Seven Novel Macrocyclic Polypeptides from *Viola arvensis*

Ulf Göransson, Teus Luijendijk, Senia Johansson, Lars Bohlin, and Per Claeson*

Division of Pharmacognosy, Department of Pharmacy, Biomedical Center, Uppsala University, Box 579, S-751 23 Uppsala, Sweden

Received September 11, 1998

Seven novel macrocyclic polypeptides, designated as varv peptides B−H, have been isolated from the aerial parts of *Viola arvensis*. Their primary structures have been elucidated by automated Edman degradation and mass spectrometry. They all consist of 29 or 30 amino acid residues, covalently cyclized via the amide backbone and by three internal disulfide bridges. Their amino acid sequences are as follows: varv peptide B, cyclo-(TCFGGTCNTPGCSCDPWPMCSRNGLPVCGE); varv peptide C, cyclo-(TCVGGTCNTPGCSCSWPVCTRNGVPICGE); varv peptide D, cyclo-(TCVGGSCNTPGCSCSWPVCTRNGLPICGE); varv peptide E, cyclo-(TCVGGTCNTPGCSCSWPVCTRNGLPICGE); varv peptide F, cyclo-(TCTLGTCYTAGCSCSWPVCTRNGVPICGE); varv peptide G, cyclo-(TCFGGTCNTPGCSCDPWPVCSRNGVPVCGE); and varv peptide H, cyclo-(TCFGGTCNTPGCSCETWPVCSRNGLPVCGE). The varv peptides B−H exhibited high degrees of homology with the hitherto known macrocyclic peptides varv peptide A, kalata B1, violapeptide I, circulins A and B, and cyclopsychotride A.

Within a research project aimed at finding novel bioactive polypeptides from plants, we have previously developed and described a fractionation protocol that provides a highly purified polypeptide fraction.[1] By utilizing this protocol we also reported the isolation of a macrocyclic polypeptide, varv peptide A, from *Viola arvensis* Murray (Violaceae). The primary structure of varv peptide A, consisting of 29 amino acid residues covalently cyclized via the amide backbone and by three disulfide bridges, is shared with a small group of plant polypeptides. This group is confined to the uterotonic peptide kalata B1 from the tropical plant *Oldenlandia affinis* DC. (Rubiaceae),[2−6] the hemolytic violapeptide I isolated from *V. arvensis*,[7] the HIV-inhibitory circulins A and B from *Chassalia parvifolia* K. Schum. (Rubiaceae),[8,9] and the neurotensin antagonist cyclopsychotride A from *Psychotria longipes* Muell. Arg. (Rubiaceae).[10] In this report we describe the primary structures of seven novel macrocyclic polypeptides, varv peptides B−H, that have been isolated from *V. arvensis* by a combination of adsorption chromatography on Sephadex LH-20 and reversed-phase HPLC.

Results and Discussion

By following an established five-step fractionation protocol,[1] a highly enriched polypeptide fraction (previously designated Fraction P) was isolated. The protocol consisted, in short, of pre-extraction with dichloromethane to remove lipophilic substances, main extraction with 50% aqueous ethanol, removal of tannins by filtration through polyamide, removal of low-molecular-weight substances by size-exclusion chromatography on Sephadex G-10 and, finally, solid-phase extraction on RP_{18} material to remove polysaccharides and salts. Analysis of the Fraction P by reversed-phase HPLC (diode array detection) indicated the presence of several tryptophan-containing peptides.

The main peptide in Fraction P, varv peptide A, was originally isolated in a rather low yield by the extensive use of reversed-phase HPLC.[1] To increase the isolation yield of the peptides in this study, column chromatography on Sephadex LH-20 with 50% MeOH−0.1% TFA (v/v) in water was attempted. Contrary to true size-exclusion chromatography, in which an exclusion limit of approximately 4 kDa could be expected,[11] the peptides eluted in the low-molecular-weight range of the column. Increasing the polarity of the eluent by adding water further increased the retention times of the peptides, indicating that the separation took place in a reversed-phase-like manner. In the mobile phase aqueous 30% MeOH−0.05% TFA all peptides eluted after the geometrical column volume and a separation into groups was clearly obtained (Figure 1). Reversed-phase adsorption is a known additive mechanism to size exclusion in chromatography on Sephadex LH-20, which has been proposed to be due to the interaction between the dextran gel matrix and aromatic solutes.[12] This has been utilized extensively for the separation of low-molecular-weight substances.[11,13] Our results with the varv peptides indicate a similar behavior for polypeptides (Figure 1). Peptides containing one aromatic amino acid in addition to the conserved tryptophan (*cf.* below) were more strongly retarded. Isolation of varv peptide A and seven novel peptides, varv peptides B−H, was finally achieved by gradient reversed-phase HPLC on C_{18} material.

