Structure, Vol. 12, 85–94, January, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.str.2003.11.019

Solution Structure of the Cyclotide Palicourein: Implications for the Development of a Pharmaceutical Framework

Daniel G. Barry,[1] Norelle L. Daly,[1] Heidi R. Bokesch,[2] Kirk R. Gustafson,[3] and David J. Craik[1,*]
[1]Institute for Molecular Bioscience
Queensland Bioscience Precinct
The University of Queensland
Brisbane, Queensland 4072
Australia
[2]Basic Research Program
SAIC-Frederick, Incorporated
Frederick, Maryland 21702
[3]Molecular Targets Development Program
Center for Cancer Research
National Cancer Institute
Frederick, Maryland 21702

Summary

The cyclotides are a family of disulfide-rich proteins from plants. They have the characteristic structural features of a circular protein backbone and a knotted arrangement of disulfide bonds. Structural and biochemical studies of the cyclotides suggest that their unique physiological stability can be loaned to bioactive peptide fragments for pharmaceutical and agricultural development. In particular, the cyclotides incorporate a number of solvent-exposed loops that are potentially suitable for epitope grafting applications. Here, we determine the structure of the largest known cyclotide, palicourein, which has an atypical size and composition within one of the surface-exposed loops. The structural data show that an increase in size of a palicourein loop does not perturb the core fold, to which the thermodynamic and chemical stability has been attributed. The cyclotide core fold, thus, can in principle be used as a framework for the development of useful pharmaceutical and agricultural bioactivities.

Introduction

A paradigm shift in the way proteins are viewed is upon us with the discovery of increasing numbers of circular proteins in the past decade. Before these discoveries, proteins had been generally thought of as linear chains of amino acids, which fold into three-dimensional shapes that define their biological functions. Circular proteins, with no identifiable termini, incorporate a continuous cycle of peptide bonds in their backbone. Backbone cyclization of a protein is thought to confer the advantage of reduced sensitivity to proteolytic cleavage (Trabi and Craik, 2002; Rosengren et al., 2003) and has been shown to improve the thermodynamic stability of some natively linear proteins (Iwai et al., 2001; Deechongkit and Kelly, 2002; Siebold and Erni, 2002) and rescue the fold of a destabilized mutant (Camarero et al., 2001). Circular proteins have been identified from diverse sources, including sunflower (Korsinczky et al., 2001), Rhesus macaque monkey (Tang et al., 1999), and bacteria (Samyn et al., 1994; Blond et al., 2001), with sizes ranging from 14 to 78 amino acids. The cyclotides (Craik et al., 1999, 2002a; Craik, 2001) are the largest and most diverse naturally occurring family of backbone cyclized proteins.

The cyclotides have been isolated from a variety of plants (Gustafson et al., 1994; Witherup et al., 1994; Saether et al., 1995; Claeson et al., 1998; Craik et al., 1999; Göransson et al., 1999; Gustafson et al., 2000) and are distinguishable from other proteins by the presence of a circular backbone that is braced by a knotted arrangement of three disulfide bonds. These structural restrictions, observed in the cyclotide solution structures published to date (Saether et al., 1995; Craik et al., 1999; Daly et al., 1999; Rosengren et al., 2003), are thought to contribute to their physiological, thermal, and chemical stability (Gran, 1973). These characteristics support the applicability of the cyclotides as potential drug candidates or scaffolds onto which new bioactivities may be engineered (Craik et al., 2002b). The prototypical cyclotide, kalata B1 (Saether et al., 1995), exhibits insecticidal activity against the caterpillar pest *Helicoverpa punctigera* (Wallengren), occurring either through cellular toxicity or an antifeedent effect (Jennings et al., 2001). The insecticidal and other activities attributed to the cyclotides (Craik, 2001; Craik et al., 2002a) suggest that the physiological advantages conferred by their unique structural features make them a promising scaffold for pharmaceutical and agricultural development (Pearce, 2001).

The surface loops of proteins and bioactive peptides have often been implicated in recognition by protein binding partners. Accordingly, it is of interest to investigate these loops as potential leads for drug discovery. Most small peptides are highly flexible and do not, therefore, adopt unique solution conformations; in particular, peptide fragments generally do not adopt the structure that the same sequence adopts in a native protein. The lack of fixed structure reduces the affinity the peptide might have for a target (for entropy reasons) and obscures any connection between its solution properties and the active, bound-state conformation. Because of this, many strategies have been developed to introduce structural constraints into peptides, including disulfide bonds and other cross-links, D-amino acids, and backbone modifications. The cyclotides have evolved some of these strategies to form the highly stable cyclic cystine knot (CCK) motif. The CCK motif presents the residues of its six loops (i.e., backbone segments between successive Cys residues) to the solvent in a defined orientation, implicating a possible role as a scaffold for bioactive moieties. Studies of acyclic permutants of the cyclotide kalata B1 have confirmed the important role of the circular backbone in stabilizing the protein scaffold (Daly and Craik, 2000).

An anti-HIV bioactivity-guided fractionation of an or-

*Correspondence: d.craik@imb.uq.edu.au

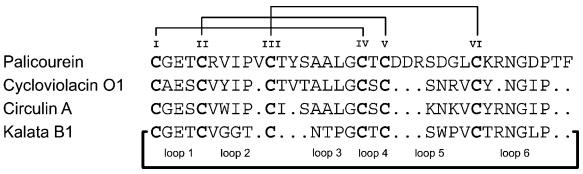

Figure 1. Amino Acid Sequence of Palicourein and Three Other Representative Cyclotide Sequences

The six conserved cysteine residues are numbered (Roman numerals), and the intercysteine loops are labeled 1–6. Cycloviolacin O1 and circulin A are from the *bracelet* subfamily of cyclotides (Craik et al., 1999), while kalata B1 is from the *Möbius* subfamily. A *cis*-Pro bond in loop 5 characterizes the latter.

ganic extract of *Palicourea condensata* (Standley), a member of the Rubiaceae plant family, identified a 37 amino acid polypeptide (Bokesch et al., 2001). This peptide, given the name palicourein, has a head-to-tail cyclic backbone that is covalently cross-linked with three disulfide bonds. A sequence alignment of palicourein with selected cyclotides, as illustrated in Figure 1, suggests that palicourein may be the largest member of this macrocyclic plant peptide family. Of particular interest is the effect that an increase in loop size has on the cystine knot framework of the cyclotides. Palicourein, with an increased loop 5 compared with the other cyclotides shown in Figure 1, allows for such a study. It was therefore of interest to determine the three-dimensional solution structure of palicourein, first to confirm its classification as a cyclotide and second to study the effect of increased loop size on the physiologically and thermodynamically stable cyclic cystine knot framework that is unique to the cyclotides. Here we report the three-dimensional solution structure of palicourein, the largest cyclotide, and discuss the implications for using the cyclotide scaffold as a framework for the stabilization and presentation of peptide epitopes.

Results

^{1}H-NMR Resonance Assignments

The peaks in the two-dimensional total correlation spectroscopy (TOCSY) spectrum of palicourein are well dispersed in the amide region, and the individual amino acid spin systems were readily identified, as illustrated in Figure 2A. These were assigned to specific amino acid residues in the sequence using two-dimensional NOE spectroscopy (NOESY) experiments (Wüthrich, 1986).

