Proceedings of the 22nd American Peptide Symposium
Michal Lebl (Editor)
American Peptide Society, 2011

Sunflower Trypsin Inhibitor-1: Chemical v. Biological Synthesis

Alysha G. Elliott, David J. Craik, and Joshua S. Mylne
Institute for Molecular Bioscience, The University of Queensland, St Lucia, 4072, Australia

Introduction

Diverse gene-encoded cyclic peptides are produced by species from all three domains of life. The 14-residue, head-to-tail cyclized plant peptide, SFTI-1 (Sunflower Trypsin Inhibitor 1) [1] (Figure 1) has attracted much attention due to its great stability and capability to potently inhibit trypsin (Ki 0.1 nM) as well as the epithelial serine protease matriptase (Ki 0.92 nM) [2], giving it exciting promise as a drug lead and a protein engineering scaffold [2]. Although peptides such as SFTI-1 are routinely produced by chemical synthesis, the biological mechanisms that enable biosynthesis of ribosomally synthesized cyclic peptides are largely unknown.

Recently Mylne *et al.* described the biosynthetic origin of SFTI-1 and a related peptide SFT-L1 [3]. Both emerge from seed storage protein precursors PawS1 and PawS2, respectively, using each seed protein's own maturing protease for their release [3]. We used transgenic constructs in the model plant *Arabidopsis thaliana* combined with proteomics and MALDI mass spectrometry to study this unusual dual-fate for PawS1 and identified the residues that are critical for SFTI-1 maturation and cyclisation [3]. Here we compare best practice for chemical synthesis of SFTI-1 with how plants biologically create the same product.

Fig. 1. Sequence of SFTI-1 (left) and a model of its backbone structure (PDB 1SFI). When chemically synthesized, SFTI-1 ends with an N-terminal Cys, but its biological ligation point is between Gly1 and Asp14.

Results and Discussion

There is a long history of the synthesis of cyclic peptides by manual solid-phase peptide synthesis [4]. Here we describe the optimized synthesis of the 14 residue cyclic peptide, SFTI-1 using BOC (t-butoxycarbonyl) chemistry with *in situ* neutralization, HBTU [2-(1-H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate] [5] and a C-terminal thioester linker.

Peptides were assembled on a PAM-Gly-Boc (phenylacetamidomethyl-glycine-t-Boc) resin on a 0.5-mmol scale and designed to contain the C-terminal thioester linker, S-trityl-β-mercaptopropionic acid, and an N-terminal cysteine. Peptides were subsequently cleaved from the resin by HF cleavage with p-cresol as the scavenger (9:1 HF:p-cresol by volume). The cleavage reaction was incubated at <5 to 0°C for 90 min; HF was removed under vacuum and the peptide precipitated with diethyl ether, then filtered and re-dissolved in 50% acetonitrile 0.045% TFA and lyophilised. The crude peptide was subsequently purified by RP-HPLC using a gradient of 0–80% acetonitrile 0.045% TFA over 80 min. Analytical HPLC and ESI-MS confirmed peptide purity and mass.

Importantly, the folding for SFTI-1 is a two-step process. The two steps ensure that the more favorable reaction, disulfide bond formation occurs secondarily to cyclization, the less favorable reaction, as the free sulfur donated from the cysteine is required to form a thiol for cyclization.

Firstly, the peptides were reduced and cyclized in 0.1 M ammonium bicarbonate pH 8.2 with 0.5 mg/mL TCEP overnight at room temperature, followed by RP-HPLC purification. Secondly, the peptides were oxidized in 0.1 M ammonium bicarbonate pH 8.2 overnight at room temperature, followed by RP-HPLC purification as above to yield fully folded and pure peptides. From ~200 mg of crude cleaved peptide ~40 mg of over 98% pure fully folded SFTI-1 may be obtained using this method.

Sunflowers produce SFTI-1 with ease. We cloned the precursor gene for SFTI-1, *PawS1,* and were surprised to find that the sequence for SFTI-1 was encoded along with a much larger

protein of entirely different function. PawS1 is a preproalbumin, which encodes a seed storage albumin protein as well as SFTI-1 [3]. Seed storage albumins are matured from proalbumins by the action of a Cys-protease called asparaginyl endopeptidase (AEP, aka legumain, vacuolar processing enzyme). AEP usually cleaves at Asn and, to a lesser extent, at Asp. Within PawS1, SFTI-1 ends with Asp and is preceded by Asn suggesting AEP matures both the albumin and SFTI-1. We used an *Arabidopsis* mutant lacking AEP to confirm that AEP was required to release SFTI-1 at both proto-termini and is the best candidate for the ligation reaction.