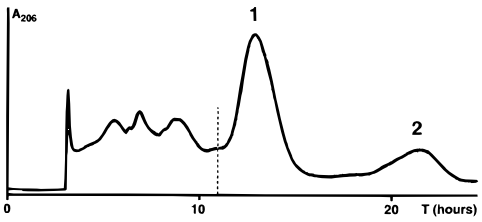

Figure 1. Chromatogram of the polypeptide fraction from *V. arvensis* on Sephadex LH-20 in aqueous 30% MeOH−0.1% TFA. The macrocyclic peptides were isolated from peak 1 (varv peptides A−E) and peak 2 (varv peptides F−H). Both peaks appeared after the geometrical volume of the column (660 mL, marked with a dotted line), which indicates that separation depends on an adsorption mechanism (see also Figure 2). Peptides containing one additional aromatic amino acid residue besides the conserved tryptophan eluted in peak 2. Varv peptide B, however, eluted in peak 1.

By including adsorption chromatography on Sephadex LH-20 in the purification procedure, the final yield of varv peptide A was improved fivefold (1750 μg from 100 mg of Fraction P, compared to the previously described 150 μg

* To whom correspondence should be addressed. Tel.: +46 18 471 44 79. Fax: +46 18 50 91 01. E-mail: per.claeson@pharmacog.uu.se.

10.1021/np9803878 CCC: $18.00
Published on Web 01/26/1999

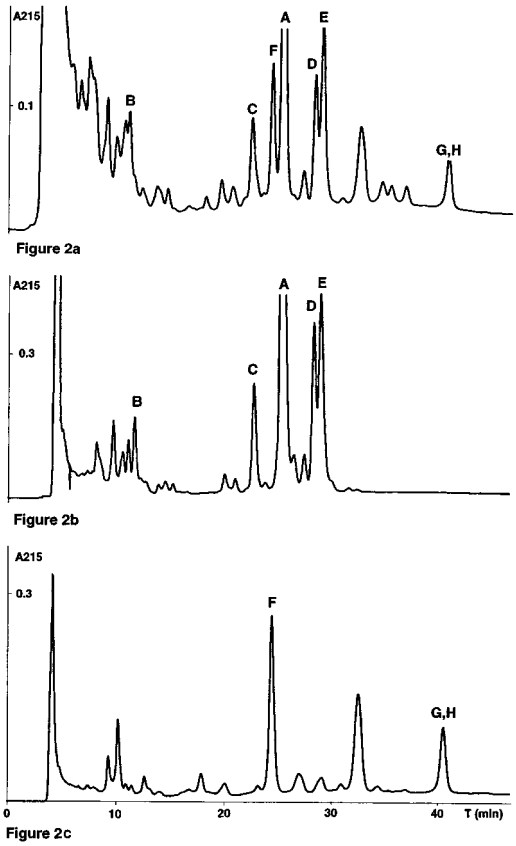

Figure 2. (a−c) Analytical reversed-phase HPLC of the polypeptide fraction before (Figure 2a) and after separation on Sephadex LH-20 (Figures 2b and 2c). The chromatogram of peak 1 in Figure 1 is shown in Figure 2b, and the chromatogram of peak 2 is shown in Figure 2c. Individual varv peptides are marked by their letters. The difference in selectivity of reversed-phase adsorption chromatography on Sephadex LH-20 and silica based C_{18} material is demonstrated clearly. Peaks overlapping each other (varv peptides A and F) in preparative HPLC on C_{18} material were separated on Sephadex LH-20. The main part of the injection peak seen in Figure 2a corresponds to the substances eluting before the end of the geometrical volume seen in Figure 1. Final purification of the peptides was achieved by rechromatography on reversed-phase HPLC (see Experimental). The peak at approximately 33 min in Figure 2c was then split up to at least five components and was not further analyzed. The analytical reversed-phase HPLC was performed on a Rainin Dynamax C18 column (250 × 4.6 mm, 5 μm, 300 Å) operated with a linear gradient from 37 to 43% organic modifier (MeCN/iPrOH, 6/4) in 0.1% TFA.

from 42 mg[1]). Moreover, isolation of individual peptides was facilitated not only by the removal of nonpeptidic substances, but also by the separation of the peptide mixture into two major groups (Figures 1 and 2a−c).