The fingerprint region of the NOESY spectrum in Figure 2B shows a complete cycle of αH_i-$NH_{(i+1)}$ sequential connectivities for the entire sequence, unbroken except at the two proline residues. However, in both cases NOEs were seen between α protons of the preceding residue and the proline δ protons, completing the sequential connectivity. The NMR data thus provide unequivocal evidence for the cyclic nature of the peptide, confirming this finding from mass spectral and enzymatic cleavage studies reported earlier (Bokesch et al., 2001). The presence of strong $\alpha_{H(i-1)}$-δ_{Hi} NOE signals and the absence of $\alpha_{H(i-1)}$-α_{Hi} NOE signals in the NOESY spectrum for both proline residues confirms that their amide bonds are in the *trans* conformation. It is not uncommon for *cis*-proline residues to exist in highly constrained peptides; however, this geometry is not present in palicourein.

Secondary Structure of Palicourein

Trends in the observed chemical shifts provided a useful first insight into the secondary structure of palicourein. Chemical shift indices (CSI) (Wishart et al., 1992) for the backbone α protons were calculated using the measured chemical shifts and random coil values (Merutka et al., 1995) and are illustrated in Figure 3A. Most of the residues have chemical shifts that differ from the random coil values by more than 0.1 ppm and hence have CSI values of $\pm$1, suggesting that palicourein is a structured peptide. Residues 3–5, 19–21, and 28–30 have consecutive positive CSI values, which are indicative of extended local structures in these regions. The remainder of the molecule displays an irregular pattern of positive and negative CSI values, which suggests the presence of turns. This is further seen from the pattern of sequential NOEs illustrated in Figure 3A. Isolated $d_{NN(i+2)}$ and $d_{\alpha N(i,i+2)}$ NOEs indicate the presence of turns (Wüthrich, 1986) in several regions of the molecule. The continuous stretch of $d_{NN(i+2)}$ and $d_{\alpha N(i,i+2)}$ NOEs and a series of $^3J_{\alpha HNH}$ coupling constants $<$6 Hz suggest that a short helical segment forms around residues 13–16 (Wüthrich, 1986). The lack of $d_{(i,i+3)}$ or $d_{(i,i+4)}$ NOEs and a lack of a continuous stretch of negative CSI values in this region suggest that it has a 3_{10}, rather than an α-helical, conformation. Figure 3B shows the interresidual NOEs observed for palicourein and confirms that there is a uniform distribution of sequential, medium, and long range NOEs throughout the sequence, consistent with a well defined structure.

Also indicated in Figure 3 are the slowly exchanging amide protons detected after dissolution of the peptide sample in ^2H_2O. Eight of the 37 amide protons are present more than 1 hr after dissolution of the sample in ^2H_2O, and five of these are still present at least 20 hr later, providing a strong indication of protection from the solvent and hence the likelihood of well-defined secondary structure for the molecule.

From a consideration of the experimental NOE, slow exchange, and coupling patterns illustrated in Figure 3, together with some longer range NOEs between backbone protons, an indication of the major elements of secondary structure was derived and is schematically illustrated in Figure 4. The major secondary structural elements are the β strands (residues 19–21 and 28–30) linked by a connecting β turn. A third β strand incorporates residues 3–5.

Electrospray ionization mass spectrometry (ESIMS) of native and reduced palicourein predicted that the six cysteines form three disulfide bonds (Bokesch et al., 2001). The connectivity of the disulfide bonds, however, could only be inferred from the sequence alignment with similar proteins of the cyclotide family, illustrated in Figure 1. Based on the sequence similarity, the six cysteine residues in palicourein were predicted to form three disulfide bonds with a I-IV, II-V, and III-VI pattern of connectivity in preference to the other 14 possibilities. A

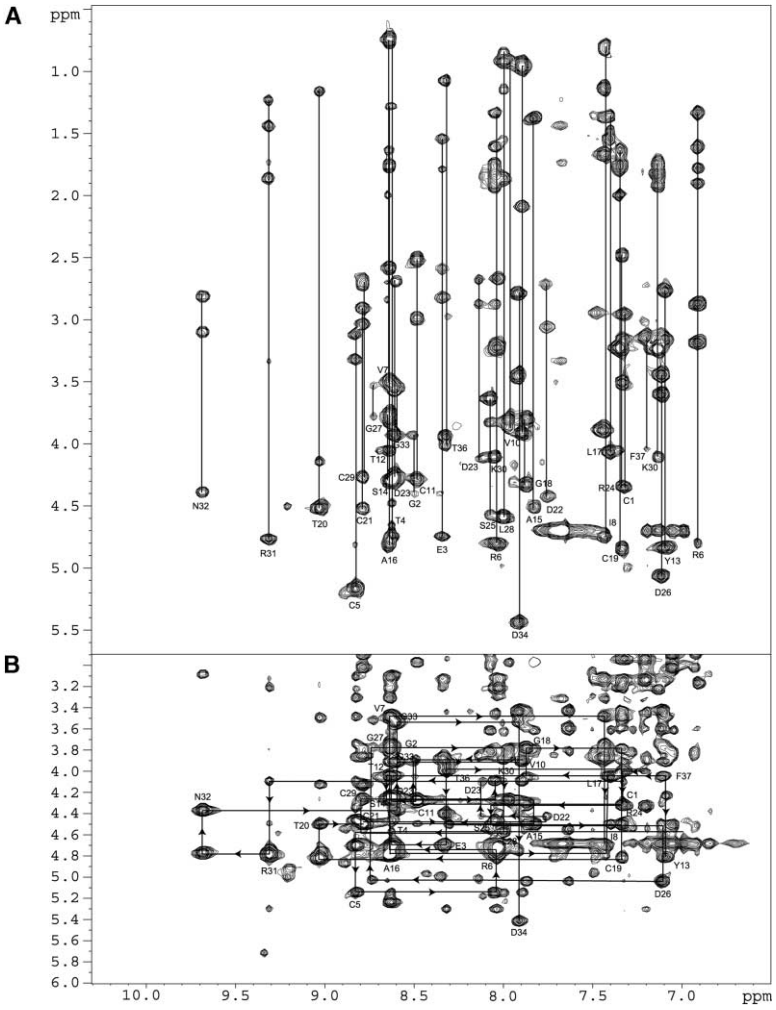

Figure 2. Two-Dimensional ^{1}H-NMR Spectra of Palicourein

(A) The 750 MHz TOCSY spectrum of palicourein at 295 K in 90% H_2O/10% ^2H_2O at pH 4.0. Each spin-system of the peptide was identified using standard protocols and is labeled accordingly.

(B) Fingerprint region of the 750 MHz NOESY spectrum of palicourein at 295 K in 90% H_2O/10% ^2H_2O at pH 4.0. The sequential connectivity pattern is broken only at the proline residues.

recent study suggested an alternative disulfide bonding pattern (I-VI, II-V, III-IV) for native kalata B1 based on an alternative interpretation of 750 MHz NMR data (Skjeldal et al., 2002). We believe that this alternative is unlikely, as studies of RTD-1 (Trabi et al., 2001), another tri-disulfide cyclic peptide with such a "laddered" arrangement of disulfide bonds, have shown that this arrangement leads to more disordered and flexible structures than is seen for kalata B1. Nevertheless, extra attention was directed at assignment of the disulfide connectivity of palicourein. In particular, dihedral angle information was valuable in discriminating between alternative disulfide bond connectivities. The side chain dihedral angles χ^1 for the Cys residues were determined by combining $^3J_{H\alpha H\beta}$ coupling constants from an ECOSY spectrum, and interproton HN-Hα distances determined from a NOESY spectrum (Wagner, 1990). From these results, together with the structure calculations described below, the disulfide connectivity best fitting the experimental data was determined to be Cys1-Cys19, Cys5-Cys21, and Cys11-Cys28, i.e., I-IV, II-V, III-VI, as for the original assignment in native kalata B1. A recent report on the high-resolution structure of kalata B1 using similar methods has reaffirmed this knotted disulfide bond connectivity for the cyclotides (Rosengren et al., 2003).

