Kalata B8, a novel antiviral circular protein, exhibits conformational flexibility in the cystine knot motif

Norelle L. DALY, Richard J. CLARK, Manuel R. PLAN and David J. CRAIK

Institute for Molecular Bioscience, Australian Research Council Special Research Centre for Functional and Applied Genomics, University of Queensland, Brisbane, QLD 4072, Australia

The cyclotides are a family of circular proteins with a range of biological activities and potential pharmaceutical and agricultural applications. The biosynthetic mechanism of cyclization is unknown and the discovery of novel sequences may assist in achieving this goal. In the present study, we have isolated a new cyclotide from *Oldenlandia affinis*, kalata B8, which appears to be a hybrid of the two major subfamilies (M"obius and bracelet) of currently known cyclotides. We have determined the three-dimensional structure of kalata B8 and observed broadening of resonances directly involved in the cystine knot motif, suggesting flexibility in this region despite it being the core structural element of the cyclotides. The cystine knot motif is widespread throughout Nature and inherently stable, making this apparent flexibility a surprising result. Furthermore, there appears to be isomerization of the peptide backbone at an Asp-Gly sequence in the region involved in the cyclization process. Interestingly, such isomerization has been previously characterized in related cyclic knottins from *Momordica cochinchinensis* that have no sequence similarity to kalata B8 apart from the six conserved cysteine residues and may result from a common mechanism of cyclization. Kalata B8 also provides insight into the structure–activity relationships of cyclotides as it displays anti-HIV activity but lacks haemolytic activity. The “uncoupling” of these two activities has not previously been observed for the cyclotides and may be related to the unusual hydrophilic nature of the peptide.

Key words: anti-HIV activity, cyclic backbone, cystine knot motif, haemolytic, kalata B8, NMR.

INTRODUCTION

The cyclotides are an expanding family of small proteins that are characterized by three disulphide bonds forming a cystine knot motif and a head-to-tail cyclized backbone [1]. They have been isolated from plants in the Rubiaceae, Violaceae and Cucurbitaceae families and display a range of biological activities, including uteoretic, antimicrobial and cytotoxic activities [2–7]. Insecticidal activity against *Helicoverpa* caterpillar species suggests that cyclotides are present in plants as a defence mechanism [8,9]. The cyclotides are highly stable proteins with compact three-dimensional structures that have potential applications as drug design templates [10,11]. Furthermore, their intrinsic insecticidal activity may also be useful in agricultural applications.

Understanding how the cyclotides are produced *in vivo* is an important question in their biology that is at present unanswered. However, it is known that they are gene-encoded [9,12] and the discovery of genes for a range of cyclotides has provided preliminary insights into the factors involved in processing of the mature circular peptides. The genes encoding the cyclotides were originally isolated from *Oldenlandia affinis* and contain an ER (endoplasmic reticulum) signal sequence, a pro-region and a mature cyclotide domain [9]. In some clones, multiple mature cyclotide domains are present. All cyclotide domains are preceded by a short conserved pro-sequence referred to as the NTR (N-terminal repeat) conserved domain. Figure 1 shows a summary of the gene organization for two of the clones isolated from a cDNA library prepared from *O. affinis* leaves [9] as well as the sequence of the prototypic cyclotide kalata B1 [13,14]. The mature peptide contains six conserved cysteine residues that form a cystine knot. The backbone sequences between the cysteine residues are referred to as loops, as shown in Figure 1. Based on the gene sequences, it appears that processing to form the mature circular proteins occurs in the region that becomes loop 6 in the mature peptide.

