Fmoc-Assisted Synthesis of a 29-Residue Cystine-Knot Trypsin Inhibitor Containing a Guaninyl Amino Acid at the P1-Position

Olga Avrutina,[a] Hans-Ulrich Schmoldt,[b] Harald Kolmar,[b] and Ulf Diederichsen*[a]

Keywords: Cyclotides / Amino acids / Protein mimetics / Solid-phase synthesis / Trypsin inhibitors

An Fmoc-assisted synthesis of a 29 amino acid trypsin inhibitor (McoTI-29) is presented, which contains a non-natural guaninyl nucleo amino acid as arginine mimetic at the P1 position. This artificial amino acid functions as a conformational restricted arginine isoster with reduced basicity. The folded cyclotide McoTI-II open-chain variant McoTI-29 has been synthesized from a linear precursor by in situ oxidation.

Introduction

Over the last decade a whole range of small disulfide-rich proteins with about 30 amino acids in length have been discovered in plants.[1−5] In these microproteins, the linear peptide backbone is cross-linked through disulfide bridges between cysteine residues, which is a common structural motif for knotted peptides.[2] These well-defined and extremely stable three-dimensional structures possess a wide spectrum of biological activities. This makes knotted microproteins attractive for functional and structural investigations, as well as potential agents for medicinal and agricultural applications.[1] As they are useful tools in peptide research, a number of these microproteins have been synthesized, both chemically by Boc-assisted synthesis and biologically.[6−9]

Here we report an Fmoc-assisted solid-phase synthesis followed by selective disulfide bond formation of an open-chain mimetic of McoTI-II. The wild-type microprotein McoTI-II is a trypsin inhibitor from *Momordica cochinchinensis* and consists of a 34 amino acid backbone − overall six cysteine residues form the knotted motif. The overall structure therefore contains six backbone loops (loop 1 − loop 6) as well as disulfide bonds between Cys I−Cys IV and Cys II−Cys V that form a ring, which is penetrated by the third disulfide bond Cys III−Cys VI (Figure 1). Moreover, McoTI-II exists as a head-to-tail macrocycle[4,5] and topologically belongs to the family of knotted microproteins known as cyclotides.[1] On the other hand, the overall fold of McoTI-II is very similar to that of the squash open-chain trypsin inhibitors.[4]

The chemical synthesis of this kind of macrocyclic microproteins is usually provided by solid-phase peptide synthesis (SPPS) based on Boc-protection strategy.[8,9] Nevertheless, yields are usually quite low and the single step monitoring during automated synthesis is hampered due to lack of a suitable chromophore. For an efficient generation of these microproteins, a stepwise control of the synthesis progression is advantageous, especially when non-natural amino acids are incorporated. Moreover, Fmoc-based synthesis eliminates strong acidic cleavage overall providing milder synthesis conditions. We therefore decided to develop a general protocol for the generation of microproteins based on the Fmoc-protection strategy being aware of possible aggregation phenomena with longer peptides. So far, only the synthesis of a 14 residue, single disulfide-containing cyclic trypsin inhibitor has been reported based on the Fmoc-OAll approach.[7]

Results and Discussion

Design of the Target Microprotein McoTI-29 with (Guaninyl)alanine in P1-Position

From the results of McoTI-II solution structure investigations, it was suggested that head-to-tail cyclization does not play a significant role in the formation of the active site conformation of the trypsin inhibitor.[4] The secondary structure is already well defined by disulfide bridges, and an open-chain mimetic generated by deletion of amino acids *S28-G32* provides microprotein McoTI-29 most likely without loss of structural information at the active site (Figure 1). The cystine knot of this microprotein is not influenced by the modification. The open chain analogue

[a] Institut für Organische und Biomolekulare Chemie, Georg-August-Universität Göttingen, Tammannstraße 2, 37077 Göttingen, Germany
Fax: (internat.) + 49-551-392944
E-mail: udieder@gwdg.de
[b] Abteilung für Molekulare Genetik, Institut für Mikrobiologie und Genetik, Griesebachstraße 8, 37077 Göttingen, Germany
Fax: (internat.) + 49-551-393805
E-mail: hkolmar@uni-molgen.gwdg.de