Structure Determination

Solution structures were determined for palicourein by simulated annealing including experimental distance restraints and dihedral angle restraints, with refinement by the explicit inclusion of water as the solvent. Of the final 50 calculated structures, the 20 lowest energy structures consistent with experimental data were chosen to represent the family of palicourein solution structures. A summary of the structural statistics for this family is given in Table 1. The structures are in agreement with the experimental data, with no distance violation exceeding 0.2 Å and no dihedral angle violation exceeding 2°. The palicourein solution structure is of high precision, as reflected in the backbone superposition of the final 20 structures in Figure 5 and the low root mean square deviation (rmsd) of backbone and heavy atoms for the final family of palicourein structures, described in Table 1. Excellent structural convergence extended to the majority of the side chains too, providing detailed structural information for the positioning of side chains.

An analysis of the family of structures using the program PROMOTIF (Hutchinson and Thornton, 1996) shows that the main element of secondary structure of palicourein is an antiparallel β sheet, consistent with the prediction from the NMR chemical shift, coupling, and

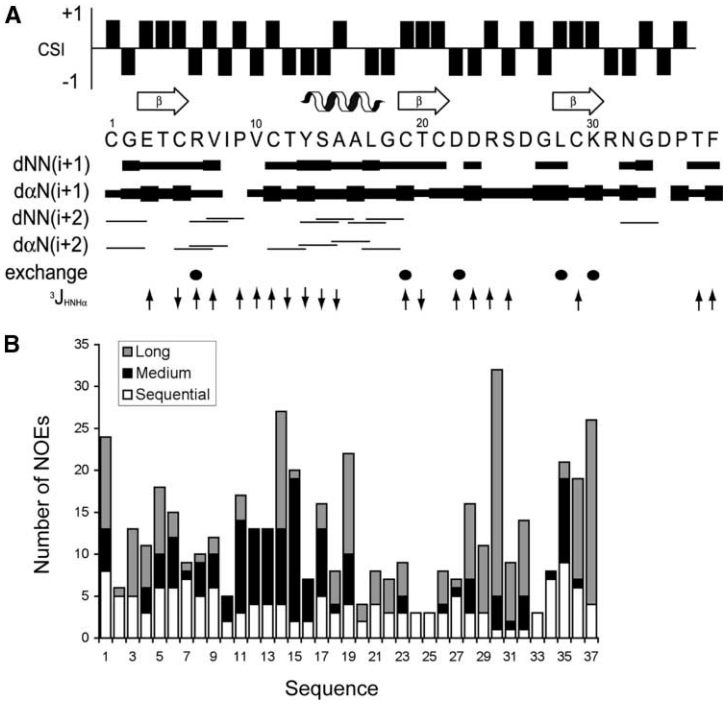

Figure 3. Summary of the Sequential, Medium, and Long-Range NOE Connectivities for Palicourein

(A) The chemical shift index (CSI; Wishart et al., 1992) for each residue is shown as a barplot, with values of ±1 indicating a shift deviation from random coil values of greater than 0.1 ppm. Short-range NOEs are shown beneath the sequence. The thickness of the filled bars indicates relative NOE intensities. Slowly exchanging backbone amide protons with $t_{1/2}$ >20 hr are indicated by filled circles. The $^3J_{HN-H\alpha}$ coupling constants above 8 Hz are shown as arrows pointing upwards, and those below 6 Hz are represented as arrows pointing downwards. Secondary structure elements are shown below the CSI barplot across the residues involved. The β strand secondary structural elements are represented as arrows, and the helix is represented as a short coil.

(B) Plot of sequential, medium, and long-range NOEs versus the palicourein sequence. Sequential NOEs refer to those from sequentially adjacent residues, medium-range NOEs refer to those from between two and five residues apart, and long-range NOEs refer to those from residues more than five residues apart.

NOE data. Strand 1 (residues 19–21) and strand 2 (residues 28–30) are joined by a type IV β turn. A third β strand, incorporating residues 3–5, is loosely associated with strand 2 but is not formally recognized as part of the β sheet by PROMOTIF (Hutchinson and Thornton, 1996). A series of well-ordered turns accompany the β sheet structure. In the view shown in Figure 5, a type I β turn comprising residues 6–9 lies directly beneath a type IV turn (residues 23–26) and a classical γ turn in loop 6 (residues 32–34). Loop 3 of palicourein adopts a 3_{10}-helical conformation for residues 14–17. These structural elements are surrounded by a stable network of hydrogen bonds and hydrophobic interactions and, together with the knotted arrangement of disulfide bonds, account for the high definition of the palicourein structure.

The high-resolution solution structure of palicourein provides an opportunity to investigate the roles of the individual amino acids in forming the native structure. Of primary importance are the six absolutely conserved cysteine residues that form three disulfide bonds. Each of the disulfide bonds adopts a single type of conformation, as classified by PROMOTIF (Hutchinson and Thornton, 1996). The disulfide bonds that form part of the embedded octapeptide ring, Cys1-Cys19 and Cys5-Cys21, adopt left-handed spiral and right-handed spiral conformations, respectively, and the Cys11-Cys29 penetrating disulfide bond adopts a left-handed spiral con-

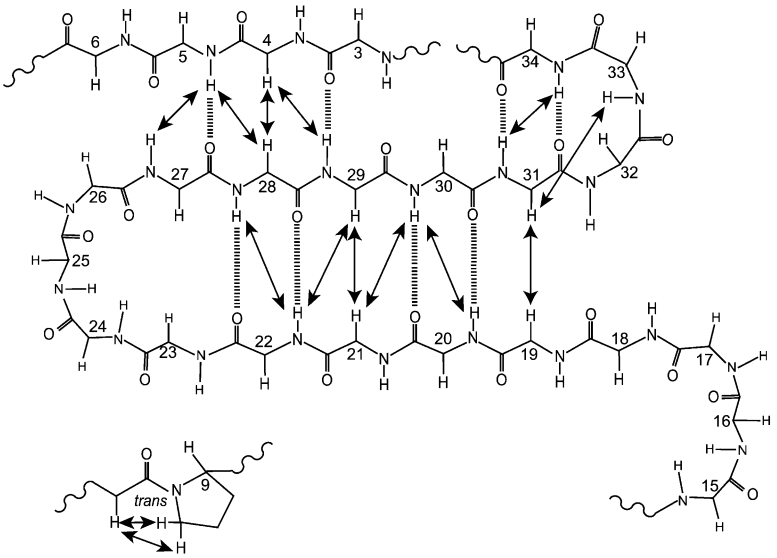

Figure 4. Schematic Diagram of the Secondary Structure of Palicourein Showing the Interstrand NOEs as Continuous Arrows and Potential Hydrogen Bonds as Broken Lines

Interresidual HN-HN, Hα-HN and Hα-Hα NOEs are shown with a double-headed arrow. The peptide bonds between the proline residues and their preceding residues are labeled *trans*. The α-carbon atoms are labeled according to their residue numbers. For clarity, sequential NOEs are omitted. The hydrogen bonds were inferred from slow-exchange data and preliminary structure calculations, as described.