The number of cyclotide sequences has been expanding recently, with more than 60 sequences now published [1]. Based on topological factors and sequence similarity, the cyclotides fall into two major subfamilies. This classification was originally proposed based on the presence or absence of a proline residue in loop 5, with those containing this proline referred to as M"obius cyclotides and those without referred to as bracelet cyclotides [15]. In two example cases, this proline has been shown to be in the *cis* conformation, introducing a conceptual twist in the otherwise all-*trans* backbone and leading to the name M"obius [8,16]. A third subfamily, which is referred to as the trypsin inhibitor cyclotides, currently contains only two members (MCoTI-I and MCoTI-II) that are quite different from the other cyclotides [17]. No sequence homology to either the M"obius or bracelet cyclotides is present in this subfamily, with the exception of the six conserved cysteine residues. The structure of MCoTI-II has been determined and reveals a similar cyclic cystine knot motif to other cyclotides. Interestingly, loop 6, putatively involved in the cyclization process, is disordered in the family of MCoTI-II structures [18,19], quite unlike the situation in all other cyclotides structurally characterized to date. The discovery of novel sequences in loop 6 may assist in finding clues to the mechanism of cyclization.

In the present study, we have characterized a novel cyclotide from *O. affinis*, kalata B8, which has sequence characteristics of both the M"obius and bracelet subfamilies, making it somewhat of a hybrid. Furthermore, kalata B8 displays disorder in loop 6 and isomerization of the cyclic peptide backbone similar to what is observed in MCoTI-II [17] but not in other cyclotides so far.

---

**Abbreviations used:** DOF-COSY, double-quantum-filtered COSY; EndoGluc, endoproteinase Glu C; ER, endoplasmic reticulum; LC-MS, liquid chromatography–MS; NOE, nuclear Overhauser effect; NTR, N-terminal repeat; RP, reverse phase.

---

1 To whom correspondence should be addressed (email d.craik@imb.uq.edu.au).

Co-ordinates of the reported protein structure have been deposited in the Protein Data Bank (PDB ID code 2B38).
chloride in 0.1 M Tris buffer (pH 8.5) with 10 µl of ethanol by incubating at 50°C for 4 h. The reduced peptide was alkylated with vinylpyridine in the dark at room temperature for 1 h and then centrifuged at 150,000 × g for 30 s in a microcentrifuge several times. The assay was performed by adding 20 µl of peptide solution to 80 µl of a 1% suspension of red blood cells in PBS. Test concentrations ranged from 0.5 to 3200 µl of peptide solution in 90% water/10% D2O at approx. pH 3.2. Spectra were recorded between 280 and 320 K on a Bruker ARX-500 spectrometer equipped with a shielded gradient unit. Two-dimensional NMR spectra were recorded in phase-sensitive mode using time-proportional phase incrementation for quadrature detection in the second dimension [20].

**The common theme of disorder, which presumably results from flexibility, and isomerization in the processing loops of both these peptides may be associated with the cyclization reaction and suggests a common processing mechanism.**

**EXPERIMENTAL**

**Isolation of kalata B8**

Kalata B8 was extracted from the above-ground parts of *O. affinis* with dichloromethane/methanol (1:1; v/v) overnight at room temperature (23°C). The extract was partitioned with water and the water/methanol layer was concentrated on a rotary evaporator prior to freeze-drying. The dried product was re-dissolved in dimethyl-2-silapentane-5-sulphonate.

**Sequence determination of kalata B8**

Kalata B8 (200 µl, 1 mg/ml) was reduced in 6 M guanidinium chloride in 0.1 M Tris buffer (pH 8.5) with 10 µl of 2-mercaptoethanol by incubating at 50°C for 4 h. The reduced peptide was alkylated with vinylpyridine in the dark at room temperature for 2 h prior to cleavage with EndoGluC (endoproteinase Glu C). Sequence determination was achieved using Edman degradation.