The amino acid compositions of the peptides were determined by quantitative amino acid analysis and MALDI-TOF MS (Table 1). Masses calculated for linear peptides with all cysteines in the reduced form were consistently approximately 24 Da higher than the molecular weights found. This is in agreement with the proposed macrocyclic structures and can be attributed to the deficit of one water molecule (−18 Da) and six hydrogens (−6 Da), due to the absence of N- and C-terminals and the formation of three disulfide bridges.

Prior to automated sequencing, all peptides were reduced and alkylated with 4-vinylpyridine to give the *S*-(*β*-4-pyridylethyl)cysteine (PEC) derivatives. Peptides containing only one glutamic acid could directly be digested with endoproteinase Glu-C, to provide single linear products amenable to sequencing in one run. Varv peptide H, however, was found to contain two glutamic acid residues and was first digested with trypsin. A single linear PEC

```
Varv peptide H  TCFGGTCNTPGCSCETWPVCSRNGLPVCGE

Trypsin:        TCFGGTCNTPGCSXET      NGLPVCGE
GluC:            TCFGGTCNTPGCSCE
GluC:                              TWPVCSRNGLPVCGE
```

Figure 3. Determination of the amino acid sequence of varv peptide H. Overlapping fragments from digestion with trypsin (starting with an asparigine, N) and endoproteinase Glu-C give the whole sequence and an additional proof of the macrocyclic structure. One cysteine residue could not be identified in the trypsin fragment due to low yield (marked X in the figure).

derivative was isolated and the molecular weight determined by MALDI-TOF MS as 3712.8 Da, which is in agreement with the theoretical mass (3710.0 Da) for this derivative [3055.4 (native cyclic peptide) + 18.0 (H_2O) + 6.0 (6H) + 630.6 (6 × 4-vinylpyridine)]. From this digest, the sequence of the first 24 amino acids was determined. The alkylated varv peptide H was then digested with Glu-C, to provide two fragments, which were sequenced. The trypsin fragment overlaps with the two Glu-C fragments, and thus they unambiguously define the complete cyclic amino acid sequence of varv peptide H (Figure 3).

In a comparison of the sequences of the varv peptides and the hitherto known family of macrocyclic polypeptides from plants (Figure 4), two groups of peptides were easily identified (Figure 5). The first of these are the peptides isolated from *V. arvensis*, which share a very high degree of sequence homology with kalata B1. The net charge for these peptides is neutral or slightly acidic (varv peptides B, G, H have a net charge of −1). The second group consists of circulins A and B, together with cyclopsychotride A. They are all basic with a net charge of +2. Comparison of the primary structures of the most abundant peptide from *V. arvensis*, varv peptide A, and the most distant basic peptide, cyclopsychotride A, shows a homology of more than 40%. This indicates that the two groups of peptides, nonetheless, are closely related.

The consensus sequence shows that the cysteines are conserved in all the macrocyclic peptides (Figure 4). In the more variable regions, substitution seems to occur with amino acids of similar polarity. This maintains the pattern of alternating hydrophobic and hydrophilic regions in all the peptides. One exception to this is the strongly basic regions of circulins A and B and cyclopsychotride A (Figure 4).

At present, nothing is known about the function or significance of the varv peptides. The similarity with the cocktail of toxic peptides produced by the cone snail, *Conus geographus*, is, however, striking.[14] The cone snail produces a series of closely related peptides that block different types and subtypes of calcium channels of its prey. By analogy, it appears reasonable to assume that these macrocyclic polypeptides provide *V. arvensis* with a combinatorial peptide library, enabling optimal effect on a hypothetical biological target of importance to the plant.