Table 1. Geometric and Energetic Statistics for Palicourein

Energies (kcal mol^{-1})	
Overall	-882 ± 32
Bond	8.48 ± 0.49
Angle	49.6 ± 3.9
Improper	13.76 ± 0.8
Van der Waal	15.8 ± 1.1
cDih	0.287 ± 0.177
Dihedral	90.3 ± 6.9
Electrostatic	-811 ± 14
Rmsd (Å)	
Bond	0.006 ± 0.0002
Angle	0.72 ± 0.047
Improper	0.68 ± 0.04
NOE	0.027 ± 0.003
cDih	0.07 ± 0.07
Pairwise rmsd (Å)	
Backbone atoms	0.48 ± 0.12
Heavy atoms	0.96 ± 0.15
Experimental data	
Distance restraints	345
Dihedral restraints	33
NOE violations >0.20 Å	0
Dihedral violations $>2.0^\circ$	0
Ramachandran statistics (%)	
Most favored	82
Additionally allowed	18

The values are given as mean $\pm$ standard deviations for the ensemble of the 20 final solution structures. Experimental distance restraints include only interresidual NOEs.

formation, as illustrated in Figure 6A. The octapeptide ring embedded in the cyclotides is the smallest hole through which a disulfide bond could thread and presumably contributes to the exceptional stability of the cyclotides.

Apart from the six cysteine residues, Glu3 is one of a few residues fully conserved throughout the cyclotide family (Craik et al., 2002a; Rosengren et al., 2003). The Glu3 side chain is oriented in a way that suggests the presence of hydrogen bonds between the oxygen atoms of the Glu3 carboxyl group and the backbone amides of residues 12, 13, and 14, and the side chain of Ser14. These interactions stabilize the short 3_{10} helix of loop 3, as illustrated in Figure 6B. The low-field amide chemical shifts of residues 12, 13, and 14 ($\sim$8.8 ppm) relative to the other amides of palicourein support the suggestion of the presence of these hydrogen bonds. Similar interactions have been reported for kalata B1 and cycloviolacin O1 (Rosengren et al., 2003). Overall, the palicourein structure adopts well-defined loop regions, including the residues of loop 5 that are atypical in composition and length when compared with the previously known cyclotides.

Discussion

The three-dimensional solution structure of the macrocyclic peptide palicourein determined here is well defined and confirms that it is a member of the cyclotide family of CCK-containing peptides. In addition to its cystine knot core, the palicourein structure contains antiparallel β strands consistent with other known cyclotide structures, even though it has a sequence length in loop 5 almost double that of the other cyclotides. It also incorporates a short 3_{10} helix in loop 3. Outside of the regular secondary structure, the turns of palicourein are well defined, a result of the topology enforced by the CCK motif. The high convergence of the family of structures suggests that the CCK motif of palicourein maintains a rigid framework. This confirms that, in principle, the loops of the CCK framework may be expanded to accommodate variable numbers and types of amino acids (charged and uncharged), offering an adaptable scaffold for epitope presentation.

The structure adopted by palicourein in solution is similar to the structures of the cyclotides kalata B1 (Saether et al., 1995; Rosengren et al., 2003), circulin A (Daly et al., 1999), and cycloviolacin O1 (Craik et al., 1999; Rosengren et al., 2003). All contain β strands forming a β hairpin, with a third strand inferred from NOE, coupling, and hydrogen bonding patterns. The third strand is not formally recognized by PROMOTIF (Hutchinson and Thornton, 1996). In each cyclotide, the cyclized peptide backbone is extensively folded back upon itself and is braced with disulfide bonds, two of which form an embedded octapeptide ring that is completed by the backbone segments between the cysteine residues. A third disulfide bond threads through the center of this ring to form the cystine knot motif. The cystine knot motif has been found in a range of biologically active proteins, including nerve growth factor (McDonald and Hendrickson, 1993) and many toxins (Craik et al., 2001).

The rmsd between the backbone atoms of the lowest energy structure of palicourein and the corresponding residues of kalata B1, circulin A, and cycloviolacin O1 is 1.21, 1.25, and 1.31 Å, respectively. The rmsd between the α carbon atoms in the cystine knot of palicourein and that of kalata B1, circulin A, and cycloviolacin O1 is 0.74, 0.81, and 0.91 Å, respectively. It is clear that the structures are very similar, particularly over the cystine knot region. It is noteworthy that in addition to the Cys residues, all of the backbone residues of loops 1 and 4, which make up the embedded ring of the cystine knot core, are highly conserved in all of the known cyclotides. In palicourein, loop 1 comprises the sequence Gly-Glu-Thr, while loop 4 contains just a Thr residue. Of these residues the Glu is completely conserved, and the only variations seen for the other residues are highly conservative substitutions, including Ala/Gly and Thr/Ser. In the case of the Glu residue, the conservation appears to be due to a structural requirement for hydrogen bonding interactions of the side chain carboxyl group with backbone amides of loop 3. Steric requirements for a small amino acid (Gly, Ala) appear to be important at the first position of loop 1 (Rosengren et al., 2003), while hydroxylated amino acids are required at the other positions.

The most significant differences between palicourein and the other cyclotides for which structures have so far been determined are in loop 5, which contains seven residues in palicourein relative to four in most other cyclotides. The overlay of palicourein with kalata B1 and cycloviolacin O1 in Figure 7 emphasizes this difference in the size of loop 5. Further, the nature of the residues is quite different. Palicourein has three acidic (Asp) residues and one basic (Arg) residue in loop 5, compared

