± 0.4; 1503.5 ± 0.2 ^c
K3A	CTASIPPICFPDGR	1456.72	1474.72	1456.4 ± 0.4
S4A	CTKAIPPICFPDGR	1497.81	1515.81	1497.7 ± 0.3; 1515.0 ± 0.4 ^c
I5A	CTKSAPPICFPDGR	1471.73	1489.73	1471.3 ± 0.5
P6A	CTKSIAPICFPDGR	1487.78	1505.78	1487.5 ± 0.3
P7A	CTKSIPAICFPDGR	1487.78	1505.78	1487.5 ± 0.3
I8A	CTKSIPPACFPDGR	1471.73	1489.73	1471.3 ± 0.5
C9A	CTKSIPPIAFPDGR	1483.75	1501.75	1483.7 ± 0.4
F10A	CTKSIPPICAPDGR	1437.72	1455.72	1437.6 ± 0.2
P11A	CTKSIPPICFADGR	1487.78	1505.78	1487.5 ± 0.3
D12A	CTKSIPPICFPAGR	1469.80	1487.80	1469.5 ± 0.2
G13A	CTKSIPPICFPDAR	1527.84	1545.84	1528.3 ± 0.4
R14A	CTKSIPPICFPDGA	1428.70	1446.70	1428.5 ± 0.3

Identical values, within error, were obtained for in vivo cyclized peptides

^a Mass corresponds to the oxidized peptide except for mutant C9A

^b Mass found for in vitro cyclized peptides

^c Mass corresponds to hydrolyzed peptide

of coupling buffer, and incubated with excess coupling buffer containing 100 mM ethanolamine (Eastman Kodak) for 3 h with gentle rocking at room temperature. Finally, the beads were washed with 50 volumes of wash buffer (200 mM sodium acetate buffer at pH 3, 250 mM NaCl) and stored in 1 volume of wash buffer.

Thirty microliters of clarified lysate was incubated with 0.2 mL of trypsin-sepharose for 1 h at room temperature with gentle rocking, and centrifuged at 3,000 rpm for 1 min. The beads were washed with 50 volumes of PBS containing 0.1% Triton X100, then rinsed with 50 volumes of PBS, and drained of excess PBS. After the addition of 0.5 mL of 8 M Gdm-HCl, the reaction was incubated in a 90°C heat block for 5 min to elute the bound peptide. The peptide was analyzed by reversed-phase HPLC and ES-MS/MS.

Determination of K_i

Sunflower trypsin inhibitor 1 concentrations were quantified by amino acid analysis (UC Davis Molecular Structure Facility). Trypsin inhibitory assay was performed in a solution containing 20 mM CaCl₂, 50 mM Tris-HCl buffer at pH 8.0. Inhibitor solutions of SFTI-1 ranging from 10 to 100 nM were preincubated with a freshly prepared solution of 2 nM trypsin in a final reaction volume of 3.0 mL. After incubation for 20 min, residual trypsin activity was measured by adding the fluorogenic substrate N^z-Cbz-L-Arg-AMC (AMC: 7-amido-4-methyl-coumarin, Cbz: benzyl-oxycarbonyl) to a final concentration of 2 μM and then following the release of AMC using a Fluorolog-3 spectrofluorometer (Jobin Yvon Inc., Edison, NJ, USA). Measurements were taken every 60 s using an excitation

wavelength of 366 nm and an emission wavelength of 465 nm. The initial velocities for the hydrolysis of substrate N^z-Cbz-L-Arg-AMC by trypsin in the presence of different concentrations of SFTI-1 were fitted to a one-site competitive binding equation using the software package Prism (GraphPad Software). K_i was calculated using the equation of Cheung and Prusoff (1973) and a K_m value of 12 μM (Baird et al. 2000) (Supplementary Figure 2).