**NMR spectroscopy**

Samples for 1H NMR measurements contained approx. 1 mM peptide in 90% water/10% 2H2O (v/v) at approx. pH 3. 2H2O (99.9 and 99.99%) was obtained from Cambridge Isotope Laboratories (Woburn, MA, U.S.A.). Spectra were recorded between 280 and 320 K on a Bruker ARX-500 spectrometer equipped with a shielded gradient unit. Two-dimensional NMR spectra were recorded in phase-sensitive mode using time-proportional phase incrementation for quadrature detection in the second dimension [20]. The two-dimensional experiments consisted of a TOCSY [21] using an MLEV-17 spin lock sequence [22] with a mixing time of 80 ms, DQF-COSY (double-quantum-filtered COSY) [23], ECOSY (exclusive COSY) [24] and NOESY [25] with mixing times of 100–250 ms. Solvent suppression was achieved using a modified WATERSAGE (water suppression by gradient-tailored excitation) sequence [26]. Spectra were acquired over 6024 Hz with 4096 complex data points in F2 and 512 increments in the F1 dimension. 1H-Jcoupling constants were measured from a one-dimensional spectrum or from the DQF-COSY spectrum.

Spectra were processed on a Silicon Graphics Indigo workstation using XWINNMR (Bruker) software. The two dimension was zero-filled to 1024 real data points, and 90° phase-shifted sine bell window functions were applied prior to Fourier transformation. Chemical shifts were referenced to internal 2,2-dimethyl-2-silapentane-5-sulphonate.

**Structure calculations**

Preliminary structures of kalata B8 were calculated using a torsion angle simulated annealing method within the program DYANA [27]. Final structures were calculated using simulated annealing and energy minimization methods within CNS version 1.1 [28]. The starting structures were generated using random (ψ, φ) dihedral angles and energy-minimized to produce structures with the correct local geometry. A set of 50 structures was generated by a torsion angle simulated annealing method [29,30]. This method 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. Structures consistent with restraints were subjected to further molecular dynamics and energy minimization in a water shell, as described by Linge and Nilges [31]. 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. Secondly, 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 analysed using PROMOTIF [32] and PROCHECK-NMR [33].

**Haemolytic assay**

Human type A red blood cells were washed with PBS and centrifuged at 150 g for 30 s in a microcentrifuge several times until a clear supernatant was obtained. The assay was performed by adding 20 µl of peptide solution to 80 µl of a 1% suspension of red blood cells in PBS. Test concentrations ranged from 0.5 to 1400 µM. Synthetic melittin (Sigma) was used as a standard, as it is a well-established haemolytic agent. The mixtures were incubated at room temperature for 1 h and then centrifuged at 150 g for 1 min. The supernatant (35 µl) was diluted 1:30 in de-ionized (Milli-Q) water, and the absorbance was measured at

---

**Figure 1** Gene organization of cyclotides and the three-dimensional structure of the prototypic cyclotide kalata B1

A schematic representation of two of the genes from *O. affinis* is shown at the bottom of the Figure. The genes comprise a signal sequence labelled ER, a pro-region, a conserved repeated fragment labelled NTR and either one or multiple copies of the mature peptides, e.g. B1, B3 and B6. A short hydrophobic C-terminal tail is coded for at the end of the gene. The amino acid sequence of kalata B1 is shown with the cysteine residues labelled with roman numerals. The six inter-cysteine loops are labelled 1–6 on the structure in the top left.
RESULTS

Kalata B8 was isolated from the above-ground parts of *O. affinis* and purified using RP-HPLC. An LC-MS (liquid chromatography–MS) trace is shown in Figure 2 and highlights the hydrophilic nature of kalata B8 compared with the prototypic cyclotides kalata B1 and B2, as reflected in its shorter retention time. Indeed, kalata B8 is the earliest eluting of the major cyclotide peaks present in *O. affinis*. Kalata B8 was reduced with dithiothreitol, alkylated with vinylpyridine and sequenced using Edman degradation. The sequence is shown in Figure 2 and compared with selected Möbius and bracelet cyclotides. Kalata B8 appears to be a hybrid between the two major subfamilies of cyclotides. Specifically, loops 2 and 3 of kalata B8 resemble those found in kalata B1 and B2, members of the Möbius subfamily, but loop 5 is similar to the bracelet cyclotides, as exemplified by circulins A and B and lacks the characteristic cis-proline present in the Möbius cyclotides.