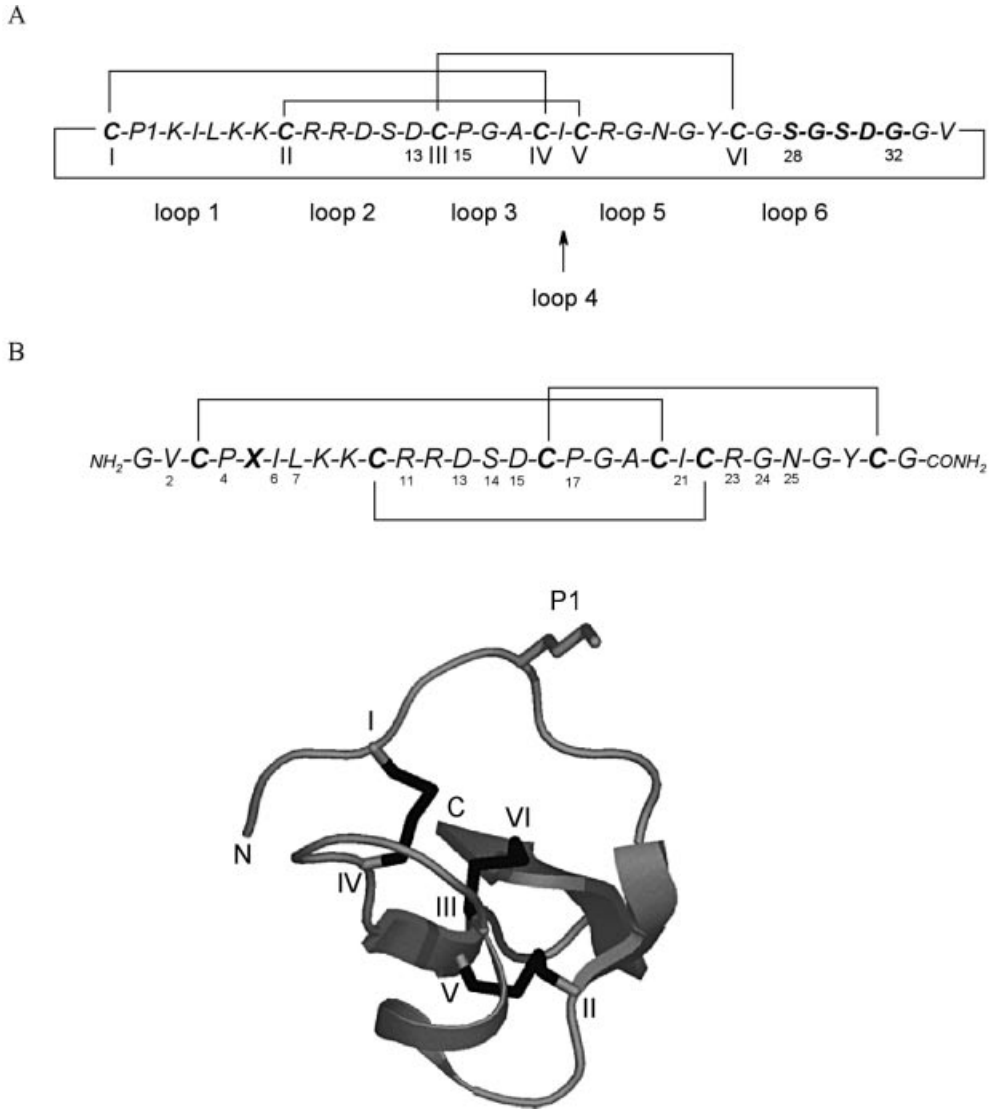

Figure 1. A: Amino acid sequence of circular McoTI-II with disulfide bridges indicated; B: amino acid sequence (***X*** = Lys in the native protein and ***X*** = AlaG in the modified microprotein) and model for the open-chain McoTI-II analogue McoTI-29, deduced from the NMR structure of cyclic McoTI;[4] the P1-position, the N- and C-termini and the disulfide bonds CI−CIV, CII−CV, CIII−CVI are indicated; β-strands are drawn as arrows; the knot-forming cysteines are marked I−VI

