Circulins A and B: Novel HIV-Inhibitory Macrocyclic Peptides from the Tropical Tree *Chassalia parvifolia*[1]

Kirk R. Gustafson, Raymond C. Sowder II,[†] Louis E. Henderson,[†] Ian C. Parsons, Yoel Kashman,[‡] John H. Cardellina II, James B. McMahon, Robert W. Buckheit, Jr.,[§] Lewis K. Pannell,[∥] and Michael R. Boyd*

Laboratory of Drug Discovery Research and Development
Developmental Therapeutics Program
National Cancer Institute
Frederick Cancer Research and Development Center
Frederick, Maryland 21702-1201

Received May 23, 1994

A number of structurally diverse peptide derivatives have been reported to inhibit the replication and cytopathic effects of the human immunodeficiency virus (HIV). These include linear peptides,[2] cyclized metabolites,[3] and peptide mimetic enzymes inhibitors.[4] Through our efforts to discover novel HIV-inhibitory natural products,[5] antiviral activity was detected in crude extracts of the tropical tree *Chassalia parvifolia* (Rubiaceae).[6] Subsequent anti-HIV bioassay-guided isolation and structure determination efforts have led to the discovery of circulins A and B, two novel macrocyclic peptides which now represent the largest naturally occurring peptides in which the entire primary amino acid chain is covalently cyclized via peptide bonds. The circulins exhibit some sequence homology with the human B-cell antigen CD22, and they may share key structural features with the defensin class of linear peptides.

Solvent–solvent partitioning of the extract concentrated the anti-HIV activity in a butanol-soluble fraction. This material was sequentially separated by gel permeation on Sephadex LH-20 (MeOH), centrifugal partition chromatography (Sanki, H_2O/*n*-BuOH/HOAc/EtOH, 10:8:1:1, descending mode), and C_{18} HPLC to give the two major active constituents.

The mobility of circulin A by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and its staining by Coomassie brilliant blue (a single band of M_r approximately 3 kDa) suggested that it was an amino acid derived metabolite. Acid hydrolysis and analysis of the resulting amino acids indicated that circulin A was composed of 30 amino acids (Table 1). Attempts to sequence the peptide directly via Edman degradation were unsuccessful, indicating that its N-terminus was blocked. Digestion of circulin A with endoproteinase Lys-C gave one 28 amino acid fragment which was amenable to Edman sequencing (Figure 1) and a dipeptide that analyzed for N and K by amino acid analysis. Fast-atom bombardment mass spectrometry (FABMS) of circulin A provided a MH^+ molecular ion at m/z 3153.3.[7] While the composition of the Lys-C products was fully consistent with the amino acid analysis of intact circulin A, the calculated average molecular weight (3176 Da) of a linear peptide containing these 30 amino acids was approximately 24 Da greater than the molecular weight of circulin A as determined by FABMS.

Table 1. Amino Acid Analysis of Circulins A and B

	residues from amino acid analysis		residues from sequencing	
amino acid	circulin A	circulin B	circulin A	circulin B
D and N	2.0	2.0	2 (2N)	2 (2N)
T	0.1	1.1	0	1
S	3.3	2.9	3	3
Q and E	1.2	1.0	1 (E)	1 (E)
P	2.0	2.0	2	2
G	3.2	2.9	3	3
A	2.0	0	2	0
V	2.0	2.8	2	3
M	0	0	0	0
I	2.9	2.7	3	3
L	1.1	2.0	1	2
Y	1.0	1.0	1	1
F	0	1.0	0	1
H	0	0	0	0
K	2.1	2.0	2	2
R	1.1	1.0	1	1
W	1.1[a]	*b*	1	0
C	5.4[c]	6.0[c]	6[c]	6[c]
			30	31 total

[a] Not determined directly. The presence of W was determined from the relative intensity of UV absorbances at 294 vs 206 nm. [b] UV absorbance at 294 vs 206 nm was insufficient to account for the presence of W. [c] C was analyzed as (pyridylethyl)cysteine following treatment of the peptide with vinylpyridine.