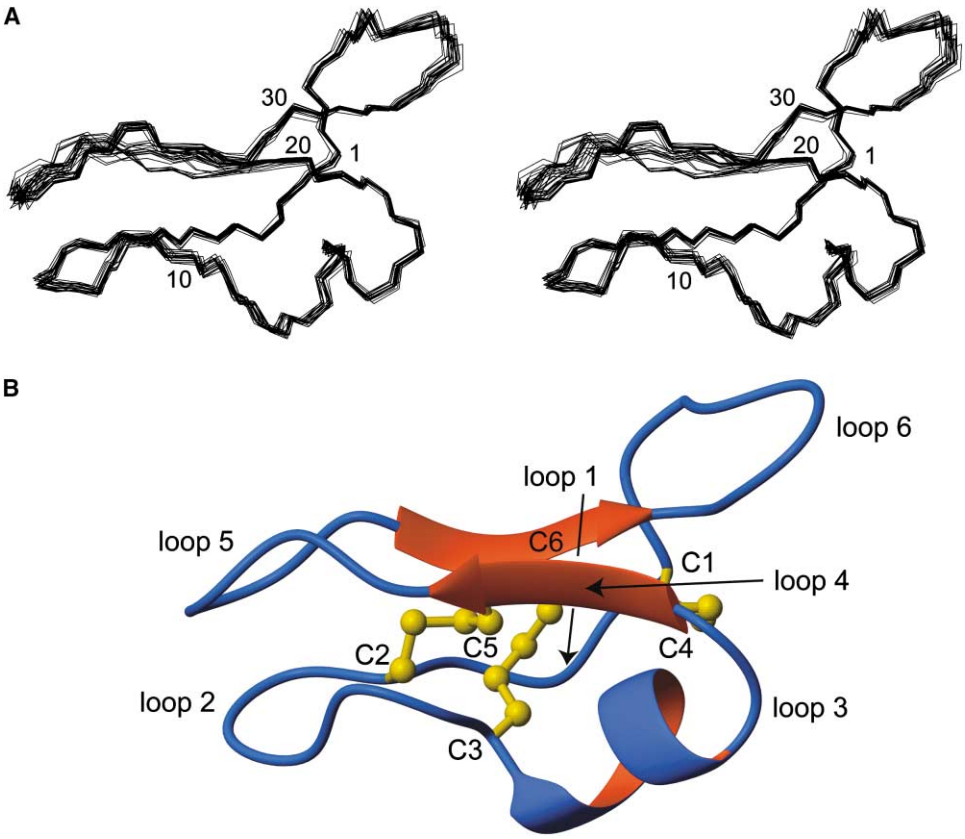

Figure 5. The Solution Structure of Palicourein

(A) The 20 lowest energy solution structures of palicourein are superimposed over the backbone atoms (N, C^{α}, and C^{η}), with every tenth residue numbered for reference in a stereoscopic view.

(B) Ribbon diagram of the lowest energy solution structure of palicourein. Secondary structural elements identified by PROMOTIF (Hutchinson and Thornton, 1996) are shown, with the β strands as red arrows and the helix as blue and red coil. The intercysteine loops are labeled. The cysteine residues are numbered according to their order within the sequence, shown in Figure 1. The three disulfide bonds are represented with yellow ball-and-stick rendering. The figure was created using the program MOLMOL (Koradi et al., 1996).

to the predominantly nonpolar nature of loop 5 in other cyclotides. It may also be significant that many cyclotides contain a Trp residue in either loop 2 or loop 5, which form part of a hydrophobic patch near the turns at the ends of the these loops. However, this Trp residue is absent from palicourein. The presence of charged residues within loop 5, specifically Asp22, Asp23, Arg24, and Asp26, and the well defined nature of the turns in loops 2 and 5 suggest that the cyclotide framework does not rely on a single type of charge distribution for folding or stability and can accommodate vastly different loop sequences without altering the overall characteristic CCK fold.

Another feature of loop 5 is that in the Möbius subfamily of cyclotides it typically contains a Pro residue preceded by a *cis* rather than the usual *trans* peptide bond (indeed, this 180° twist in the ω angle of the backbone is what gives the subfamily its name). The Pro residue is not present in loop 5 of palicourein, and all peptide bonds are in the *trans* conformation, making palicourein a member of the bracelet subfamily of cyclotides. With the structure of palicourein now available, another distinguishing feature of the bracelet subfamily becomes apparent: the 3_{10} helix seen in loop 3 also occurs in the bracelet subfamily members cycloviolacin O1 (Rosengren et al., 2003) and circulin A (Daly et al., 1999) and thus appears to be a characteristic of the family. It is interesting that chimeric sequences that contain bracelet-like loop 3 sequences together with, for example, Möbius-like loop 5 sequences have so far not been discovered. This is surprising since loops 3 and 5 are spatially well resolved from each other (e.g., see Figure 5) and a mutual dependency seems unlikely. As further sequences are discovered, the significance of these trends in explaining the evolutionary development of cyclotides should become apparent.

Loop 2 of palicourein contains a segment of three nonpolar residues (Val7, Ile8, and Pro9) preceded by the basic Arg6 residue. Palicourein is the only known cyclotide that contains a charged residue within loop 2, despite many cyclotides containing charged residues within loop 5. The side chain of Arg6 adopts a conformation placing it near Asp26, permitting a possible stabilizing interaction.

The greater sequence homology between palicourein and cycloviolacin O1 (and circulin A), relative to kalata B1, should be reflected in the amino acid orientations (i.e., dihedral angles). The loop 2 residues of palicourein

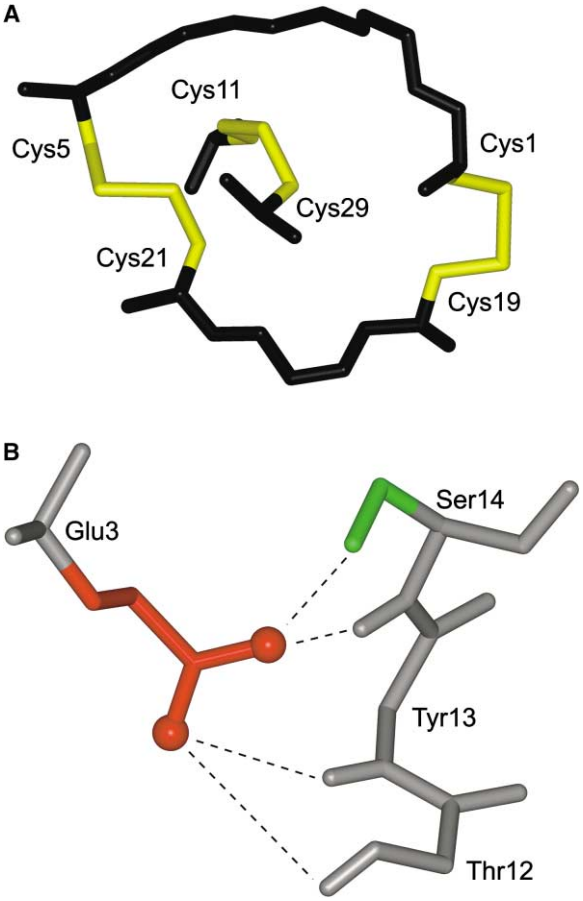

Figure 6. Structural Features of Palicourein

The geometry of the embedded ring of the cystine knot within palicourein is shown in (A), which is made up of the backbone segments of loop 1 (comprising residues Cys1-Cys5) and loop 4 (comprising residues Cys19-Cys21) and the adjoining disulfide bonds, and is highly conserved throughout the cyclotide family. (B) Illustration of the putative hydrogen bonding pattern between the Glu3 side chain and residues of loop 3, including the backbone amide protons of Thr12, Tyr13, and Ser14, and the side chain of Ser14 (side chain shown in green). The residues are labeled next to the α-carbon atoms.

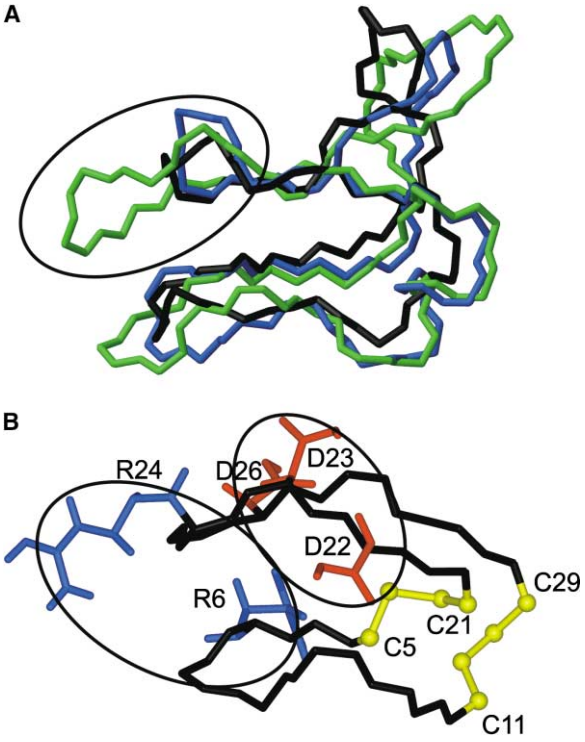

Figure 7. Comparison of Palicourein with the Other Known Cyclotides

(A) A backbone superposition of the lowest energy structures of palicourein (green), kalata B1 (black), and cycloviolacin O1 (blue), highlighting the extended loop 5 of palicourein. The structures were superimposed over the cysteine α-carbon atoms.