McoTI-29 was therefore expected to fold correctly. The intention of our investigation was to establish an automated synthesis protocol for linear microproteins that opens up the possibility of including noncanonical amino acids. Furthermore, it was necessary to prove that synthetic linear microproteins fold in the same manner as the wild-type macrocycles. Finally, (guaninyl)alanine (**2**) (Figure 2) was introduced as an arginine mimetic, which sets the ground for upcoming biological studies. Recently, **2** was introduced as an isosteric, conformationally restricted arginine mimetic with reduced basicity of the guanidino group.[10] Furthermore, additional proton donor and acceptor functionalities are provided with **2**. We needed to evaluate if these structural alterations result in increased bioavailability, higher binding affinity or different receptor selectivity.

Mixed Automated and Manual Solid-Phase Synthesis and Folding

The synthesis of McoTI-29 with AlaG in the P1 position (Figure 1) is challenging with regard to coupling of six cysteines without racemization, the generation of sequences that are known to be difficult to synthesize (*RGNGYC*, *DSDC*, *GSGSDG*),[4] aggregation and solubility problems during chain elongation, incorporation of artificial nucleo amino acids and, finally, the oxidative folding process. These problems were addressed by the interplay between the manual and the automated solid-phase peptide synthesis based on the Fmoc-strategy:

The aggregation phenomena were based on sequence length, significant number of amino acids with a high

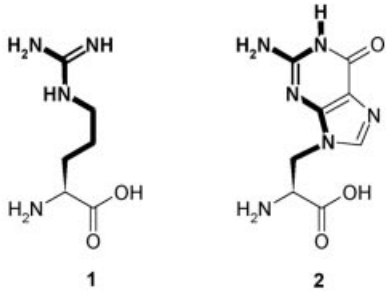

Figure 2. Nucleo amino acid AlaG (**2**) as an isoster for Arg (**1**) in the P1-position of McoTI-29; the isosteric relationship between Arg and AlaG is marked with highlighted bonds

aggregation probability, the ability to form secondary structures, and difficult sequences. Thus, β-sheet interactions between the McoTI-II C-terminal hairpin and residues 13−15 were in fact observed during the secondary structure analysis.[4] Moreover, the synthesis of the microprotein family of the structurally related open-chain trypsin inhibitors is characterized by a low yield of the linear precursor oligomer.[11] For the synthesis of McoTI analogues, aggregation can be expected as the wild-type secondary structure, which consists of several β-strands and turns, has a high folding propensity but may also be prone to intermolecular aggregation. To check the influence of the resin on aggregation processes, several automated test syntheses on a 0.015-mmol scale were conducted. For a polystyrene-type resin (0.74 mmol/g loading capacity), it was found that the polypeptide chain stopped growing after six residues were attached. Fortunately, this problem was possible to overcome by the use of NovaSyn® TGR resin with a PEG spacer between the growing peptide chain and the polystyrene matrix, and by a low loading capacity (0.18−0.23 mmol/g).[12] The secondary structure disrupting surrogate Fmoc-Asp-(O*t*Bu)-Ser($\psi^{Me,Me}$)pro-OH was used in large-scale synthesis in spite of sequential coupling of *D13* and *S14* in order to affect aspartimide formation and to decrease aggregation.[13]

The Fmoc-assisted synthesis of long cysteine-containing peptides is always difficult because of substantial racemization under the conditions of standard protocols used in automated peptide syntheses.[14] The manual procedure was therefore used for all cysteines, preferentially with DIC/HOBt, 5 min preactivation and a coupling time of about 1 h. This also allowed for a better monitoring of the respective coupling step by the Kaiser test. Double coupling was necessary for amino acids *C28*, *C22* and *C3*. After each cysteine coupling, a small portion of the peptide resin was subjected to TFA cleavage in order to ensure the quality of these intermediates by HPLC ESI MS analysis of the resulting crude products.

The guaninyl nucleo amino acid **2** was introduced as the *N*α-Fmoc-protected amino acid. From the incorporation of AlaG in peptide nucleic acids, it was known that further protection on the guanine was not required.[15−17] The relatively low solubility of the Fmoc-protected amino acid and the inherent aggregation potential of the guaninyl residue also led to manual incorporation of the guaninyl nucleo amino acid. Nevertheless, the bioavailability of the microprotein McoTI-29 should not been altered by incorporation of one AlaG unit **2**, as the overall hydrophilicity of the microprotein is quite high.