(A) Circulin A

Lys-C VCYRNGIPCGESCVWIPCISAALGCSCK + NK

Glu-C SCVWIPCISAALGCSCKNKVCYRNGIPCGE

(B) Circulin B

Arg-C NGVIPCGESCVFIPCISTLLGCSCKNKVCYR

Figure 1. (A) Amino acid sequences of peptides generated by digestion of circulin A (19 h, 24 °C, pH 8.5) with endoproteinases Lys-C or Glu-C. (B) Amino acid sequence of the peptide resulting from treatment of circulin B with Arg-C (20 h, 37 °C, pH 7.6). Following digestion, resultant peptides were characterized by amino acid analysis and sequenced by N-terminal Edman degradation.

A second enzymatic digest of circulin A with endoproteinase Glu-C[8] (Figure 2) provided a linear 30 amino acid peptide that could be sequenced directly (Figure 1). The amino acid composition of this peptide was identical to that determined for circulin A, and it showed sequence homology with the Lys-C products. These enzymatically generated peptides which share common sequences could only arise if the entire amide backbone of circulin A was cyclized via peptide bonds to give the 30 amino acid macrocycle shown in Figure 3. A smaller cyclic structure which blocked the N-terminus could not give the peptides we observed in the enzyme digests.

Reaction of reduced circulin A (EtSH, 37 °C, 1 h) with 4-vinylpyridine (24 °C, 2 h, pH 8.5, Tris-HCl) generated the *S*-(β-4-pyridylethyl)cysteine (PEC) derivative of circulin A. A FABMS MH^+ molecular ion at m/z 3790.5 supported the presence of six disulfide-linked cysteines in the nonreduced starting material (calculated net 636.4 Da increase in molecular weight due to six pyridylethyl groups). Treatment of the PEC derivative of circulin A with Glu-C gave a single linear product with a molecular ion at MH^+ m/z 3807.9, which resulted from addition

* Corresponding author.
† AIDS Vaccine Program, Program Resources, Inc./DynCorp, NCI-FCRDC, Frederick, MD.
‡ On leave from the Department of Chemistry, Tel-Aviv University, Tel-Aviv, Israel.
§ Southern Research Institute, Frederick Research Center, Frederick, MD.
∥ Laboratory of Analytical Chemistry, NIDDK, NIH, Bethesda, Maryland.

(1) HIV Inhibitory Natural Products. 17. Part 16: Fuller, R. W.; Bokesch, H. R.; Gustafson, K. R.; McKee, T. C.; Cardellina, J. H., II; McMahon, J. B.; Cragg, G. M.; Soejarto, D. D.; Boyd, M. R. *BioMed. Chem. Lett.* **1994**, in press.

(2) Nakashima, H.; Masuda, M.; Murakami, T.; Koyanagi, Y.; Matsumoto, A.; Fujii, N.; Yamamoto, N. *Antimicrob. Agents Chemother.* **1992**, *36*, 1249–1255.

(3) (a) Helynck, G.; Dubertret, C.; Mayaux, J. F.; Leboul, J. *J. Antibiot.* **1993**, *46*, 1756–1757. (b) Fréchet, D.; Guitton, J. D.; Herman, F.; Faucher, D.; Helynck, G.; Monegier du Sorbier, B.; Ridoux, J. P.; James-Surcouf, E.; Vuilhorgne, M. *Biochemistry* **1994**, *33*, 42–50.

(4) Darke, P. L.; Huff, J. R. *Adv. Pharmacol.* **1994**, *25*, 399–454.

(5) Boyd, M. R. *AIDS Etiology, Diagnosis, Treatment and Prevention*; Lippincott: Philadelphia, PA, 1988; pp 305–319.

(6) Samples of *C. parvifolia* Schumann (Rubiaceae) were collected in the Iringa Region of Tanzania, Dec 1988, by J. Lovett.

(7) Masses quoted are average molecular weights.

(8) Also commonly referred to as endoproteinase V8.