Experimental Section

Plant Material. The aerial parts of *Viola arvensis* Murray (Violaceae) were collected in July 1996, near the Ångström Laboratory, Uppsala, Sweden. A voucher specimen (labeled VM-107) was identified by Dr. Ö. Nilsson, the Botanical Garden, Uppsala University, Uppsala, and deposited at the herbarium of Uppsala University. The plant material used in this study was identical with the one used previously.[1]

Isolation Procedure. The polypeptide fraction (Fraction P) was isolated from the plant material as previously described.[1] Fraction P (100 mg, corresponding to 7.3 g dried and powdered plant material[1]) was subjected to chromatography on Sephadex LH 20 with aqueous 30% MeOH containing 0.05%

Table 1. The Amino Acid Composition and Molecular Weights of the Varv Peptides. For Each Peptide, the Residues from Amino Acid Analysis Are Listed to the Left and the Residues from Sequencing to the Right

amino acid	varv B		varv C		varv D		varv E		varv F		varv G		varv H	
Asp/Asn (D/N)	3.1	1D, 2N	2.1	2 N	2.0	2 N	2.1	2 N	1.0	1 N	3.1	1D, 2N	2.0	2 N
Thr (T)	2.9	3	4.0	4	3.0	3	3.8	4	4.7	5	2.9	3	3.8	4
Ser (S)	2.1	2	2.1	2	2.7	3	2.0	2	2.1	2	2.0	2	2.0	2
Glu/Gln (E/Q)	1.1	1 E	1.0	1 E	1.0	1 E	1.0	1 E	1.0	1 E	1.0	1 E	1.9	2 E
Pro (P)	3.7	4	3.1	3	2.8	3	2.7	3	1.9	2	3.7	4	3.0	3
Gly (G)	4.8	5	5.0	5	5.0	5	5.1	5	4.0	4	4.9	5	4.8	5
Ala (A)									1.0	1				
Cys (C)	5.4[a]	6[b]	4.8[a]	6[b]	4.9[a]	6[b]	4.9[a]	6[b]	4.9[a]	6[b]	4.9[a]	6[b]	5.0[a]	6[b]
Val (V)	1.0	1	3.0	3	2.1	2	2.1	2	2.0	2	3.0	3	2.0	2
Met (M)	1.0	1												
Ile (I)			1.0	1	1.0	1	0.9	1	1.0	1				
Leu (L)	0.9	1			1.0	1	1.1	1	1.0	1			1.0	1
Tyr (Y)									0.9	1				
Phe (F)	1.0	1									1.1	1	1.0	1
Arg (R)	1.0	1	1.0	1	1.0	1	1.0	1	1.0	1	1.0	1	1.0	1
Trp (W)	1[d]	1	1[d]	1	1[c]	1	1[c]	1	1[c]	1	1[d]	1	1[d]	1
No aa:s	30		29		29		29		29		30		30	
Mol wt														
Calculated[e]	3085.5[g]		2878.3		2878.3		2892.3		2959.4		3023.4		3055.4	
MS[f]	3087.0[g]		2878.0		2879.5		2894.9		2956.2		3023.5		3053.2	

[a] Half-cystine was determined as cysteic acid with a separate sample following oxidation with performic acid. [b] Cysteine was determined as (pyridylethyl)cysteine following alkylation with 4-vinylpyridine. [c,d] Tryptophan was determined photometrically either through separate measurements by spectrophotometer,[c] or by utilizing the HPLC on-line spectra obtained from the diode-array detection. The presence of tryptophan was then estimated from the relative intensity of UV absorbance at 280 vs 250 nm. [d,e] Calculated with average masses and with the total sum from amino acid composition adjusted to the macrocyclic structure (−18 Da) and three disulfide bridges (−6 Da). [f] Masses determined by MALDI-TOF MS with an experimental accuracy of ±0.1% (see Experimental Section). [g] Including oxidation of the methionine residue.

```
varv peptide F       cyclo-( TCTLGTC---YTAGCSCS-WPVCTRNGVPICGE )
varv peptide C       cyclo-( TCVGGTC---NTPGCSCS-WPVCTRNGVPICGE )
varv peptide D       cyclo-( TCVGGSC---NTPGCSCS-WPVCTRNGLPICGE )
varv peptide E       cyclo-( TCVGGTC---NTPGCSCS-WPVCTRNGLPICGE )
varv peptide A       cyclo-( TCVGGTC---NTPGCSCS-WPVCTRNGLPVCGE )
varv peptide B       cyclo-( TCFGGTC---NTPGCSCDPWPMCSRNGLPVCGE )
varv peptide G       cyclo-( TCFGGTC---NTPGCSCDPWPVCSRNGVPVCGE )
varv peptide H       cyclo-( TCFGGTC---NTPGCSCETWPVCSRNGLPVCGE )

Consensus[a]                 TC  G C    T GCSC  WP C RNG P CGE
Identity[b]                  **  *:*--- *.****. **:*:***:*:***

kalata B1            cyclo-( TCVGGTC---NTPGCTCS-WPVCTRNGLPVCGE )
violapeptide I       cyclo-( TCVGGTC---NTPGCSCS-RPVCTXNGLPVCGE )

circulin A           cyclo-( SCVWIPC-ISAALGCSCK-NKVCYRNG-IPCGE )
circulin B           cyclo-( SCVFIPC-ISTLLGCSCK-NKVCYRNGVIPCGE )
cyclopsychotride A   cyclo-( SCVFIPCTVTALLGCSCK-SKVCYKNS-IPCGE )

Consensus[a]                  C    C      GC C     C  N    CGE
Identity[b]                  :*   .*      **:*.   :*  *.:  ***
```

