# Total synthesis of the macrocyclic cysteine knot microprotein MCoTI-II†

**Panumart Thongyoo, Edward W. Tate and Robin J. Leatherbarrow***

*Received (in Cambridge, UK) 23rd May 2006, Accepted 1st June 2006*
*First published as an Advance Article on the web 20th June 2006*
**DOI: 10.1039/b607324g**

**The first total synthesis of MCoTI-II, a cysteine knot microprotein and potent trypsin inhibitor, is described; a synthetic strategy has been developed that combines efficient backbone construction *via* optimised solid phase peptide synthesis with one-pot 'thia-zip' native chemical ligation and refolding to yield the natural product.**

Cyclotides are a unique class of cysteine-rich microproteins up to 37 residues in length that exhibit a wide range of biological properties, ranging from anti-HIV to insecticidal activity.[1] They are characterised by a cyclic peptide backbone that adopts an intricate cysteine knot topology, whereby two disulfide bridges form a ring through which a third is threaded, resulting in a highly stable secondary structure. The three-dimensional structures of five cyclotides have been determined by NMR spectroscopy,[2] and chemical syntheses of several small-to-medium size cyclotides have been reported.[3]

MCoTI-II is a 34-residue cyclotide first isolated from the seeds of *Momordica cochinchinensis*,[4] and is the founding member of a small cyclotide family that exhibits very potent inhibition of trypsin. Whilst the trypsin inhibitor cyclotides do not share significant sequence homology with other cyclotides beyond the presence of three cysteine bridges, solution NMR spectroscopy has shown that MCoTI-II does adopt a cysteine knot topology (Fig. 1).[2d] We present here the first total synthesis of MCoTI-II.

Our approach to the synthesis of MCoTI-II (Scheme 1) involves three distinct stages: backbone synthesis and generation of a C-terminal thioester; ring-closing macrolactam formation by native chemical ligation (NCL); and final refolding to the native cysteine knot topology. The macrocyclic 34-residue peptide backbone of MCoTI-II was assembled prior to disulfide bond formation,‡ and the Ala22–Cys23 amide bond was selected as the disconnection point for NCL in order to provide minimal steric hindrance to lactamisation. Thus Fmoc-protected alanine was loaded onto a polystyrene resin bearing a sulfonamide (or 'safety-catch') linker[5] and the backbone assembled according to Fmoc/$^{t}$Bu solid phase peptide synthesis (SPPS) protocols.[6] Several optimisations were required for efficient backbone SPPS, including the double-coupling of residues Val7, Lys10 and Ile11. In addition, Asp5–Gly6 was introduced as a dipeptide building block incorporating a 2-hydroxy-4-methoxybenzyl (Hmb) protecting group on the glycine backbone nitrogen to suppress aspartimide formation at this position,[7] which otherwise resulted in the recovery of only trace quantities of 34-mer product. Activation of the sulfonamide linker with iodoacetonitrile[5] followed by cleavage with ethyl 3-mercaptopropionate[8] provided the linear fully-protected backbone bearing a C-terminal thioester. Subsequent acidolysis, and purification by precipitation with *tert*-butyl methyl ether, followed by reverse phase HPLC (RP-HPLC) rendered the free N-terminal cysteine thioester **1** in 15% overall yield, based on the initial resin loading (0.62 mmol $g^{-1}$).

Ring-closing NCL proceeded smoothly in 100 mM phosphate buffer at pH 7.4 in the presence of excess tricarboxyethylphosphine (TCEP) to yield the head-to-tail cyclic backbone **2** with >95% conversion by RP-HPLC. This reaction is thought to proceed *via* a so-called thia-zip (or cysteine zipper) mechanism[3a,9] whereby a dynamic equilibrium of intermediate cyclic thioesters is trapped by an irreversible S-to-N rearrangement at the N-terminal cysteine. The thia-zip approach overcomes the otherwise unfavourable formation of the 102-atom macrolactam, and enables the reaction to be performed in the absence of protecting groups and at relatively high concentration (0.5 mg $mL^{-1}$). To our knowledge, **2** is the largest macrolactam prepared to date by a thia-zip reaction.