Close inspection of HPLC fractions and associated mass spectra of the partially purified extracts of *O. affinis* revealed two minor components eluting very close to kalata B8, one with the same mass (3282 Da) and one with a mass 18 Da lower. Incubation of the lower molecular mass form in 0.1 M ammonium bicarbonate buffer at pH 8 resulted in interconversion into kalata B8 and the isobaric form in an approx. 20:80 ratio, i.e. with the relative proportions of the two isobaric species reversed. Kalata B8 contains a single aspartate residue and the interconversion behaviour is consistent with the lower-molecular-mass form being a succinimide derivative and the two isobaric forms being α- and β-aspartic acid derivatives, as indicated in Figure 3. Conversion of succinimides into α- and β-aspartic acids generally favours the β-aspartic acid form [35,36], indicating that the minor form present in the *O. affinis* extract is the β-aspartic acid form. The major form in the extract was analysed using NMR spectroscopy and the assignments confirmed that it contains an α-aspartic acid.

NMR spectral assignments for kalata B8 were made using established techniques [37] and the 1H chemical shifts are supplied as Supplementary data (http://www.BiochemJ.org/bj/393/bj3930619add.htm). A comparison of the αH secondary shifts of kalata B8 and kalata B1 is shown in Figure 4 and illustrates the broad structural similarities between the molecules. However, the shifts also indicate that there are substantial local structural
differences, mainly associated with the loops where amino acid substitutions occur, in particular in loops 5 and 6. Figure 4 also shows that while kalata B8 has many of the same slowly exchanging amide protons as kalata B1, it is missing several slow-exchange amides in loop 6. This provides a first indication that loop 6 of kalata B8 is more solvent accessible and perhaps more flexible than that of kalata B1.

For kalata B8, the amide signals in the TOCSY and NOESY spectra for residues 1, 2 and 31 are weak and broadened in the temperature range analysed (280–320 K) and the amide signals of Glu1 are so broadened as to be visible only at temperatures > 312 K, as shown in Figure 5. This provides further evidence of a degree of flexibility in kalata B8 that is not seen in kalata B1 and other cyclotides structurally characterized so far. Significantly, all of the broadened residues are in or near loop 1, which forms a segment of the embedded ring in the structure that is an integral part of the cystine knot. Despite the broadening of these residues, chemical shifts in the amide region are well dispersed and the large number of resolved cross peaks in the NOESY spectrum allowed determination of a well-defined structure for the majority of the molecule.

The three-dimensional structure of kalata B8 was calculated with 203 distance restraints and 23 angle restraints using a simulated annealing method in CNS. Ten restraints for five hydrogen bonds were included, based on the slowly exchanging amide protons and preliminary structures (5HN-21O, 16HN-24O, 18HN-22O, 22HN-18O and 24HN-16O). No hydrogen bond acceptor was apparent in the preliminary structures for the slowly exchanging Cys15 amide proton. The resulting family of structures accepted annealing method in CNS. Ten restraints for five hydrogen bonds were included, based on the slowly exchanging amide protons and preliminary structures (5HN-21O, 16HN-24O, 18HN-22O, 22HN-18O and 24HN-16O). No hydrogen bond acceptor was apparent in the preliminary structures for the slowly exchanging Cys15 amide proton. The resulting family of structures, the secondary shifts were calculated by subtracting the random coil shifts \([54]\) from the experimental \(\alpha H\) chemical shift. The slowly exchanging amide protons for kalata B8 are highlighted with open circles on the sequence and those for kalata B1 with filled circles. The putative ligation site for cyclization of kalata B8 is highlighted with an asterisk and an arrow indicating the direction of the peptide chain from N- to C-terminus. The boxed regions show the sequence differences between kalata B1 (lower part of boxes) and B8.
> 11 µM. Tricyclon A did not display anti-HIV or cytotoxic activity. In contrast, other cyclotides have much more potent anti-HIV activities. For example, kalata B1 and circulins A–F have EC₅₀ values of approx. 140 nM [42] and 40–275 nM [4,43] respectively.