Results

Biosynthesis of SFTI-1 via Cys¹ and Cys⁹

Sunflower trypsin inhibitor 1 is a relatively simple cyclic peptide that contains two cysteine residues that may be used for NCL. Two different SFTI-1 linear precursors were cloned in frame with a modified Sce VMA intein. The N-terminal methionine residue is efficiently removed by the endogenous MAP, which yields the required N-terminal Cys residue required for intramolecular NCL. At the same time, the modified intein generates the α-thioester at the junction between C-terminal end of the SFTI-1 and the N-terminal Cys of the intein (Fig. 1). Linear versions of SFTI-1 have shown to adopt a native-like fold, so we anticipated that this conformational predisposition would bring the N-terminal Cys in close proximity to the α-thioester thus facilitating the intramolecular cyclization through NCL.

Sunflower trypsin inhibitor 1 intein fusion proteins (SFTI-C1 and SFTI-C9) were expressed in *E. coli* BL21(DE3) cells, purified by affinity chromatography using a CBD and analyzed by SDS-PAGE. As shown in

Fig. 2 Biosynthesis of SFTI-1 using a modified Sce VMA intein. **a** SDS-PAGE analysis of the expression levels and in vivo cleavage of precursors SFTI-C1 and SFTI-C9. **b** Analytical reversed-phase HPLC trace of the in vitro cyclization/folding of SFTI-C1 with 50 mM GSH after 24 h. **c** Analytical reversed-phase HPLC trace of the soluble cellular fraction of *E. coli* BL21(DE3) cells expressing precursor SFTI-C1 after it was purified on trypsin-immobilized sepharose beads. The arrow denotes the natively folded SFTI-1 peptide. The large peak eluting at 16–17 min and small peaks around SFTI-1 are not related to SFTI-1 or SFTI-intein precursor. **d** ES-MS spectrum of affinity purified SFTI-1 peptide. HPLC gradients are 0–70% solvent B over 20 min (see Sect. "Materials and methods" for buffer composition)