Figure 4. Sequences of reported macrocyclic polypeptides in the range of 29−31 amino acid residues, tabulated starting with the cleavage site of Glu-C and vertically ordered to illustrate homologies. Inserted gaps for maximum alignment are marked with (−). The consensus sequence[a] and the identity based on matching codons[b] show a very high degree of homology within the group of varv peptides. This sequence is conserved with only minor changes also in kalata B1 and violapeptide I. Including all sequences in the comparison limits the conserved amino acids mainly to the regions of the cysteine residues. The pattern of hydrophobic/hydrophilic regions of the sequences is maintained, with exception for the strongly basic regions of circulins A and B and cyclopsychotride A. [The alignment was made with the CLUSTAL W (1.7) program.[15]]

TFA as mobile phase. Three columns [550 × 30 (i. d.), 380 × 25 and 480 × 15 mm], packed with 170 g of gel in total and connected in series, were used. The columns were attached to an LKB 2150 HPLC pump operated at a flow rate of 1 mL/min. The effluent was monitored at 206 nm using an LKB 2158 Uvicord SD. The fractions corresponding to peak 1 and 2 in Figure 1 were collected, concentrated, and lyophilized, yielding 17.7 and 6.3 mg, respectively. Preparative HPLC was performed using a Shimadzu LC10 system, equipped with a Shimadzu SPD-M10Avp diode array detector. Portions of up to 4 mg of the fractions were repeatedly injected on a 250 × 10 (i. d.) mm Rainin Dynamax column (C_{18}, 5 μm, 300 Å), eluted by a gradient of 37−43% organic modifier (MeCN/iPrOH, 6/4) in 0.1% TFA. Analysis and final purification of individual peptides were performed using either a 250 × 4.6 (i. d.) mm Rainin Dynamax column (C_{18}, 5 μm, 300 Å) or a 150 × 4.6 (i. d.) mm Vydac (polymer-RP, 259VHP5415, 5 μm, 300 Å) column, eluted with concentrations of MeCN between 30 and 40% in aqueous 0.1% TFA. The final yield of the peptides was: varv peptide A, 1750 μg (0.0240% w/w dried plant material); B, 30 μg (0.0004%); C, 200 μg (0.0027%); D, 300 μg (0.0041%); E, 315 μg (0.0043%); F, 260 μg (0.0036%); G, 135 μg (0.0018%); H, 150 μg (0.0021%), as determined by quantitative amino acid analysis.

Structure Determination. Quantitative amino acid analysis was performed at the Amino Acid Analysis Centre, Department of Biochemistry, Uppsala University. The peptides were hydrolyzed for 24 h at 110 °C with 6 N HCl containing 2 mg/mL phenol, and the hydrolysates were analyzed with an LKB model 4151 Alpha Plus amino acid analyzer using ninhydrin

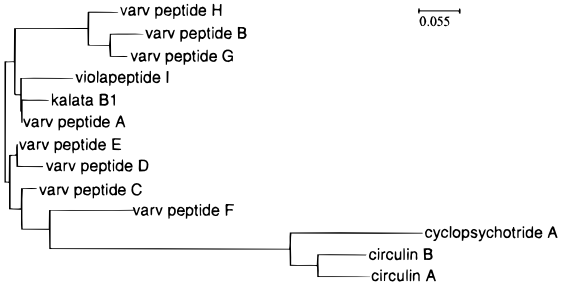

Figure 5. A phylogenetic tree of macrocyclic polypeptides from plants. The relationships based on the alignment of amino acids were created with CLUSTAL W (1.7)[15] and are graphically presented by the program NJPLOT.[16]

