THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 278, No. 48, Issue of November 28, pp. 48188–48196, 2003

Printed in U.S.A.

Disulfide Mapping of the Cyclotide Kalata B1

CHEMICAL PROOF OF THE CYCLIC CYSTINE KNOT MOTIF*

Received for publication, August 8, 2003, and in revised form, September 5, 2003
Published, JBC Papers in Press, September 5, 2003, DOI 10.1074/jbc.M308771200

Ulf Göransson‡§¶ and David J. Craik‡‖

From the ‡Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia and §Department of Medicinal Chemistry, Division of Pharmacognosy, Biomedical Centre, Uppsala University, Box 574, Uppsala S-751 23, Sweden

The cyclotides are a recently discovered family of plant proteins that have the fascinating structural feature of a continuous cyclic backbone and, putatively, a knotted arrangement of their three conserved disulfide bonds. We here show definite chemical proof of the I-IV, II-V, III-VI knotted disulfide connectivity of the prototypic cyclotide kalata B1. This has been achieved by a new approach for disulfide analysis, involving partial reduction and stepwise alkylation including introduction of charges and enzymatic cleavage sites by aminoethylation of cysteines. The approach overcomes the intrinsic difficulties for disulfide mapping of cyclotides, *i.e.* the cyclic amide backbone, lack of cleavage sites between cysteines, and a low or clustered content of basic amino acids, and allowed a direct determination of the disulfide bonds in kalata B1 using analysis by mass spectrometry. The established disulfide connectivity is unequivocally shown to be cystine knotted by a topological analysis. This is the first direct chemical determination of disulfides in native cyclotides and unambiguously confirms the unique cyclic cystine knot motif.

The cystine knot is now a well established motif seen in a wide variety of proteins from organisms as diverse as fungi, plants, insects, and animals. It comprises an embedded ring formed by two disulfide bonds and their connecting backbone segments that is threaded by a third disulfide bond (1, 2). The cyclic cystine knot (CCK)[1] motif contains an additional level of topological complexity in that it is embedded in proteins containing a head-to-tail cyclized backbone (3). So far the CCK motif is known only to occur in certain plant proteins known as the cyclotides (4). These are typically about 30 amino acids in size, contain six conserved Cys residues involved in the CCK motif, and have a diverse spectrum of biological actions ranging from uterotonic action (5, 6), cytotoxic activity (7), antimicrobial (8), inhibition of trypsin (9, 10), and human immunodeficiency virus-inhibitory (11, 12) to inhibition of neurotensin binding (13).

The sequences of ~50 cyclotides have so far been reported (4, 6, 10–20), and many more are currently being characterized. They represent the largest known family of circular proteins. Their natural function in plants seems to be defense molecules as they show potent activity against *Helicoverpa* insect pests (20). They also have activity against a range of bacteria and fungi (8).

It is believed that the exceptional stability of the cyclotides (which retain biological activity after boiling) is because of the strongly braced structure built around the CCK motif. Their exceptional stability has led to suggestions that they may make a valuable template in peptide-based drug design applications (21). Further, their potent activity against insect pests (20) makes them potentially valuable crop protection agents if their genes can be transferred into crops subject to predation by susceptible pest species. For these reasons and because of their unique and fascinating topological properties there is much interest in the structural characterization of the cyclotide molecular framework.

Determination of the disulfide connectivity and topology of the cyclotides is intrinsically very difficult, because the close packing of the six Cys residues (I-VI) and the resistance to enzyme digestion have made it impossible to obtain suitable digestion fragments for classical disulfide mapping. In the original structural studies (22) we used a geometric and energetic analysis of the calculated NMR models for each of the 15 possible disulfide connectivities to deduce the preferred connectivity for kalata B1. This led to the proposed I-IV, II-V, III-VI connectivity, although two other low energy alternatives could not be completely eliminated. A subsequent analysis by mass spectroscopy of fragments obtained after partial acid hydrolysis of circulins A and B supported the I-IV, II-V, III-VI connectivity (23). Further, a synthetic study of the circulins and cyclopsychotride A in which one disulfide bond was unequivocally formed via directed chemical synthesis supported the previously proposed connectivity (8). These cyclotides, however, represent only one of the two major subfamilies of cyclotides identified so far, *i.e.* the "bracelet" one, whereas kalata B1 belongs to the other, the Möbius subfamily. Fig. 1 (*panels A* and *B*) shows sequence homology diagrams for the two subfamilies based on the published sequences of ~50 cyclotides.

Recently Skjeldal *et al.* (24) recorded new NMR data for kalata B1 at 750 MHz and proposed an alternative disulfide connectivity based in part on the existence of certain NOEs between Cys residues presumed to be connected and in part on an energetic analysis of all 15 possible disulfide isomers as was done in the original structural study (22). It was suggested that the differences in sequence between the two subfamilies might

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ Supported by a fellowship from the IF Foundation, Swedish Academy of Pharmaceutical Sciences.

‖ Australian Research Council Senior Fellow. To whom correspondence should be addressed: Inst. for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia. Tel.: 61-7-3346-2019; Fax: 61-7-3346-2029; E-mail: d.craik@imb.uq.edu.au.

[1] The abbreviations used are: CCK, cyclic cystine knot; 1S, disulfide species with one disulfide bond; 2S, disulfide species with two disulfide bonds; Ae, aminoethyl; des(CysII-CysVI), disulfide species with this particular disulfide reduced; HPLC, high performance liquid chromatography; MS, mass spectrometry; NEM, *N*-ethylmaleimide; RP, reverse phase; TCEP, tris(2-carboxyethyl)phosphine; NOE, nuclear Overhauser effect; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; MS/MS, tandem MS.