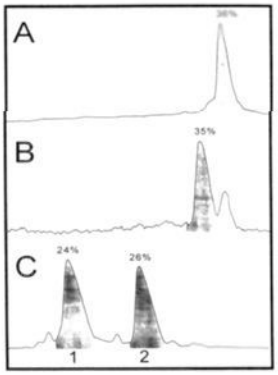

Figure 2. CH_3CN/H_2O gradient elution C_{18} HPLC purification, monitored at 206 nm, of (A) circulin A, (B) Lys-C digest of circulin A, and (C) Glu-C digest of the PEC derivative of circulin A. CH_3CN elution percentages are shown above the peaks; fractions collected under the shaded regions were pooled for sequencing and amino acid analysis. The peak that eluted with 35% CH_3CN in B was the 28 amino acid linear peptide; the later eluting peak was not analyzed. Peak 1 in C was the 30 amino acid linear peptide; peak 2 was undigested starting material. The linear elution gradients were 0–60% CH_3CN over 60 min, 0–40% CH_3CN over 120 min, and 10–35% CH_3CN over 120 min for A, B, and C, respectively.

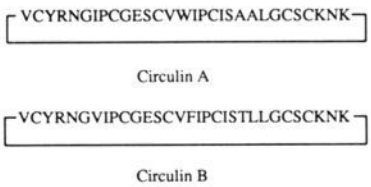

Figure 3. Cyclic amino acid sequences of circulins A and B.

of H_2O when the macrocyclic ring was opened. The molecular weight of native circulin A determined by MS analysis (3152 Da) was in excellent agreement with the calculated mass of the corresponding linear 30 amino acid peptide (3176 Da) minus 18 due to cyclization and minus 6 due to formation of three disulfide bridges. The deduced cyclic structure of circulin A is consistent with the molecular weight determined by MS analysis, the peptide products of endoproteinase digestion, and the intractability of the intact molecule to direct sequencing.

Circulin B was clearly related to circulin A on the basis of its SDS–PAGE characteristics and molecular weight (FABMS, MH^+ m/z 3284.7). The native peptide was resistant to N-terminal Edman sequencing, and amino acid analysis indicated that it consisted of 31 residues (Table 1). The calculated molecular weight of a linear peptide composed of these 31 amino acids (3308 Da) was approximately 24 Da higher than the molecular weight determined by MS. This suggested that circulin B was also macrocyclic with three intramolecular disulfide bonds. The PEC derivative of circulin B (FABMS, MH^+ m/z 3922.3) was digested with endoproteinase Arg-C (Figure 2) to give a 31 amino acid, linear product that was suitable for direct N-terminal sequencing (Figure 1). Amino acid analyses of the cyclic starting material and linear digestion product were virtually identical. This allowed definition of the macrocyclic structure of circulin B as shown in Figure 3. Circulins A and B share a very high degree of sequence homology, differing only by three amino acid substitutions and insertion of one additional valine residue in circulin B.

An *in vitro* XTT-based anti-HIV assay[9] was used to screen for the effects of circulins A and B on virus-induced cell killing in diverse HIV-infected cultures. Ten different strains of HIV and two different host cell lines were employed. Confirmatory assays[9] examined the effects of the compounds upon viral replicative indices, including p24 antigen, supernatant reverse transcriptase and syncytium-forming units. For both compounds, the antiviral cytoprotective concentrations (EC_{50}'s) ranged from about 40 to 260 nM, depending upon the particular virus strain and host cell line used in the assay.[10] The cytotoxicity index concentrations (IC_{50}'s) were quite constant at about 500 nM in all assays. The concentration-dependent increase in cytoprotection observed in the XTT assays was paralleled with decreases in supernatant levels of viral p24, reverse transcriptase and infectious virions. The mechanism of anti-HIV activity for circulins A and B has not yet been defined; however, in a reverse transcriptase (RT) assay[9] they had no activity at concentrations as high as 660 nM. The highly constrained, three-dimensional structures of circulins A and B appear to be critical to their anti-HIV properties, since cytoprotective activity was lost if their disulfide bonds were reduced prior to analysis.