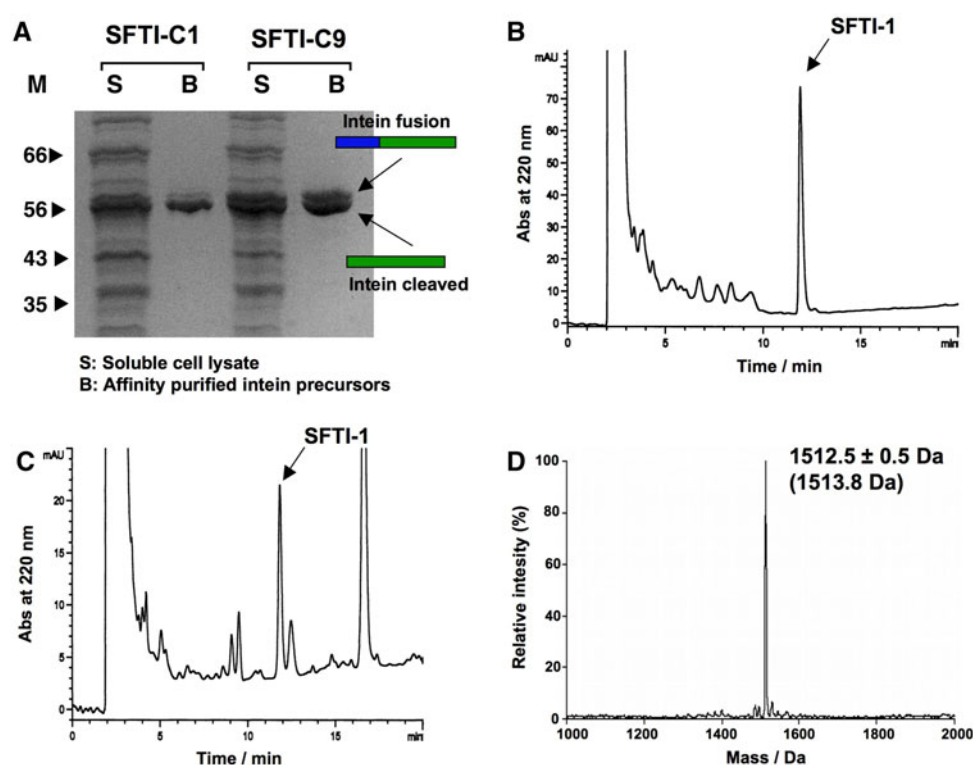


Fig. 2a, both intein precursors had comparable levels of expression in *E. coli* cells, however they showed different rates and propensities for in vivo cleavage. SFTI-C1 shows $\approx 80\%$ in vivo cleavage, and SFTI-C9 shows $\approx 70\%$ in vivo cleavage (Fig. 2a). This difference could be attributed to the close proximity of two consecutive Pro residues to the extein–intein junction in the precursor SFTI-C9, which may partially affect the equilibrium between the thioester and amide bonds at the junction.

Next, we investigated in vitro cyclization of intein-containing precursor proteins. Affinity purified fusion precursors SFTI-C1 and SFTI-C9 were incubated in phosphate buffer at pH 7.2 containing either 1% EtSH, 50 mM sodium MESNA or 50 mM reduced GSH to facilitate intramolecular cyclization. After 12 h of reaction, SFTI-C1 was completely cleaved in vitro by any of the thiols tested, whereas SFTI-C9 showed no additional cleavage even after 48 h of reaction.

Reversed-phase HPLC and ES-MS analysis confirmed the presence of both reduced-circular SFTI-1 and native SFTI-1 when using SFTI-C1, but neither reduced nor cyclic peptide using the SFTI-C9 precursor. This initial in vitro analysis indicated that the SFTI-C1 linear precursor is more efficient for intramolecular cyclization. Interestingly, when the cyclization reaction was performed in the presence of GSH, the only product obtained was the native SFTI-1 as indicated by ES-MS (Fig. 2b, d). This could be attributed to the partial air oxidation of GSH during the

reaction to give a GSSG/GSH redox buffer that facilitated the formation of the native disulfide between Cys¹ and Cys⁹. This effect has been also observed in the cyclization of cyclotides using GSH-containing buffers (Kimura et al. 2006; Camarero et al. 2007). The obtained SFTI-1 was purified and its biological activity tested using a trypsin inhibitory assay (Jaulent and Leatherbarrow 2004; Camarero et al. 2007). In vitro biosynthesized SFTI-I was able to inhibit pancreatic trypsin with a K_i of 26 ± 2 nM (Supplementary Figure 2). This K_i constant closely parallels previously reported values, thus confirming the native fold of recombinant SFTI-1.

In vivo biosynthesis of SFTI-1

Due to the highly efficient cleavage of precursor SFTI-C1 and the high propensity to cyclize and fold in the presence of GSH-containing buffers, we explored whether any folded SFTI-1 was formed in vivo. We used trypsin-sepharose to capture folded SFTI-1 formed inside living cells expressing the SFTI-C1 precursor. After extensive washings, the bound fraction was eluted with 8 M guanidinium hydrochloride solution. HPLC analysis of the bound fraction revealed the presence of a major peak with the same retention time and mass as that the folded SFTI-1, thus confirming the in vivo biosynthesis of SFTI-1 inside living *E. coli* cells (Fig. 2c). Approximately 250 μ g of folded SFTI-1 were produced per gram of wet *E. coli* cells. This

corresponds to an intracellular concentration of 100 nM (Camarero et al. 2007). Interestingly, although SFTI-C9 is cleaved efficiently in vivo ($\approx 70\%$), no cyclic folded SFTI-1 was found in the corresponding lysate (data not shown).

Encouraged by these results, we tested in vivo cyclization of SFTI-C1 using the modified Mxe Gyrase intein. This mini-intein is a bacterial intein that can be expressed very efficiently in *E. coli* cells at 30°C. The Sce VMA intein starts to accumulate to high levels in the insoluble cellular fraction when expressed at 30°C, thus reducing the yield of folded intein. The expression of the Mxe Gyrase intein, however, can be induced at this temperature for 6 h without any significant aggregation. On average, the Gyrase intein is able to produce between two and three times more precursor protein than the corresponding VMA intein precursor.