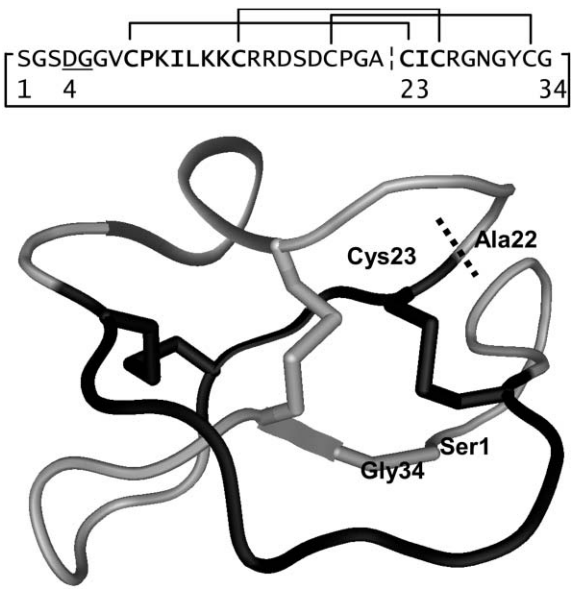
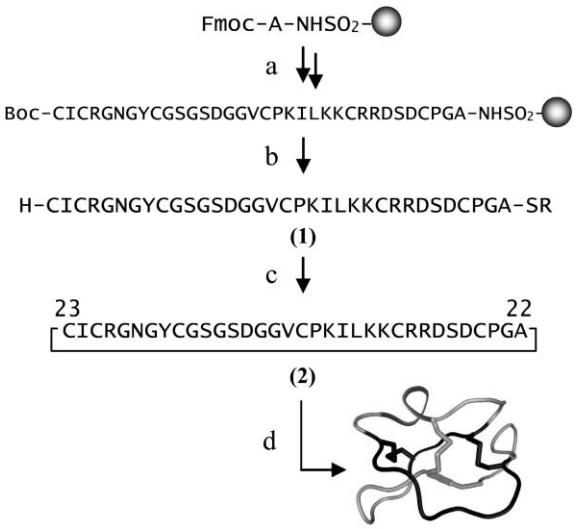



**Fig. 1** Amino acid sequence and three-dimensional solution NMR structure of MCoTI-II,[2d] indicating the backbone and disulfide connectivity and the Ala22–Cys23 backbone disconnection used in our synthetic strategy. The central ring system is highlighted in bold. Presented with PyMol.[14]

*Biological and Biophysical Chemistry Section, Department of Chemistry, Imperial College London, UK SW7 2AZ. E-mail: r.leatherbarrow@imperial.ac.uk; Fax: +44 (0)20 7594 1139; Tel: +44 (0)20 7594 5752*

† Electronic supplementary information (ESI) available: TOCSY spectrum of MCoTI-II. See DOI: 10.1039/b607324g




**Scheme 1** Strategy for the total synthesis of MCoTI-II. (a) Assembly of the fully-protected peptide backbone by Fmoc/$^t$Bu SPPS; (b) (i) $ICH_2CN$–*N*-methylpyrrolidone (NMP)–DIPEA (8:1:1), rt, 24 h, (ii) ethyl 3-mercaptopropionate–DMF 2:5, cat. PhSNa, rt, 24 h; (iii) trifluoroacetic acid (TFA)–water–triisopropylsilane–ethanedithiol 94:2.5:2.5:1, 3 h, rt; (c) 6 equiv. TCEP, 0.1 M phosphate buffer, pH 7.4, rt, 20 min; (d) 1 mM glutathione, 0.1 M carbonate buffer, pH 7.8, rt, 24 h.

It was found that **2** is prone to rapid and spontaneous disulfide bond formation to give a complex mixture of isomers. However, as described in a recent study on the refolding pathway of natural MCoTI-II,[2d] we found that mild air oxidation and refolding in the presence of glutathione allows the microprotein to adopt the native cysteine knot topology. Indeed, we found that ring-closing NCL and refolding could be achieved in a one-pot reaction by dissolving **1** and glutathione in carbonate buffer at pH 7.5, resulting in excellent conversion to MCoTI-II (>95% by RP-HPLC).

Selected characterisation data for synthetic MCoTI-II are presented in Fig. 2. The correspondence of the RP-HPLC retention time, MALDI-TOF, 1D and TOCSY[10] proton NMR spectral data to that of natural MCoTI-II, coupled with potent inhibitory activity against trypsin,[11] demonstrate unequivocally that the synthetic product adopts the native cysteine knot configuration.

In summary, we have achieved the total synthesis of MCoTI-II in a strategy that combines optimised SPPS of the peptide backbone bearing an N-terminal cysteine and C-terminal thioester, native chemical ligation/cyclisation *via* an efficient thia-zip reaction, and refolding under mild conditions to yield the native cysteine knot topology. To our knowledge this represents the first total synthesis of MCoTI-II, which is the largest synthetic cyclotide reported to date. We anticipate that the work reported here will be useful for the development of combinatorial library approaches for optimisation of cyclotide activity,[12] and in studies directed towards the use of MCoTI-II as a scaffold to present structured peptide motifs for chemical genetics.[13]

We are grateful to the Royal Thai Government for a research studentship (to PT). We also acknowledge preliminary work carried out in our laboratory by Dr A. Jaulent and Mr S. Hyung (currently at the University of Cambridge, UK).