detection. Molecular weights were determined with a Kratos Kompact IV MALDI-TOF mass spectrometer. The spectrometer was externally calibrated and operated in the linear mode (experimental accuracy: ±0.1%). Average isotopic masses were used for all calculated molecular weights. For the sequence analysis, the peptides were reduced with either mercaptoethanol or dithioerythritol in 0.25 M Tris-HCl containing 1 mM EDTA and 6 M guanidine-HCl (pH 8.5, 24 °C, 2 h). The reduced peptides were subsequently alkylated to their PEC derivatives by adding 4-vinylpyridine to the solution (37 °C, 1 h). Desalting and isolation of the alkylated peptides were performed with gel filtration on a Superdex Peptide HR 10/30 column (Amersham Pharmacia Biotech, Uppsala, Sweden) in 40% AcN, 0.1% TFA (v/v) or by reversed-phase HPLC [Rainin Dynamax, C_{18}, 5 μm, 300 Å, 250 × 4.6 (i. d.) mm]. Enzymatic digests were generated by endoproteinase Glu-C (Calbiochem−Novabiochem Co., La Jolla, CA) and, for varv peptide H, trypsin (Sigma Chemical Co., St. Louis, MO) in 0.1 M ammonium bicarbonate (pH 8.1, 37 °C, 4 h). Reversed-phase HPLC [Vydac polymer-RP, 259VHP5415, 5 μm, 300 Å, 150 × 4.6 (i. d.) mm or Brownlee Aquapore, OD300, 7 μm, 300 Å, 30 × 2.1 (i. d.) mm] and H_2O−MeCN in 0.1 TFA gradient elution were used for isolation of digested peptides. The amino acid sequences were determined by automated Edman degradation, using a Protein/Peptide sequencer model 477 A, coupled online to a model 120A analyzer (Applied Biosystems).

Acknowledgment. We thank Dr. Å. Engström, Department of Medical and Physiological Chemistry, Uppsala University, for assistance with MS and amino acid sequence analyses. We also thank Amersham Pharmacia Biotech, Uppsala, for support with separation equipment.

References and Notes

(1) Claeson, P.; Göransson, U.; Johansson, S.; Luijendijk, T.; Bohlin, L. *J. Nat. Prod.* **1998**, *61*, 77−81.
(2) Gran, L. *Lloydia* **1973**, *36*, 174−178.
(3) Gran, L. *Lloydia* **1973**, *36*, 207−208.
(4) Gran, L. *Acta Pharmacol. Toxicol.* **1973**, *33*, 400−408.
(5) Sletten, K.; Gran, L. *Medd. Nor. Farm. Selsk.* **1973**, *35*, 69−82.
(6) Saether, O.; Craik, D. J.; Campbell, I. D.; Sletten, K.; Juul, J.; Norman, D. G. *Biochemistry* **1995**, *34*, 4147−4158.
(7) Schöpke, T.; Hasan Agha, M. I.; Kraft, R.; Otto, A.; Hiller, K. *Sci. Pharm.* **1993**, *61*, 145−153.
(8) Gustafson, K. R.; Sowder, R. C., II.; Henderson, L. E.; Parsons, I. C.; Kashman, Y.; Cardellina, J. H., II.; McMahon, J. B.; Buckheit, R. W., Jr; Pannell, L. K.; Boyd, M. R. *J. Am. Chem. Soc.* **1994**, *116*, 9337−9338.
(9) Derua, R.; Gustafson, K. R.; Pannell, L. K. *Biochem. Biophys. Res. Commun.* **1996**, *228*, 632−638.
(10) Witherup, K. M.; Bogusky, M. J.; Anderson, P. S.; Ramjit, H.; Ransom, R. W.; Wood, T.; Sardana, M. *J. Nat. Prod.* **1994**, *57*, 1619−1625.
(11) Henke, H. *Preparative Gel Chromatography on Sephadex LH-20*; Hüthig GmbH: Heidelberg, 1995.
(12) Janson, J. C. *J. Chromatogr.* **1967**, *28*, 12−20.
(13) Wilk, M.; Rochlitz, J.; Bende, H. *J. Chromatogr.* **1966**, *24*, 414−416.
(14) Olivera, B. M.; Hillyard, D. R.; Marsh, M.; Yoshikami, D. *TiBtech* **1995**, *13*, 422−426.
(15) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.
(16) Perrière, G.; Gouy, M. *Biochimie* **1996**, *78*, 364−369.

NP9803878