Using the Mxe Gyrase intein system, close to 70% of the SFTI-C1 precursor was cleaved in vivo as indicated by SDS-PAGE analysis (Supplementary Figure 1). This value is similar to that found for the Sce VMA intein. Analysis of the cell lysate using trypsin-sepharose affinity chromatography showed the presence of folded SFTI-1 in *E. coli* cytoplasm (Supplementary Figure 1). The intracellular concentration of folded SFTI-1 was estimated to be approximately 1 μM . A similar value has been observed for intracellular biosynthesis of the cyclotide MCotI-II in *E. coli* cells using the Mxe Gyrase intein (Camarero et al. 2007).

Biosynthesis of an SFTI-based library

In order to evaluate whether this biosynthetic approach may produce genetically encoded libraries of SFTI mutants, we performed alanine scanning mutagenesis to determine whether this stable double β -hairpin scaffold could accommodate amino acid substitutions throughout its primary structure, retain its bioactivity and importantly, whether these mutants can be biosynthesized in vivo.

A small library of 14 Ala mutants was designed, such that every amino acid, with the exception of Cys¹, was sequentially substituted by Ala (Table 2). Since Cys¹ was found to be the better of the two cysteines for NCL, the wild type member as well as each Ala-mutant member incorporates this cysteine at the N-terminal end of the corresponding intein precursors. Synthetic dsDNA fragments encoding the different SFTI-1 mutants were ligated into plasmid pTXB1 in frame with Mxe Gyrase intein. The resulting plasmid library was transformed into competent DH5 α *E. coli* cells obtaining approximately 10^4 colonies (data not shown). All colonies were pooled and the corresponding plasmid library was transformed into different *E. coli* expression cell lines.

Expression of the library in *E. coli* BL21(DE3) cells produced the corresponding Gyrase intein precursors with similar yields to that of the wild-type SFTI-1. The level of in vivo cleavage was estimated to be $\approx 70\%$ following induction for 5 h at 30°C. Interestingly, when SFTI-Gyrase encoded plasmid library was expressed in *E. coli* Origami(DE3) cells under the same conditions (5 h at 30°C), only $\approx 20\%$ in vivo cleavage was observed (data not shown). Origami cells have a more oxidizing cytoplasm, which may favor the formation of disulfide bonds over the competing intramolecular NCL reaction. We have found, however, that in both cellular backgrounds (BL21 and Origami cells), the in vivo cleavage ratio can be increased by inducing the culture at lower temperatures for longer times (Camarero et al. 2007). This allows for experimental to accommodate different screening methods. For in vitro screening, cyclization can be accomplished in vitro under controlled conditions, and therefore short induction times at relatively higher temperatures will yield more uncleaved linear precursor. Alternatively, for high throughput in vivo screening, cyclization yields must be maximized. Levels of in vivo cleavage may be increased by longer induction times and lower induction temperatures (18–20 h at 20°C for example).

In order to characterize the SFTI-based library, uncleaved SFTI-Gyrase fusion proteins were obtained from *E. coli* BL21(DE3) cells which were induced at 30°C for 5 h. The uncleaved fusion precursors ($\approx 30\%$ of the total induced protein) were cyclized with phosphate buffer at pH 7.2 containing 50 mM GSH for 24 h. This treatment resulted in nearly 100% cleavage of the intein precursors. The soluble fraction was analyzed by HPLC and ES-MS/MS to determine the presence of relative representation of library members (Fig. 3a). All library members were observed with the exception of D12A. This mutant was found in the cell lysate (see results below). Mutants with the same or similar molecular weight were distinguished by MS/MS and/or differences in their HPLC retention times.