Many proteins and peptides have been described in the literature as cyclic, due to intramolecular disulfide bridges[3,11] or cyclization with one or more side chain functional groups.[3,12] However, with *30 and 31 amino acids, respectively, circulins A and B represent* the largest known naturally occurring macrocyclic peptides which result from peptide bond cyclization of the entire amino acid backbone.[13] While the cyclic nature and amino acid sequences of circulins A and B are novel, a search of protein and DNA data bases[14] revealed, interestingly, that they have significant sequence homology with a region of the human B-cell antigen CD22.[15] Eight out of 11 amino acids were identical within the homologous regions of circulin A and CD22. In the corresponding segment of circulin B, six of 11 amino acids were identical.[15]

Circulins A and B also show some similarities to a widely distributed class of antimicrobial, antiviral, and cytotoxic peptides commonly referred to as defensins.[16] These peptides typically contain a conserved cysteine motif with three intramolecular disulfide bonds, are composed of 29–35 amino acids, and carry a net positive charge. They are active against enveloped viruses,[17] and at least one defensin has been reported to inhibit HIV.[18] The six disulfide-bonded cysteines, molecular size range, net charge (2+), and antiviral activities of circulins A and B suggest that these cyclic peptides share some structural properties with the defensin class of peptides.

(9) Gulakowski, R. J.; McMahon, J. B.; Staley, P. G.; Moran, R. A.; Boyd, M. R. *J. Virol. Methods* **1991**, *33*, 87–100.

(10) For example, with the viral isolates HIV-1_{RF}, HIV-2_{ROD}, and HIV-1 (strain A17), the EC_{50}'s were 40, 266, and 260 nM, respectively.

(11) For example, see: (a) Selsted, M. E.; Harwig, S. S. L. *J. Biol. Chem.* **1989**, *264*, 4003–4007. (b) Head, S.; Ramotar, K.; Lingwood, C. *Infect. Immun.* **1990**, *58*, 1532–1537. (c) Posthumus, W. P. A.; Lenstra, J. A.; van Nieuwstadt, A. P.; Schaaper, W. M. M.; van der Zeust, B. A. M.; Meloen, R. H. *Virology* **1991**, *182*, 371–375. (d) Léonetti, M.; Pillet, L.; Maillère, B.; Lamthanh, H.; Frachon, P.; Couderc, J.; Ménez, A. *J. Immunol.* **1990**, *145*, 4214–4221.

(12) For example, see: (a) McDowell, R. S.; Gadek, T. R. *J. Am. Chem. Soc.* **1992**, *114*, 9253–9265. (b) Heavner, G. A.; Audhya, T.; Doyle, D.; Tjoeng, F.; Goldstein, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 198–209. (c) Su, C. M.; Jensen, L. R.; Heimer, E. P.; Felix, A. M.; Pan, Y. C. E.; Mowles, T. F. *Horm. Metab. Res.* **1991**, *23*, 15–21.

(13) During the course of this work, the preliminary report of a 31 amino acid cyclic peptide from *Psychotria longipes* (Rubiaceae) was made in a poster presentation: Sardana, M.; Bogusky, M.; Anderson, P.; Ransom, R.; Wood, T.; Witherup, K. The Seventh Annual Symposium of the Protein Society, July 24–28, 1993, San Diego, CA.

(14) Linear representations of the amino acid sequences of circulins A and B were compared to the Allprot Data Base.

(15) The amino acid sequence of the CD22 protein was deduced from the nucleotide sequence of the cloned gene. Seed, B.; Stamenkovic, I. *Nature* **1990**, *345*, 74–77. The 11 amino acid regions of sequence homology of circulin A, circulin B, and CD22 are **CVWIPCISAAL**, **CVFIPCISTLL**, AND **CVWIPCTYRAL**, respectively.

(16) (a) Lehrer, R. I.; Lichtenstein, A. K.; Ganz, T. *Annu. Rev. Immunol.* **1993**, *11*, 105–128. (b) Ganz, T.; Selsted, M. E.; Lehrer, R. I. *Bacterial-Host Cell Interaction*; Alan R. Liss, Inc.: New York, 1988; pp 3–14. (c) Selsted, M. E.; Brown, D. M.; DeLange, R. J.; Lehrer, R. I. *J. Biol. Chem.* **1983**, *258*, 14485–14489.

(17) Daher, K. A.; Selsted, M. E.; Lehrer, R. I. *J. Virol.* **1986**, *60*, 1068–1074.

(18) Nakashima, H.; Yamamoto, N.; Masuda, M.; Fujii, N. *AIDS* **1993**, *7*, 1129.