For all amino acids except cysteine and the non-natural amino acids, chain elongation was performed with the use of an Applied Biosystems peptide synthesizer ABI 433 with conditional monitoring (FastMoc 0.1 MPP CondMon mode). As difficult coupling usually follows difficult deprotection, the coupling time was extended automatically when the difference in intensities between the last deprotection loop and the one before extended to 5%. Activation was performed by HBTU/HOBt/DIEA with a standard coupling time of 30 min. For amino acids *L7*, *D15*, *G24*, *I21* and *V2*, coupling was conditionally continued for another 50 min, followed by acetic anhydride capping. Conditional double coupling followed by acetic anhydride capping was used for amino acids *N25*, *G24*, *R23*, *P17*, *R11* and *P4*. Amino acids *S14* and *D13* were introduced in one standard coupling step as Mutter's pseudo-proline dipeptide.[13] A large-scale synthesis provided a higher yield and an improved purity of the resulting peptide than the small-scale synthesis conducted without the use of Mutter's dipeptide.

Finally, the ESI MS analysis of the linear target peptide McoTI-29 in a crude reaction mixture showed that the synthesis of the desired microprotein was quite successful. Peaks for peptides with incomplete coupling of *C28* and *Y27*, *Pro*, and incomplete deprotection of one side-chain *t*Bu group are detected near the main peak that is assigned to linear McoTI-29 (Figure 3). Attempts to extend the TFA treatment for complete deprotection led to decomposition of the peptide.

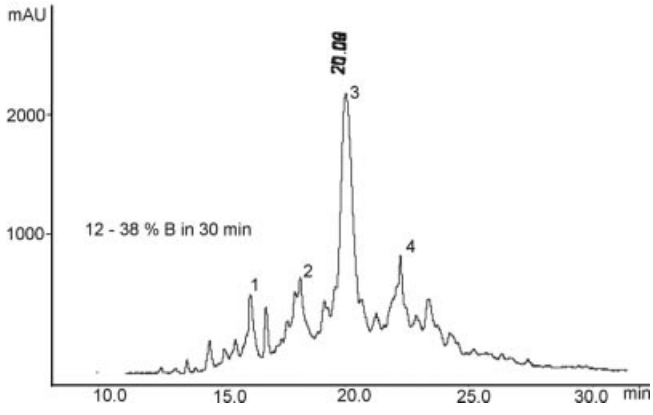

Figure 3. HPLC profile of crude linear peptide McoTI-29; peak 3: linear target peptide; as products from incomplete coupling were detected: peak 1: missing *Pro*; peak 2: missing *TyrCys*; peak 4: additional *t*Bu group; for HPLC conditions see Exp. Sect.

After purification of the linear McoTI-29 precursor by RP HPLC, gentle folding to the oxidized microprotein was observed under NH_4HCO_3 buffer conditions. The knotted microprotein McoTI-29 with AlaG in the P1-position was obtained as the major component (Figure 4) despite the

various intra- and intermolecular possibilities for disulfide formation with six cysteine residues. Finally, it was possible to convert 7 mg of the linear precursor to give 2 mg of the desired microprotein after HPLC purification. The experimentally determined biological trypsin inhibitor activity of McoTI-29 with AlaG in the P1-position provided evidence for the correct folding. Interestingly, in comparison with the "wild type" McoTI-29 by exchange of AlaG for Lys as P1 residue, the trypsin binding ability was retained, albeit with reduced affinity.[18]