In vivo biosynthesis and screening of SFTI-based library

In vivo biosynthesis of the SFTI-library was performed in *E. coli* BL21(DE3) cells by inducing the expression of the SFTI-intein fusion library at 20°C for 18 h. Under these conditions, in vivo cleavage was close to 95%. The soluble cellular fraction was incubated with trypsin-sepharose, and the bound fraction was analyzed by HPLC and ES-MS. As shown in Fig. 3b, the MS analysis was able to detect the SFTI-1 wt, and the mutants G13A, R14A, S4A, I5A, I8A, P7A, P11A, F10A and D12A. Visual inspection of the structural complex formed between SFTI-1 and trypsin (Luckett et al. 1999) indicates that the side-chains of these

residues are solvent exposed. Thus, these residues may be more tolerable to amino acid substitution with relatively minor consequences to biological activity (Luckett et al. 1999). For example, a study by (Descours et al. 2002) showed that the K_i values of mutants P8A and P13A are only slightly affected, and I7A and I9A are only moderately affected with K_i values ranging from picomolar to low micromolar. These findings are also corroborated by the results reported by (Hilpert et al. 2000) on the complete substitutional analysis of SFTI-1 with different serine proteases and more recently by (Daly et al. 2006), where a complete suite of Ala mutants of SFTI-1 was chemically synthesized and characterized structurally and functionally.

As expected, the SFTI-1 K3A was not found in the trypsin-bound fraction. This residue determines binding affinity and specificity, and can only be replaced by Arg to maintain biological activity (Hilpert et al. 2000). SFTI-1 mutants P6A and C9A were also not found in the trypsin-bound fraction. In agreement with Hilpert et al. (2000), bioactivity of Pro⁶ mutants is maintained only by Asp or

Glu substitutions. A more recent study by (Daly et al. 2006) showed the importance of this residue as well. Mutation of Cys⁹ to Ala is also known to disrupt the biological activity due to the necessary disulfide bond (Korsinczky et al. 2005).

Interestingly, two masses corresponding to linear Ala mutants were also found in the *in vivo* trypsin-bound fraction. Linear oxidized S4A binds trypsin beads demonstrating not only tight binding of this library member, but suggests an equilibrium between circular and linear (hydrolyzed) forms. In addition, a mass corresponding to the linear-oxidized T2A was also detected. Previously, it has been determined that an active acyclic permutant of SFTI-1 which opens at its scissile bond (i.e., Lys³-Ser⁴) can be enzymatically backbone-cyclized by trypsin. The resulting ratio of cyclic SFTI-1 to hydrolyzed SFTI-1 is $\approx 10:1$ regardless of whether trypsin is incubated with the acyclic or cyclic form of SFTI-1 (Marx et al. 2003). Therefore, the results of mutants T2A and S4A may be attributed to the proximity of these mutations with respect to the scissile bond, which may stabilize the hydrolyzed form.