**DISCUSSION**

In the present study, we have determined the sequence of a novel circular protein (kalata B8) belonging to the family of plant cyclotides. The cyclic nature of the backbone was confirmed by MS analysis following EndoGluC cleavage and the sequence was determined using Edman sequencing. A significant difference between kalata B8 and most of the published cyclotide sequences is that it does not fall into either of the two subfamilies commonly found in cyclotides, namely Möbius or bracelet. Instead, it appears to be a hybrid between the two subfamilies. In addition to its hybrid characteristics, kalata B8 is the most hydrophilic of any of the cyclotides expressed by *O. affinis* and therefore is an interesting target for structural and bioactivity studies.

The discovery of kalata B8 suggests that the sequence homologies within the Möbius and bracelet subfamilies may not be as distinct as previously thought, and with the discovery of more cyclotides the boundaries between the two subfamilies may merge further. Although the cyclotide family is rapidly expanding, there still remains a significant degree of sequence similarity amongst known members. This is in contrast with other disulphide-rich peptide families such as the conotoxins, which display significant sequence variation and multiple disulphide bond frameworks [44,45]. In the case of the conotoxins, which are used for prey capture by marine snails, it is thought that the rich diversity in structures in the cocktail of venom peptides is important for the targeting of multiple receptor sites simultaneously, to result in the rapid immobilization of prey. The rationale for a suite of broadly similar host defence peptides such as the cyclotides in a
single plant is less clear, but the existence of such suites is clear. In a recent report, at least 57 new cyclotide masses were seen in the single plant *Viola hederacea*, for example [46].

The structure of kalata B8 contains the cyclic cystine knot motif characteristic of cyclotides but has some interesting differences, including disordered loops 1 and 6, most probably as a result of conformational flexibility that is manifest in the broadening of NMR signals from residues in these regions. It is apparent from the present study that the dynamics of cyclotides can vary significantly despite a high degree of sequence conservation. For example, kalata B1 and B8 contain 17 identical residues, including an identical loop 1, and their overall structures are very similar. However, without determining their structures experimentally, the differences in flexibility would not have been apparent. Clearly, only subtle differences in sequence are sufficient to result in different dynamic properties.

The apparent flexibility observed in kalata B8 is particularly interesting as it occurs in a loop directly involved in the cystine knot motif, previously shown to be critical in maintaining the structural stability of the cyclotides [47]. Such disorder in the cystine knot motif is unprecedented in the cyclotides that have been structurally characterized so far. The disorder in loop 1 of kalata B8 is associated with broadening of several amide protons, most notably from Glu3. In all cyclotides structurally characterized so far, the side-chain carboxyl of this absolutely conserved residue hydrogen-bonds to the backbone amides of the first two residues in loop 3, thus mutually stabilizing loops 1 and 3. The equivalent residues in loop 3 for kalata B8 are Tyr11 and Thr12. The chemical shifts of these residues are influenced by pH as occurs for other cyclotides but Tyr11 is broadened beyond detection at high pH, indicating conformational exchange. Thus it appears that the interaction between loops 1 and 3 is reduced in kalata B8 compared with other cyclotides and the Glu residue may be in conformational exchange between the hydrogen-bonded conformation and one or more other conformations.

More generally, the cystine knot motif is present in a wide range of peptides and proteins, including many toxins and in general is a very stable motif [38,39], making the apparent flexibility observed in kalata B8 a surprising result. Disorder in cystine knot-containing peptides has previously been observed in only a few cases, for example in the conotoxin MrVIB [48], but this disorder was restricted to one large loop and did not involve the loops directly involved in the cystine knot. Disorder in loop 6 has also been observed for the trypsin inhibitor cyclotide MCoTI-II [18,19]. In that case, several isoforms, involving isomerization of the backbone at an Asp-Gly sequence in loop 6, are present. Kalata B8 also contains Asp-Gly in loop 6 and this sequence is well known to be susceptible to backbone isomerization [35,36]. In the case of kalata B8, the native extract contained small amounts of a succinimide form and a β-aspartate form in addition to the predominant α-aspartate isomer. The presence of these forms raises the question of whether they are degradation products of the mature peptide, or possibly by-products from the cyclization process that may provide clues as to its mechanism.