(B) This panel identifies the residues of loop 2 and loop 5 in palicourein that are significantly different from the other known cyclotides. Palicourein contains multiple cationic (Arg6 and Arg24) and anionic (Asp22, Asp23, and Asp26) residues within loop 2 and loop 5, whereas the other cyclotides exhibit large hydrophobic patches in this region.

exhibit backbone and side chain orientations that are comparable to other bracelet cyclotides, in spite of minor sequence variations. The high sequence homology of loop 3 in palicourein, circulin A, and cycloviolacin O1 is associated with the formation of a short 3_{10} helix in these three cyclotides. Palicourein and cycloviolacin O1 have very similar dihedral angles for the nonhomologous residues of loop 3, suggesting that the short helical segment within loop 3 can tolerate amino acid redundancy. Consistent with a lack of sequence homology, the residues of loop 5 of palicourein adopt backbone and side chain orientations that are quite different from those of the other cyclotides. Nevertheless the overall molecular folds are similar. This clearly demonstrates that the CCK motif can accommodate varied loop regions with no disruption to the core fold.

The electrostatic potential surfaces of palicourein and two other representative cyclotides are illustrated in Figure 8. In general, proteins have their hydrophobic side chains packed tightly together in a hydrophobic core. However, in the cyclotides the core is occupied by the cysteine residues, forcing the surface exposure of many hydrophobic residues. Figure 8 makes it clear that the previously studied cyclotides have a large proportion of solvent-exposed hydrophobic residues, indicated by the large areas of uncolored (white) surfaces. These hydrophobic patches are prominent within kalata B1, involving residues Val6, Pro20, Val21, Pro28, and the methyl group of Thr4 (Rosengren et al., 2003), with equivalent patches present in cycloviolacin O1 and circulin A, corresponding to residues within loops 2, 5, and 6. Palicourein differs from the other cyclotides with an increased number of charged amino acids. Palicourein is the only known cyclotide to contain a cationic amino acid within loop 2 (Arg6), which in general is hydrophobic. Of particular interest is the concentration of charges within loops 5 and 6 of palicourein. Four anionic charges Asp22, Asp23, Asp26, and Asp34 and three cationic charges Arg24, Lys30, and Arg31 form an interspersed arrangement of charges within loops 5 and 6. It is interesting that despite their quite different surface charge

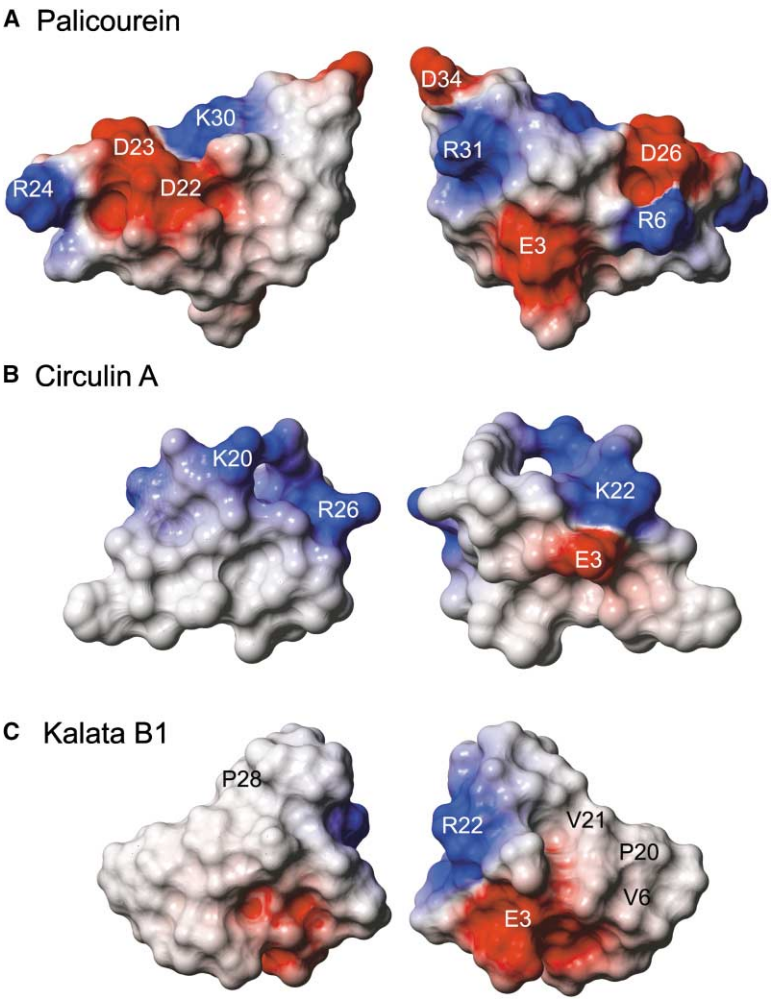

Figure 8. Electrostatic Potential Surface Comparison of the Cyclotides

(A) Palicourein, (B) circulin A, and (C) kalata B1 electrostatic potential surfaces are represented with two views (rotated 90° about the *y* axis). Anionic charges are colored red, cationic charges are blue, and the hydrophobic regions are white. Residues that contribute to the electrostatic potential surface are labeled according to the sequences in Figure 1. Electrostatic potential surfaces were prepared using the program GRASP (Nicholls et al., 1991).

distributions, both palicourein and circulin A (Figure 8) have potent anti-HIV activity. They inhibit the in vitro cytopathic effects of HIV-1RF infection of CEM-SS cells with EC_{50} values of 40–100 nM, relative to an AZT control EC_{50} of 50 nM (Gustafson et al., 1994; Bokesch et al., 2001). While the mechanism of antiviral activity for the cyclotides has not been elucidated, it appears that HIV inhibition may be more related to one specific region of the cyclotide framework, rather than the overall charge distribution. Further mutagenesis studies will be necessary to locate the region of the framework responsible for the anti-HIV activity.

Use of the Cyclotides as Scaffolds

The highly constrained cyclotides are interesting candidates for protein engineering. Because the disulfide bonds and cyclic peptide backbone provide exceptional stability, a large part of the protein surface is available for residue replacement or *grafting* experiments. Grafting the sequence of bioactive peptide fragments into natural structural scaffolds represents an effective means to engineer novel proteins displaying a specific function with a predetermined structure (Vita, 1997; Martin et al., 2003). In the case of the cyclotides, loop regions not involved in the embedded octapeptide ring could potentially be used for grafting experiments. Cyclotides are likely to be most applicable for the grafting of short (<10 amino acids) peptide epitopes. They have the advantage that multiple epitopes may be simultaneously grafted into exposed loops to provide a contiguous binding surface (loops 2 and 5 are proximate). The highly constrained cyclotide structure will likely ensure only a small entropy loss upon binding and thus allow high binding affinities of the grafted peptide sequence(s). Additionally, the cyclotide scaffold has the potential to present at its surface hydrophobic residues that would otherwise be buried in a more flexible peptide.