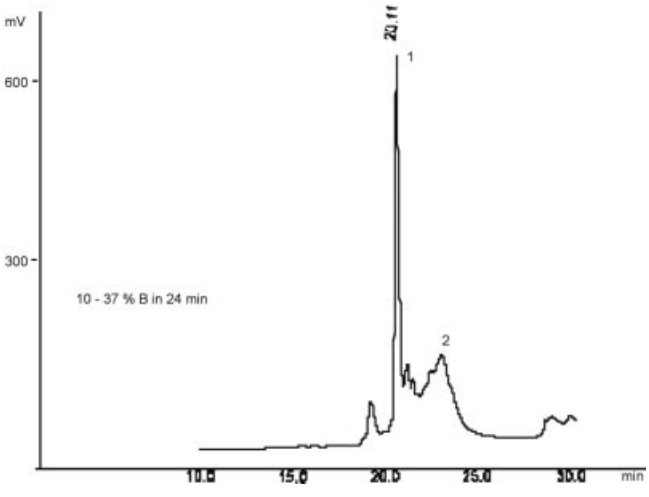

Figure 4. Result of folding procedure: Peak 1 indicates the folded peptide McoTI-29, peak 2 a mixture of incomplete or not properly folded intermediates; for HPLC conditions see Exp. Sect.

Conclusion

The SPPS of an analogue of the open-chain trypsin inhibitor McoTI-II by using Fmoc-strategy with subsequent cysteine oxidation in situ is described. A (guaninyl)alanine amino acid was introduced as a conformational restricted and less basic arginine isoster at the P1-position. The synthetic approach combines advantages of automated and manual Fmoc-synthesis, and easily allows the preparation of these well-folded microproteins in a multimilligram scale. The solubility in water of the synthesized peptide McoTI-29 with a guaninyl nucleo amino acid in the P1-position was high, both in linear and oxidized form. Finally, the inhibitory activity against trypsin is quite remarkable with a K_i in the nanomolar range and high selectivity for trypsin.[18]

Experimental Section

General Remarks: All chemicals used were of the highest grade available. Solvents were of analytical grade and used as supplied. Fmoc-protected amino acids were used with the following side-chain protecting groups: *t*Bu (Asp, Tyr), Boc (Lys), Trt (Cys, Asn), Pbf (Arg). Pseudo-proline dipeptide Fmoc-Asp(O*t*Bu)-Ser($\psi^{Me,Me}$)pro-OH was purchased from Calbiochem-Novabiochem GmbH. ESI mass spectra were measured with a Finnegan LCQ spectrometer. High-resolution ESI mass spectra were recorded with a Bruker APEX-Q III 7T. HPLC was performed with a Pharmacia Äcta basic system using YMC J'sphere ODS H-80, RP C-18 (250 × 20 mm, 4 μm, 8 nm) and Phenomenex Synergi 4u Hydro-RP 80A (250 × 10 mm, 4 μm, 8 nm) columns for preparative separations and YMC J'sphere ODS H-80, RP C-18 (250 × 4.6 mm, 4 μm, 8 nm) and Phenomenex Synergi 4u Hydro-RP 80A (250 × 4.6 mm, 4 μm, 8 nm) for analytical samples.

Synthesis of Fmoc-AlaG-OH: The preparation of the Fmoc-protected nucleo amino acid Fmoc-AlaG-OH was performed as reported earlier.[10]