In support of the latter, the Asp-Gly sequence associated with the isomerization in kalata B8 is almost certainly located at the putative cyclization site, based on the gene sequences of a range of cyclotides [9,12]. In nearly all cyclotides sequenced so far, either an Asp-Gly sequence or, more commonly, an Asn-Gly sequence is present in loop 6 and it appears that this sequence is directly involved in the cyclization mechanism, with the Asn/Asp residue deriving from the C-terminus of the excised precursor fragment and the Gly from the N-terminus. Figure 6 shows the cyclization loops from a range of naturally occurring circular peptides and highlights the location of the processing site in kalata B8 by analogy with those in kalata B1 and tricyclon A, where the precursor sequences are known.

Another example of a naturally occurring circular peptide containing an Asp-Gly sequence is the 14-amino acid sunflower trypsin inhibitor, SFTI-1 [49]. The recently discovered precursor protein for this peptide suggests that in vivo cyclization involves ligation to form an Asp-Gly sequence [50], as indicated in Figure 6. It has also been shown that, in vitro, an acyclic analogue of SFTI-1 is susceptible to cyclization and isomerization of the Asp-Gly sequence, confirming the high reactivity of this sequence [51]. Also shown in Figure 6 is loop 6 of MCoTI-II. Although the gene sequence of MCoTI-II has not been determined, loop 6 of the mature peptide (SGSDGGV) shares high homology with the corresponding part of the precursor (SGRHGGI) of a related acyclic trypsin inhibitor TGTI-II [52]. However, in this case, even though the loop contains an Asp-Gly sequence, alignment of the TGTI-II and MCoTI-II sequences suggests that cyclization may instead occur at a Gly-Ser sequence. In this case, the observation of both α- and β-aspartic acid linkages as well as a succinimide derivative in plant extracts is probably an indicator of post-cyclization degradation of the Asp-Gly bond rather than direct involvement in the cyclization mechanism. The flexible nature of loop 6 may facilitate this degradation.

The sizes of the processing loops for the various naturally occurring circular peptides shown in Figure 6 vary from 5 to 10 residues and the loops vary substantially in sequence. By definition, the loops are flanked by Cys residues at either end and it is interesting to note that the processing point in most cases appears to be three residues from the upstream Cys. If a common mechanism of cyclization is involved, then formation of a C-terminal succinimide intermediate via a side chain to backbone

![Figure 6 Processing loop sequences for plant-derived circular proteins](image)
The lowest-energy structure of kalata B8 is shown in Figure 7 and compared with kalata B1 and tricyclon A. The hydrophobic patch in kalata B1 is not as continuous in tricyclon A and this has been suggested to account for the lack of haemolytic activity in tricyclon A [41]. In kalata B8, the degree of surface-exposed hydrophobic residues is even less pronounced than tricyclon A, consistent with the lack of haemolytic activity for kalata B8. Further studies are required to determine the features important for anti-HIV activity.

In summary, we have characterized a novel cyclotide from O. affinis that has provided insights into structure–activity relationships and the mechanism of cyclization. Kalata B8 displays anti-HIV activity but not haemolytic activity, indicating that the mechanism by which cyclotides act is dependent on the composition of the membrane and is influenced by the hydrophilicity of the peptide. The structure of kalata B8 is disordered in loops 1 and 6, most likely as a result of flexibility, indicating that very similar sequences and structural motifs can display vastly different dynamic properties.

N. L. D. is an NHMRC (National Health and Medical Research Council) Industry Fellow. D. J. C. is an ARC (Australian Research Council) Professorial Fellow. We thank Lillian Sando for performing the haemolytic assays, Kirk Gustafson for anti-HIV assays and Elizabeth Westbury for assistance with NMR analysis.

REFERENCES