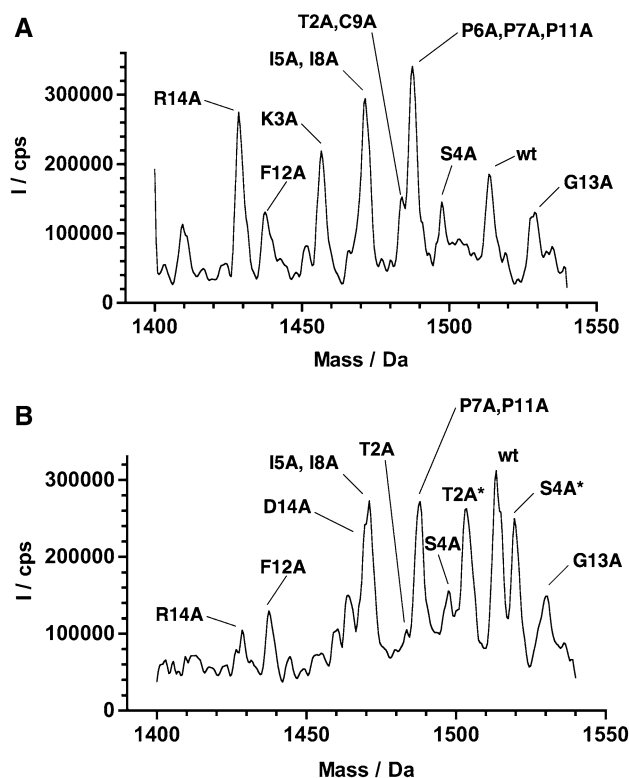


Fig. 3 Analysis of the SFTI-based library expressed in *E. coli* BL21(DE3) cells. **a** ES-MS spectrum of the SFTI-based library generated *in vitro* by cyclization/folding using 50 mM GSH. **b** ES-MS spectrum of the soluble cellular fraction of *E. coli* BL21(DE3) cells expressing precursor SFTI-C1 using a modified Mxe Gyrase after it was purified on trypsin-immobilized sepharose beads. Peptides hydrolyzed by trypsin are indicated with an asterisk

Discussion

A substantial number of natural products with wide range of pharmacological activities are cyclic polypeptides. Peptide cyclization is widely adopted in medicinal chemistry to improve the biochemical and biophysical properties of peptide-based drug candidates (Hruby and Al-Obeidi 1990; Rizo and Gierasch 1992). Among the different approaches used to cyclize polypeptides, backbone or head-to-tail cyclization remains one of the most extensively used to introduce structural constraints into biologically active peptides. Cyclization generally increases stability and rigidifies structure, thereby minimizing the entropic cost of receptor binding.

Chemical synthesis of cyclic peptides has been well explored and a number different approaches involving solid-phase or liquid-phase have been reported (Camarero and Muir 1997; Zhang and Tam 1997; Camarero et al. 1998a, b; Shao et al. 1998). Recent developments in the fields of molecular biology and protein engineering also allow biosynthesis of cyclic peptides.

The biosynthesis of cyclic polypeptides offers many advantages over purely synthetic methods. Using the tools of molecular biology, large combinatorial libraries of cyclic peptides may be generated and screened *in vivo*. A typical chemical synthesis may generate 10^4 different molecules. It is not uncommon for a recombinant library to contain as many as 10^9 members. The molecular diversity generated by this approach is analogous to phage-display

technology. Moreover, this approach takes advantage of the enhanced pharmacological properties of backbone-cyclized peptides as opposed to linear peptides or disulfide-stabilized polypeptides. However, unlike phage-display, backbone-cyclized polypeptides are not fused to or displayed on the surface, but remain on the inside of the living cell where they can be further screened for biological activity. The complex cellular cytoplasm provides the appropriate environment to address the physiological relevance of potential leads.

Protein *trans*-splicing has been successfully used by Benkovic and co-workers to generate backbone cyclized or polypeptides in vivo (Scott et al. 1999). In this approach, the peptide to be cyclized is nested between the two split intein fragments of the naturally occurring *Ssp* DnaE split intein (Wu et al. 1998) (usually referred as N- and C-inteins) in such way that the N-terminus of the peptide template is fused to C-intein fragment and vice versa. Protein splicing of this chimeric protein leads to the formation of the desired cyclic peptide inside *E. coli* cells. A potential limitation of this approach, however, is the requirement for specific N- and C-extein residues at the intein junction sites (Evans et al. 2000). These amino acids are necessary for efficient protein splicing to occur, which restricts the sequence diversity within the sequence of the cyclic peptide library members.

An attractive alternative approach is the use of modified protein splicing units to perform an intramolecular version of NCL reaction in vivo (Camarero and Muir 1999a, b; Iwai and Pluckthum 1999; Evans et al. 2000; Camarero et al. 2001). Here we have shown the biosynthesis inside living *E. coli* cells of the cyclic peptide SFTI-1 and a small SFTI-based peptide library containing different Ala mutants using a modified protein-splicing unit. SFTI-1 is a bicyclic 14-amino acid peptide isolated from sunflower seeds that is related to the BBI and is one of the most potent inhibitors of trypsin of any naturally occurring peptide (Luckett et al. 1999). SFTI-1 forms a tightly folded scaffold, either when complexed with trypsin or free in solution (Luckett et al. 1999; Korsinczky et al. 2001). Its compact structure (Fig. 1) and high potency have led to suggestions that it may serve as a scaffold for the design of novel peptide-based drug leads (Korsinczky et al. 2004).