Synthesis of the Linear Precursor McoTI-29 with AlaG at P1-Position: The linear peptide was assembled using a combination of automated and manual Fmoc-assisted SPPS on PEG-type amide NovaSyn® TGR resin (Calbiochem-NovaBiochem GmbH) with a loading capacity of 0.23 mmol/g. The C-terminal glycine (5 equiv. with respect to resin loading) was activated by HATU/DIEA (5 equiv. and 10 equiv., respectively) and attached without preactivation onto the resin over six hours. The loading with the first amino acid was determined by UV spectroscopy by monitoring Fmoc-deprotection.[19] Chain elongation was performed using the Applied Biosystems peptide synthesizer ABI 433. Starting from (435 mg, 0.10 mmol) the amide resin, FastMoc 0.1 MPP CondMon mode with conditional monitoring[20] was used for all amino acid cycles including pseudo-proline dipeptide, with the exception of cysteine and Fmoc−AlaG−OH. After preswelling in DCM and then NMP, each amino acid cycle was programmed to include basically 16 min Fmoc-deprotection with conditional conductivity monitoring, 8 min activation with HBTU/HOBt/DIEA, and 30 min coupling. Depending on the Fmoc-monitoring, an additional 20 min of Fmoc-deprotection, and 50 min coupling and acetic anhydride capping were automatically included in the regular cycles. Cysteine was coupled manually in DCM/NMP with the standard DIC/HOBt protocol.[21] The peptide resin was transferred from the peptide synthesizer reaction vessel into an SPPS reactor and treated by nitrogen bubbling in the presence of activated cysteine until a negative Kaiser test was obtained. The resin was then thoroughly washed with NMP and DCM, and transferred back to the peptide synthesizer reaction vessel. After the peptide was assembled up to *I6* and the N-terminal Fmoc-group was removed, the peptide resin was washed thoroughly with DCM, methanol and dried until constant weight was reached. A portion of the resin (150 mg, 0.016 mmol according to a calculating program of ABI 433 peptide synthesizer) was subjected to coupling of Fmoc−AlaG−OH, which was coupled manually within 2 h with the use of the HBTU/HOBt/DIEA activation protocol. A solution of Fmoc−AlaG−OH, HBTU and HOBt (0.064 mmol each, 4 equiv.) in 3 mL of NMP was preactivated for 10 min, than added to the peptide resin, together with Hünig base (0.16 mmol, 10 equiv.). After 2 h of coupling, the peptide resin was washed three times with NMP, and the Fmoc group was cleaved with piperidine (20% in NMP). Two deprotection steps were conducted for 10 and 15 min followed by NMP and DCM washes; the peptide resin was then transferred into the reaction vessel, and the usual chain elongation cycles were repeated. After completion of the synthesis, the peptide resin was transferred into a polypropylene syringe, washed with DCM, methanol, and dried. Cleavage of the peptide from the resin was achieved using TFA/DTT/TIS/anisole (23:1:0.5:0.5 v:w:v:v, 5 mL per 200 mg peptide resin). The resin was gently shaken for 2.5 h, filtered, and washed with TFA (3 × 1 mL). The

combined filtrates were concentrated in vacuo and the peptide was precipitated with MTBE. The peptide was centrifuged, washed with MTBE, dried, dissolved in water containing 0.1% TFA, and lyophilized. The crude peptide (18 mg) was purified by preparative RP-HPLC and after lyophilization 10 mg (56%) of the linear McoTI-29 with AlaG modification was obtained. The linear peptide was characterized by ESI mass spectrometry as well as by high resolution ESI FT-ICR MS. Linear McoTI-29 with AlaG in the P1-position: HPLC (RP C-18, gradient: 5−60% 0.1% TFA in water (eluent A) and 0.1% TFA in 90% acetonitrile (eluent B) in 30 min, flow rate of 10 mL/min, monitoring at 215 and 280 nm) 10.82 min. ESI-MS (CH_3CN/H_2O/TFA): *m*/*z* = 1055.6 (100) $[M + 3H]^{3+}$, 792.2 (26) $[M + 4H]^{4+}$, 1583.3 (18) $[M + 2H]^{2+}$. High resolution ESI-MS of $[C_{126}H_{212}N_{47}O_{37}S_6]^{5+}$; calcd. 791.60941; found 791.60887, ΔM: 0.68 ppm; $[C_{126}H_{210}N_{47}O_{37}S_6]^{3+}$; theory: 1055.14345; found 1055.14422, ΔM: 0.73 ppm.

Folding Procedure: Oxidation of the linear McoTI-29 to the cystine-knot was performed by dissolving the reduced lyophilized peptide in 50 μL 10 mM HCl per mg of peptide, followed by addition of NH_4HCO_3 (200 mM, pH 9.1) to a final concentration of 1−1.5 mg/mL.[22] The reaction mixture was incubated overnight in a PET container under vigorous shaking at room temperature. Purification of the folded peptide was done by RP-HPLC with Phenomenex C_{18} columns (analytical: 250 × 4.6 mm; preparative: 250 × 10 mm) using eluent A: 0.1% TFA in H_2O, eluent B: 50% acetonitrile, 50% 2-propanole containing 0.1% (v/v) TFA. A linear gradient of 10−37% B was performed with flow rates of 1 mL/min for analytical purposes and 3.5 mL/min for preparative separations. The fractions containing the oxidized peptide (monitoring at 217 nm) were combined and lyophilized to yield 29% of the desired knotted protein. Successful oxidation was confirmed by ESI mass spectrometry and by measurement of the inhibitory activity against trypsin, which is strictly dependent on correct disulfide bond formation.[22] Folded McoTI-29 with AlaG at the P1-position: ESI-MS (MeOH/CH_3COOH): *m*/*z* = 790.8 (100) $[M + 4 H]^{4+}$, 1053.8 (26) $[M + 3 H]^{3+}$.