Furthermore, an in-depth mutagenesis of *PawS1* and subsequent testing of these constructs *in vivo* revealed residues essential for SFTI-1 processing. These experiments combined with the wealth of information about albumin maturation [6] allowed us to propose a model for the processing of SFTI-1 from within PawS1 (Figure 2) [3].

PawS preproalbumin is sent to the ER where its signal peptide is cleaved and with the aid of hairpin formation from hydrophobic clustering within SFTI-1 the Cys38-Cys46 disulfide bond is formed. The proalbumin is then matured at several points by AEP in multivesicular bodies as they traffic to protein storage vacuoles. During this processing SFTI-1 is released at the N-terminus (PawS1 Gly36) and during the final cleavage at its C-terminal P1 aspartic acid, a reactive, thioester acyl-intermediate is created. Instead of the typical attack by water and bond hydrolysis, we propose this reactive intermediate is instead attacked at its carbonyl carbon by the unmasked amino terminus of the glycine, held in proximity to the thioester by the disulfide bond. This attack results in peptide bond formation and reconstitutes the AEP active site thiol [3].

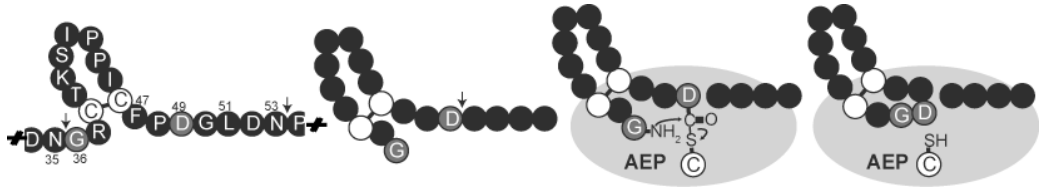

Fig. 2. Proposed model for SFTI-1 biosynthesis (adapted from Mylne et al. 2011) [3]. AEP cleavage occurs firstly at N35 and N53, later at D49 followed by ligation of G36 and D49 (grey) to form the cyclic peptide, SFTI-1.

Solid phase peptide synthesis is the best way to study peptides for applications in drug design. However, to produce them on a large-scale by chemical synthesis becomes costly. We propose to use plants as a cost effective manufacturing process for producing cyclic peptide therapeutics based on molecules produced naturally by plants. To do this, requires an in-depth knowledge of their *in planta* processing so that we may manipulate this processing to vary the peptides produced. To enhance this understanding we will continue to approach this problem by combining synthetic chemistry with plant genetic engineering.

Acknowledgments

We would like to thank the other authors of reference [3], M.L. Colgrave, N.L. Daly, A.H. Chanson, E.J. McCallum and A. Jones, as well as R. Clark for synthesis advice. A.G. Elliott holds an Australian Postgraduate Award Scholarship, D.J. Craik is a National Health and Medical Research Council Professorial Fellow and J.S. Mylne is an Australian Research Council Queen Elizabeth II Fellow and IMB's John S. Mattick Fellow.

References

1. Luckett, S., et al. *J. Mol. Biol*. **290**, 525-533 (1999).
2. Long, Y., et al. *Bioorg. Med. Chem. Lett.* **11**, 2515-2519 (1999); Daly, N., et al. *J. Biol. Chem*. **281**, 23668-23675 (2006); Swedberg, J., et al. *Chem. Biol.* **16**, 633-643 (2009).
3. Mylne, J., et al. *Nat. Chem. Biol.* **7**, 257-259 (2011).
4. White, C., Yudin, A. *Nat. Chem.* **3**, 509-524 (2011).
5. Schnolzer, M., et al. *Int. J. Pept. Protein Res.* **40**, 180-193 (1992).
6. Hara-Hishimura, I., et al. *Plant J.* **4**, 793-800 (1993); Hiraiwa, N., et al. *FEBS Lett.* **447**, 213-216 (1999); Gruis, D., et al. *Plant Cell* **16**, 270-290 (2004); Otegui, M.; et al. *Plant Cell* **18**, 2567-2581 (2006).