The evolutionarily conserved cyclic cystine knot motif appears to be quite permissive for sequence variability, since few residues are conserved in the cyclotides other than the six cysteines involved in the disulfide bridges. Apart from these residues, the cyclotide scaffold has the ability to accommodate numerous amino acid substitutions, and the only requirement is that the residues are solvent exposed. The most destabilizing amino acid substitutions in proteins usually involve buried residues, and these are the very residues that constitute the conserved core of the cyclotide scaffold. These features suggest that permissive loops in the physiologically stable cyclotide structure could serve as a platform for

displaying biologically active fragments of other proteins and peptides.

Experimental Procedures

Isolation of Palicourein

Palicourein was extracted from the bark of the tree *Palicourea condensata* Standley (Rubiaceae) as described previously (Bokesch et al., 2001). Dried plant material was ground and extracted with 1:1 MeOH-CH_2Cl_2 followed by 100% MeOH, and the combined organic extracts were evaporated in vacuo. The organic extract was subjected to solvent-solvent partitioning that concentrated the anti-HIV activity in the water-soluble fraction. Size exclusion chromatography of the active material on Sephadex LH-20 with MeOH-H_2O (7:3) provided an early eluting fraction, which was further separated by vacuum-liquid chromatography on a wide pore C4 media with a 0%–100% MeOH in H_2O gradient. The 100% MeOH fraction containing palicourein was purified by reversed-phase HPLC eluting with a 25%–60% CH_3CN in H_2O gradient.

NMR Sample Preparation

The sample of palicourein was dissolved in either 90% H_2O/10% ^2H_2O or 100% ^2H_2O to a final concentration of 3.8 mM at pH 3.8. All spectra were recorded on Bruker ARX 500 and Bruker DMX 750 spectrometers with sample temperatures in the range 280–330K. All spectra were acquired in phase-sensitive mode using time proportional phasing incrementation (Marion and Wüthrich, 1983). For resonance assignment and structure determination, a set of two-dimensional spectra including double quantum-filtered COSY (Rance et al., 1983), TOCSY (Braunschweiler and Ernst, 1983) with a MLEV17 (Bax and Davis, 1985) isotropic mixing period of 80 ms, ECOSY (Griesinger et al., 1987), and NOESY (Jeener et al., 1979) with mixing times of 100, 150, and 200 ms were recorded. The water signal in the COSY spectrum was suppressed by low-power irradiation during the relaxation delay. For the TOCSY and NOESY spectra, water suppression was achieved using a modified WATERGATE (water suppression by gradient-tailored excitation; Piotto et al., 1992) sequence. All two-dimensional spectra were collected over 4096 data points in the f2 dimension and 512 or 600 increments in the f1 dimension and processed using XWINNMR (Bruker). The f1 dimension was generally zero-filled to 2048 real data points, with the f1 and f2 dimensions being multiplied by a sine-squared function shifted by 90 degrees prior to Fourier transformation. Chemical shifts were internally referenced to sodium 2,2-dimethyl-2-silapentane-5-sulfonate.

Structure Calculations

Interproton distance restraints for palicourein were derived from cross-peaks in a NOESY spectrum recorded with a mixing time of 200 ms. The cross-peaks were analyzed and integrated within the program XEASY (Eccles et al., 1991), and structures were calculated within the DYANA package (Güntert et al., 1997). After an iterative process where preliminary structures were used to resolve ambiguities, a set of 345 interresidual distance restraints, including 141 sequential, 72 medium, and 132 long-range restraints, was derived for palicourein. Additionally, the spectral data allowed the inclusion of 20 backbone dihedral angle restraints based on $^3J_{HNH\alpha}$ coupling constants derived from the splitting of the amide signals in the one-dimensional spectra and 13 χ^1 dihedral angles based on $^3J_{H\alpha H\beta}$ coupling constants derived from an ECOSY spectrum together with NOE intensities derived from a 100 ms NOESY spectrum. Following the analysis of preliminary structures, restraints for five hydrogen bonds, from amide exchange experiments, were included in calculations. After initial structure calculations using DYANA (Güntert et al., 1997), sets of 50 structures were calculated using a torsion angle simulated annealing protocol within CNS 1.0 (Brünger et al., 1997). This protocol involves a high-temperature phase comprising 4000 steps of 0.015 ps of torsion angle dynamics, a cooling phase with 4000 steps of 0.015 ps of torsion angle dynamics during which the temperature is lowered to 0 K, and finally an energy minimization phase comprising 500 steps of Powell minimization. The resultant structures were subjected to further molecular dynamics and energy minimization in a water shell (Linge and Nilges, 1999). The refinement in explicit water involves the following steps: first, heating to 500 K via steps of 100 K, each comprising 50 steps of 0.005 ps of Cartesian dynamics; second, 2500 steps of 0.005 ps of Cartesian dynamics at 500 K before a cooling phase where the temperature is lowered in steps of 100 K, each comprising 2500 steps of 0.005 ps of Cartesian dynamics; finally, the structures were minimized with 2000 steps of Powell minimization. Structures were analyzed using the programs PROMOTIF (Hutchinson and Thornton, 1996) and PROCHECK-NMR (Laskowski et al., 1996).

Acknowledgments

D.G.B. is supported by an Australian Postgraduate Award. D.J.C. is an Australian Research Council Senior Fellow. This project has been funded in part with federal funds from the U.S. National Cancer Institute, National Institutes of Heath, under contract no. NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Received: September 4, 2003
Revised: September 26, 2003
Accepted: October 5, 2003
Published: January 13, 2004

References

Bax, A., and Davis, D.G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J. Magn. Reson. *65*, 355–360.

Blond, A., Cheminant, M., Segalas-Milazzo, I., Peduzzi, J., Barthelemy, M., Goulard, C., Salomon, R., Moreno, F., Farias, R., and Rebuffat, S. (2001). Solution structure of microcin J25, the single macrocyclic antimicrobial peptide from *Escherichia coli*. Eur. J. Biochem. *268*, 2124–2133.

Bokesch, H.R., Pannell, L.K., Cochran, P.K., Sowder, R.C., 2nd, McKee, T.C., and Boyd, M.R. (2001). A novel anti-HIV macrocyclic peptide from *Palicourea condensata*. J. Nat. Prod. *64*, 249–250.

Braunschweiler, L., and Ernst, R.R. (1983). Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. J. Magn. Reson. *53*, 521–528.

Brünger, A.T., Adams, P.D., and Rice, L.M. (1997). New applications of simulated annealing in X-ray crystallography and solution NMR. Structure *5*, 325–336.

Camarero, J.A., Fushman, D., Sato, S., Giriat, I., Cowburn, D., Raleigh, D.P., and Muir, T.W. (2001). Rescuing a destabilized protein fold through backbone cyclization. J. Mol. Biol. *308*, 1045–1062.

Claeson, P., Göransson, U., Johansson, S., Luijendijk, T., and Bohlin, L. (1998). Fractionation protocol for the isolation of polypeptides from plant biomass. J. Nat. Prod. *61*, 77–81.

Craik, D.J. (2001). Plant cyclotides: circular, knotted peptide toxins. Toxicon *39*, 1809–1813.

Craik, D.J., Daly, N.L., Bond, T., and Waine, C. (1999). Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. J. Mol. Biol. *294*, 1327–1336.

Craik, D.J., Daly, N.L., and Waine, C. (2001). The cystine knot motif in toxins and implications for drug design. Toxicon *39*, 43–60.