In vivo biosynthesis of small backbone cyclic peptides like SFTI-1 containing a disulfide bond may have tremendous potential for drug discovery. This study shows that SFTI-1 may provide an ideal scaffold for the biosynthesis of large combinatorial libraries inside of living *E. coli* cells. Coupled to an appropriate in vivo reporter system, this library may rapidly be screened using high throughput technologies such as fluorescence activated cell sorting (Kimura et al. 2007).

Acknowledgments Work was supported by funding from the School of Pharmacy at the University of Southern California and Lawrence Livermore National Laboratory.

References

- Baird T, Wang B, Lodder M, Hecht S, Craik CS (2000) Generation of active trypsin by chemical cleavage. *Tetrahedron* 56:9477–9485
- Camarero JA, Muir TW (1997) Chemoselective backbone cyclization of unprotected peptides. *J Chem Soc Chem Commun* 1997:1369–1370
- Camarero JA, Muir TW (1999a) Biosynthesis of a head-to-tail cyclized protein with improved biological activity. *J Am Chem Soc* 121:5597–5598
- Camarero JA, Muir TW (1999b) Native chemical ligation of polypeptides. *Curr Protoc Protein Sci* 18(4):1–21
- Camarero JA, Cotton GJ, Adeva A, Muir TW (1998a) Chemical ligation of unprotected peptides directly form a solid support. *J Pept Res* 51:303–316
- Camarero JA, Pavel J, Muir TW (1998b) Chemical synthesis of a circular protein domain: evidence for folding-assisted cyclization. *Angew Chem Int Ed* 37:347–349
- Camarero JA, Fushman D, Cowburn D, Muir TW (2001) Peptide chemical ligation inside living cells: in vivo generation of a circular protein domain. *Bioorg Med Chem* 9:2479–2484
- Camarero JA, Kimura RH, Woo YH, Shekhtman A, Cantor J (2007) Biosynthesis of a fully functional cyclotide inside living bacterial cells. *Chembiochem* 8:1363–1366
- Cheng Y, Prusoff WH (1973) Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I₅₀) of an enzymatic reaction. *Biochem Pharmacol* 22:3099–3108
- Craik DJ, Daly NL, Bond T, Waine C (1999) Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J Mol Biol* 294:1327–1336
- Craik DJ, Simonsen S, Daly NL (2002) The cyclotides: novel macrocyclic peptides as scaffolds in drug design. *Curr Opin Drug Discov Dev* 5:251–260
- Craik DJ, Daly NL, Saska I, Trabi M, Rosengren KJ (2003) Structures of naturally occurring circular proteins from bacteria. *J Bacteriol* 185:4011–4021
- Craik DJ, Cemazar M, Daly NL (2006a) The cyclotides and related macrocyclic peptides as scaffolds in drug design. *Curr Opin Drug Discov Dev* 9:251–260
- Craik DJ, Cemazar M, Wang CK, Daly NL (2006b) The cyclotide family of circular miniproteins: nature's combinatorial peptide template. *Biopolymers* 84:250–266
- Crovella S, Antcheva N, Zelezetsky I, Boniotti M, Pacor S, Verga Falzacappa MV, Tossi A (2005) Primate beta-defensins—structure, function and evolution. *Curr Protein Pept Sci* 6:7–21
- Daly NL, Chen YK, Foley FM, Bansal PS, Bharathi R, Clark RJ, Sommerhoff CP, Craik DJ (2006) The absolute structural requirement for a proline in the P3'-position of Bowman-Birk protease inhibitors is surmounted in the minimized SFTI-1 scaffold. *J Biol Chem* 281:23668–23675
- Descours A, Moehle K, Renard A, Robinson JA (2002) A new family of beta-hairpin mimetics based on a trypsin inhibitor from sunflower seeds. *Chembiochem* 3:318–323
- Evans TC, Benner J, Xu M-Q (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* 7:2256–2264
- Evans TC Jr, Martin D, Kolly R, Panne D, Sun L, Ghosh I, Chen L, Benner J, Liu XQ, Xu MQ (2000) Protein *trans*-splicing and