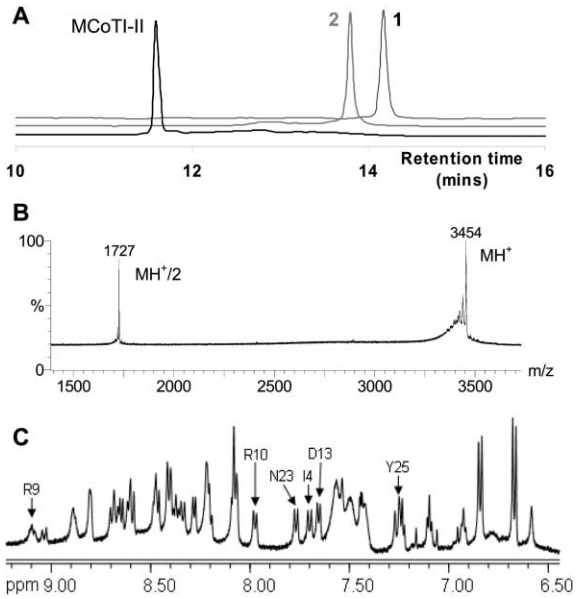



**Fig. 2** Selected characterisation data for synthetic MCoTI-II. **A**: Overlay of RP-HPLC traces (C-18 column, gradient 0 to 85% MeCN in 0.1% aq. TFA over 35 min) showing purified **1**, **2** and synthetic MCoTI-II; **B**: positive ion MALDI-TOF mass spectrum; **C**: 1D proton NMR spectrum (500 MHz, 9:1 $H_2O$–$D_2O$, 0.1% TFA) indicating selected residues in the amide region.[2d]

## Notes and references

‡ We and others have found that efficient and correct folding of the cyclotide cysteine knot is contingent on initial formation of the macrocyclic peptide backbone.

1 (*a*) D. J. Craik, N. L. Daly, T. Bond and C. Waine, *J. Mol. Biol.*, 1999, **294**, 1327–1336; (*b*) K. R. Gustafson, R. C. Sowder, L. E. Henderson, I. C. Parsons, Y. Kashman, J. H. Cardellina, J. B. Mcmahon, R. W. Buckheit, L. K. Pannell and M. R. Boyd, *J. Am. Chem. Soc.*, 1994, **116**, 9337–9338; (*c*) C. Jennings, J. West, C. Waine, D. Craik and M. Anderson, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10614–10619.

2 (*a*) M. E. Felizmenio-Quimio, N. L. Daly and D. J. Craik, *J. Biol. Chem.*, 2001, **276**, 22875–22882; (*b*) A. Heitz, J. F. Hernandez, J. Gagnon, T. T. Hong, T. T. C. Pham, T. M. Nguyen, D. Le-Nguyen and L. Chiche, *Biochemistry*, 2001, **40**, 7973–7983; D. G. Barry, N. L. Daly, H. R. Bokesch, K. R. Gustafson and D. J. Craik, *Structure*, 2004, **12**, 85–94; (*c*) N. L. Daly, A. Koltay, K. R. Gustafson, M. R. Boyd, J. R. Casas-Finet and D. J. Craik, *J. Mol. Biol.*, 1999, **285**, 333–345; (*d*) M. Cemazar, N. L. Daly, S. Haggblad, K. P. Lo, E. Yulyaningsih and D. J. Craik, *J. Biol. Chem.*, 2006, **281**, 8224–8232.

3 (*a*) J. P. Tam, Y. A. Lu and Q. T. Yu, *J. Am. Chem. Soc.*, 1999, **121**, 4316–4324; (*b*) N. L. Daly, S. Love, P. F. Alewood and D. J. Craik, *Biochemistry*, 1999, **38**, 10606–10614.

4 (*a*) L. Chiche, A. Heitz, J. C. Gelly, J. Gracy, P. T. T. Chau, P. T. Ha, J. F. Hernandez and D. Le-Nguyen, *Curr. Protein Pept. Sci.*, 2004, **5**, 341–349; (*b*) J. F. Hernandez, J. Gagnon, L. Chiche, T. M. Nguyen, J. P. Andrieu, A. Heitz, T. T. Hong, T. T. C. Pham and D. L. Nguyen, *Biochemistry*, 2000, **39**, 5722–5730.

5 B. J. Backes and J. A. Ellman, *J. Am. Chem. Soc.*, 1994, **116**, 11171–11172.

6 E. Atherton and R. C. Sheppard, *Solid-phase peptide synthesis, a practical approach*, IRL Press, Oxford, 1989.

7 M. Quibell, D. Owen, L. C. Packman and T. Johnson, *J. Chem. Soc., Chem. Commun.*, 1994, 2343–2344.

8 R. Ingenito, E. Bianchi, D. Fattori and P. Antonello, *J. Am. Chem. Soc.*, 1999, **121**, 11369–11374.

9 J. P. Tam and Y. A. Lu, *Tetrahedron Lett.*, 1997, **38**, 5599–5602.
10 See Fig. 1 in supplementary information†.
11 O. Avrutina, H. U. Schmoldt, D. Gabrijelcic-Geiger, D. Le Nguyen, C. P. Sommerhoff, U. Diederichsen and H. Kolmar, *Biol. Chem.*, 2005, **386**, 1301–1306.
12 D. J. Craik, S. Simonsen and N. L. Daly, *Curr. Opin. Drug Discovery Dev.*, 2002, **5**, 251–260.
13 E. W. Tate, *Signal Transduction*, 2006, **6**, 144–159.
14 W. L. DeLano, *The PyMOL molecular graphics system*, DeLano Scientific, San Carlos, CA, USA, 2002, http://www.pymol.org.