Acknowledgments

Support of the Deutsche Forschungsgemeinschaft (SFB 416) is gratefully acknowledged.

[1] M. Trabi, D. Craik, *Trends Biochem. Sci.* **2002**, *27*, 132−138.
[2] D. J. Craik, N. L. Daly, T. Bond, C. Waine, *J. Mol. Biol.* **1999**, *294*, 1327−1336.
[3] C. Jennings, J. West, C. Waine, D. Craik, M. Anderson, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10614−10619.
[4] M. E. Felizmenio-Quimio, N. L. Daly, D. J. Craik, *J. Biol. Chem.* **2001**, *276*, 22875−22882.
[5] J.-F. Hernandes, J. Gagnon, L. Chiche, T. M. Nguen, J.-P. Andrieu, A. Heits, T. T. Hong, T. T. C. Pham, D. Le Nguyen, *Biochemistry* **2000**, *39*, 5722−5730.
[6] N. L. Daly, D. J. Craik, *J. Biol. Chem.* **2000**, *275*, 19068−19075.
[7] M. L. J. Korsinczky, H. J. Schirra, K. J. Rosengren, J. West, B. A. Condie, L. Otvos, M. Anderson, D. J. Craik, *J. Med. Biol.* **2001**, *311*, 579−591.
[8] J. P. Tam, Y.-A. Lu, *Tetrahedron Lett.* **1997**, *38*, 5599−5602.
[9] J. P. Tam, Y.-A. Lu, Q. Yu, *J. Am. Chem. Soc.* **1999**, *121*, 4316−4324.
[10] K. B. Lorenz, U. Diederichsen, *Lett. Pept. Sci.* **2003**, *10*, 111−117.
[11] D. Le-Nguyen, D. Nalis, B. Castro, *Int. J. Peptide Protein Res.* **1989**, *34*, 492−497.
[12] M. Quibell, T. Johnson, in *Fmoc solid phase synthesis. A practical approach* (Eds.: W. C. Chan, P. D. White), Oxford University Press, New York, **2000**, pp.115−135.
[13] T. Wöhr, F. Wahl, A. Nefzi, B. Rohwedder, T. Sato, X. Sun, M. Mutter, *J. Am. Chem. Soc.* **1996**, *118*, 9218−9227.
[14] ABI 433A peptide synthesizer. User's manual, V.1, PE Applied Biosystems, Sec. 4-8-4-21.
[15] U. Diederichsen, *Angew. Chem.* **1997**, *109*, 1966−1969; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 1886−1889.
[16] U. Diederichsen, D. Weicherding, *Synlett* **1999**, 917−920.
[17] M. F. H. Hoffmann, A. M. Brückner, T. Hupp, B. Engels, U. Diederichsen, *Helv. Chim. Acta* **2000**, *83*, 2580−2593.
[18] H.-U. Schmoldt, O. Avrutina, U. Diederichsen, H. Kolmar, manuscript in preparation.
[19] W. C. Chan, P. D. White, in *Fmoc solid phase synthesis. A practical approach* (Eds.: W. C. Chan, P. D. White), Oxford University Press, New York, **2000**, p. 63.
[20] ABI 433A peptide synthesizer. User's manual, V.1, PE Applied Biosystems, Sec. 5-24.
[21] F. Albericio, I. Annis, M. Royo, G. Barany, in *Fmoc solid phase synthesis. A practical approach* (Eds.: W. C. Chan, P. D. White), Oxford University Press, New York, **2000**, pp. 81−91.
[22] A. Wentzel, A. Christmann, R. Krätzner, H. Kolmar, *J. Biol. Chem.* **1999**, *274*, 21037−21074.

Received June 22, 2004