- cyclization by a naturally split intein from the *dnaE* gene of *Synechocystis* species PCC6803. *J Biol Chem* 275:9091–9094
- Hilpert K, Hansen G, Wessner H, Schneider-Mergener J, Hohne W (2000) Characterizing and optimizing protease/peptide inhibitor interactions, a new application for spot synthesis. *J Biochem* 128:1051–1057
- Hruby VJ, Al-Obeidi F (1990) Emerging approaches in the molecular design of receptor-selective peptide ligands: conformational, topographical and dynamic considerations. *J Biochem* 268:249–262
- Iwai H, Pluckthum A (1999) Circular β -lactamase: stability enhancement by cyclizing the backbone. *FEBS Lett* 459:166–172
- Jaulent AM, Leatherbarrow RJ (2004) Design, synthesis and analysis of novel bicyclic and bifunctional protease inhibitors. *Protein Eng Des Sel* 17:681–687
- Kimura RH, Tran AT, Camarero JA (2006) Biosynthesis of the cyclotide kalata B1 by using protein splicing. *Angew Chem Int Ed Engl* 45:973–976
- Kimura RH, Steenblock ER, Camarero JA (2007) Development of a cell-based fluorescence resonance energy transfer reporter for *Bacillus anthracis* lethal factor protease. *Anal Biochem* 369:60–70
- Korsinczky ML, Schirra HJ, Rosengren KJ, West J, Condie BA, Otvos L, Anderson MA, Craik DJ (2001) Solution structures by 1H NMR of the novel cyclic trypsin inhibitor SFTI-1 from sunflower seeds and an acyclic permutant. *J Mol Biol* 311:579–591
- Korsinczky ML, Schirra HJ, Craik DJ (2004) Sunflower trypsin inhibitor-1. *Curr Protein Pept Sci* 5:351–364
- Korsinczky ML, Clark RJ, Craik DJ (2005) Disulfide bond mutagenesis and the structure and function of the head-to-tail macrocyclic trypsin inhibitor SFTI-1. *Biochemistry* 44:1145–1153
- Luckett S, Garcia RS, Barker JJ, Konarev AV, Shewry PR, Clarke AR, Brady RL (1999) High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *J Mol Biol* 290:525–533
- Marx UC, Korsinczky ML, Schirra HJ, Jones A, Condie B, Otvos L Jr, Craik DJ (2003) Enzymatic cyclization of a potent Bowman-Birk protease inhibitor, sunflower trypsin inhibitor-1, and solution structure of an acyclic precursor peptide. *J Biol Chem* 278:21782–21789
- Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci USA* 95:6705–6710
- Rizo J, Gierasch LM (1992) Constrained peptides: models of bioactive peptides and protein substructures. *Annu Rev Biochem* 61:387–418
- Scott CP, Abel-Santos E, Wall M, Wahnou D, Benkovic SJ (1999) Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci USA* 96:13638–13643
- Severinov K, Muir TW (1998) Expressed protein ligation, a novel method for studying protein–protein interactions in transcription. *J Biol Chem* 273:16205–16209
- Shao Y, Lu WY, Kent SBH (1998) A novel method to synthesize cyclic peptides. *Tetrahedron Lett* 39:3911–3914
- Trabi M, Craik DJ (2002) Circular proteins—no end in sight. *Trends Biochem Sci* 27:132–138
- Wu H, Hu Z, Liu XQ (1998) Protein *trans*-splicing by a split intein encoded in a split *DnaE* gene of *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci USA* 95:9226–9231
- Zhang L, Tam JP (1997) Synthesis and application of unprotected cyclic peptides as building blocks for peptide dendrimers. *J Am Chem Soc* 119:2363–2370