Craik, D.J., Anderson, M.A., Barry, D.G., Clark, R.J., Daly, N.L., Jennings, C.V., and Mulvenna, J. (2002a). Discovery and structures of the cyclotides: novel macrocyclic peptides from plants. Lett. Pept. Sci. *8*, 119–128.

Craik, D.J., Simonsen, S., and Daly, N.L. (2002b). The cyclotides: novel macrocyclic peptides as scaffolds in drug design. Curr. Opin. Drug Discov. Dev. *5*, 251–260.

Daly, N.L., and Craik, D.J. (2000). Acyclic permutants of naturally occurring cyclic proteins. Characterization of cystine knot and β-sheet formation in the macrocyclic polypeptide kalata B1. J. Biol. Chem. *275*, 19068–19075.

Daly, N.L., Koltay, A., Gustafson, K., Boyd, R., Casas-Finet, M.R.,

Jr., and Craik, D.J. (1999). Solution structure by NMR of circulin A: a macrocyclic knotted peptide having anti-HIV activity. J. Mol. Biol. *285*, 333–345.

Deechongkit, S., and Kelly, J.W. (2002). The effect of backbone cyclization on the thermodynamics of β-sheet unfolding: stability optimization of the PIN WW domain. J. Am. Chem. Soc. *124*, 4980–4986.

Eccles, C., Güntert, P., Billeter, M., and Wüthrich, K. (1991). Efficient analysis of protein 2D NMR spectra using the software package EASY. J. Biomol. NMR *1*, 111–130.

Göransson, U., Luijendijk, T., Johansson, S., Bohlin, L., and Claeson, P. (1999). Seven novel macrocyclic polypeptides from *Viola arvensis*. J. Nat. Prod. *62*, 283–286.

Gran, L. (1973). Oxytocic principles of *Oldenlandia affinis*. Lloydia *36*, 174–178.

Griesinger, C., Sørensen, O.W., and Ernst, R.R. (1987). Practical aspects of the E.COSY technique, measurement of scalar spin-spin coupling constants in peptides. J. Magn. Reson. *75*, 474–492.

Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. *273*, 283–298.

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., and Boyd, M.R. (1994). Circulins A and B: novel HIV-inhibitory macrocyclic peptides from the tropical tree *Chassalia parvifolia*. J. Am. Chem. Soc. *116*, 9337–9338.

Gustafson, K.R., Walton, L.K., Sowder, R.C.I., Johnson, D.G., Pannell, L.K., Cardellina, J.H.I., and Boyd, M.R. (2000). New circulin macrocyclic polypeptides from *Chassalia parvifolia*. J. Nat. Prod. *63*, 176–178.

Hutchinson, E.G., and Thornton, J.M. (1996). PROMOTIF: a program to identify and analyze structural motifs in proteins. Protein Sci. *5*, 212–220.

Iwai, H., Lingel, A., and Pluckthun, A. (2001). Cyclic green fluorescent protein produced *in vivo* using an artificially split PI-Pful intein from *Pyrococcus furiosus*. J. Biol. Chem. *276*, 16548–16554.

Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. J. Chem. Phys. *71*, 4546–4553.

Jennings, C., West, J., Waine, C., Craik, D., and Anderson, M. (2001). Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis*. Proc. Natl. Acad. Sci. USA *98*, 10614–10619.

Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. *14*, 51–55, 29–32.

Korsinczky, M.L., Schirra, H.J., Rosengren, K.J., West, J., Condie, B.A., Otvos, L., Anderson, M.A., and Craik, D.J. (2001). Solution structures by ^{1}H NMR of the novel cyclic trypsin inhibitor SFTI-1 from sunflower seeds and an acyclic permutant. J. Mol. Biol. *311*, 579–591.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR *8*, 477–486.

Linge, J.P., and Nilges, M. (1999). Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. J. Biomol. NMR *13*, 51–59.

Marion, D., and Wüthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. Biochem. Biophys. Res. Commun. *113*, 967–974.

Martin, L., Stricher, F., Misse, D., Sironi, F., Pugniere, M., Barthe, P., Prado-Gotor, R., Freulon, I., Magne, X., and Roumestand, C. (2003). Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. Nat. Biotechnol. *21*, 71–76.

McDonald, N.Q., and Hendrickson, W.A. (1993). A structural superfamily of growth factors containing a cystine knot motif. Cell *73*, 421–424.

Merutka, G., Dyson, H.J., and Wright, P.E. (1995). "Random Coil" ^{1}H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. J. Biomol. NMR *5*, 14–24.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

Pearce, J. (2001). Going round in circles to avoid proteolysis. Trends Biochem. Sci. *26*, 282.

Piotto, M., Saudek, V., and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J. Biomol. NMR *2*, 661–665.

Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., and Wüthrich, K. (1983). Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. Biochem. Biophys. Res. Commun. *117*, 479–485.

Rosengren, K.J., Daly, N.L., Plan, M.R., Waine, C., and Craik, D.J. (2003). Twists, knots and rings in proteins: structural definition of the cyclotide framework. J. Biol. Chem. *278*, 8606–8616.

Saether, O., Craik, D.J., Campbell, I.D., Sletten, K., Juul, J., and Norman, D.G. (1995). Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. Biochemistry *34*, 4147–4158.

Samyn, B., Martinez-Bueno, M., Devreese, B., Maqueda, M., Galvez, A., Valdivia, E., Coyette, J., and Van Beeumen, J. (1994). The cyclic structure of the enterococcal peptide antibiotic AS-48. FEBS Lett. *352*, 87–90.

Siebold, C., and Erni, B. (2002). Intein-mediated cyclization of a soluble and a membrane protein *in vivo*: function and stability. Biophys. Chem. *96*, 163–171.

Skjeldal, L., Gran, L., Sletten, K., and Volkman, B.F. (2002). Refined structure and metal binding site of the kalata B1 peptide. Arch. Biochem. Biophys. *399*, 142–148.

Tang, Y.-Q., Yuan, J., Ösapay, G., Ösapay, K., Tran, D., Miller, C.J., Ouellette, A.J., and Selsted, M.E. (1999). A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated α-defensins. Science *286*, 498–502.

Trabi, M., and Craik, D.J. (2002). Circular proteins—no end in sight. Trends Biochem. Sci. *27*, 132–138.

Trabi, M., Schirra, H.J., and Craik, D.J. (2001). Three-dimensional structure of RTD-1, a cyclic antimicrobial defensin from Rhesus macaque leukocytes. Biochemistry *40*, 4211–4221.

Vita, C. (1997). Engineering novel proteins by transfer of active sites to natural scaffolds. Curr. Opin. Biotechnol. *8*, 429–434.

Wagner, G. (1990). NMR investigations of protein structure. Prog. NMR Spectrosc. *22*, 101–139.

Wishart, D.S., Sykes, B.D., and Richards, F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry *31*, 1647–1651.

Witherup, K.M., Bogusky, M.J., Anderson, P.S., Ramjit, H., Ransom, R.W., Wood, T., and Sardana, M. (1994). Cyclopsychotride A, a biologically active, 31-residue cyclic peptide isolated from *Psychotria Longipes*. J. Nat. Prod. *57*, 1619–1625.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids (New York: Wiley-Interscience).

Accession Numbers

The atomic coordinates for the solution structure of palicourein have been deposited in the Protein Data Bank (accession code 1R1F).