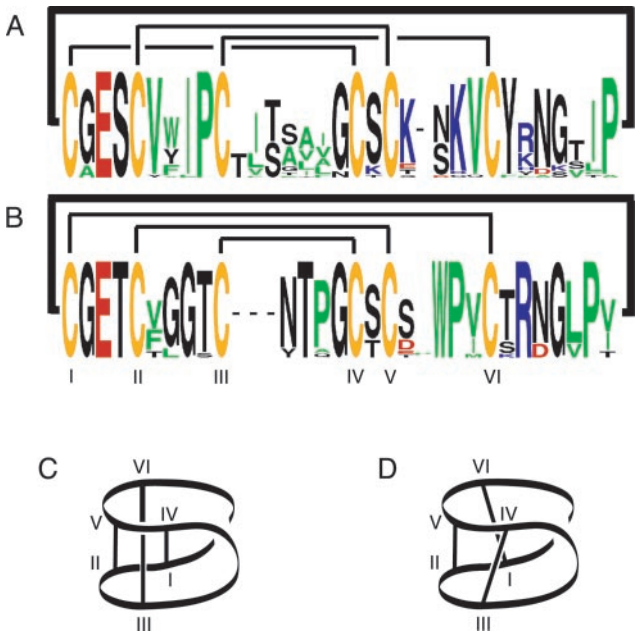

FIG. 1. **Cyclotide sequences and structures.** *Panels A* and *B* show the sequences of the bracelet and the Möbius subfamily as summarized in sequence logos (44). These display the degree of sequence conservation as the total height of the stack of *letters* and the frequency of individual amino acids as the relative height of *letters* at each position. The Möbius cyclotides, named after the topological twist caused by a *cis*-Trp-Pro bond in the loop between cysteines V and VI, were suggested recently (24) to contain a I-VI, II-V, III-IV cystine pattern as illustrated by the *brackets* in *B*, instead of the I-IV, II-V, III-VI as proposed originally (22) and shown in *A*. *C* and *D* schematically show the topological implications of the two proposed disulfide connectivities, resulting in a cystine knotted arrangement for the I-IV, II-V, III-VI connectivity (*C*) and a laddered arrangement for the I-VI, II-V, III-IV connectivity (*D*), respectively. The logos in *A* and *B* were created from an alignment of cyclotides reported in Refs. 4, 6, and 10–20; the conserved cysteines are marked in *yellow*, hydrophobic residues in *green*, acidic in *red*, and basic in *blue*.

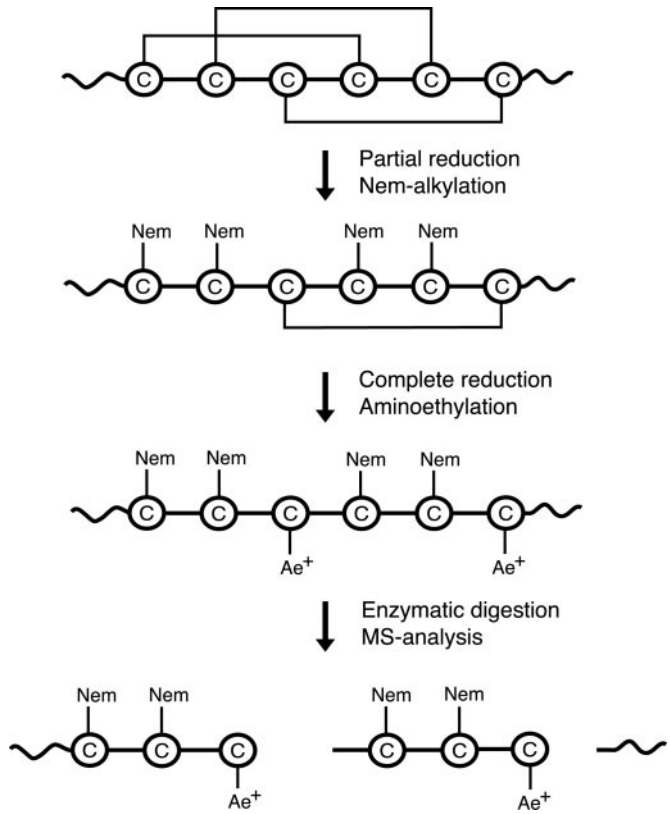

FIG. 2. **Schematic presentation of the strategy used in this study for disulfide determination.** Partial reduction and a first alkylation with NEM are performed at pH 3 to avoid reshuffling of native disulfides. The remaining disulfides are then reduced, and positive charges and enzymatic cleavage sites are introduced simultaneously at these cysteine residues by aminoethylation (Ae^+). Determination of molecular weights by MS identifies cleaved fragments and thus those cysteines that are involved in the disulfide bond.

be responsible for this alternative connectivity, which, as illustrated in Fig. 1 (*panels C* and *D*), would imply a laddered rather than a knotted topology.

We have noted recently (25) that local Cys-Cys NOEs can be very misleading in a molecule such as kalata B1 as often pairs of protons in non-connected Cys residues can be closer in space than those paired within a disulfide bond. However, an analysis of coupling constants between the αH and βH protons of the Cys residues, combined with NOE analysis, strongly supported our original proposition that there is a knotted topology in kalata B1 and cycloviolacin O1. These molecules were chosen for study as they represent prototypic examples of cyclotides from both of the two major subfamilies, *i.e.* the bracelet and Möbius forms (4).

Although we believe that our recent NMR structural studies of kalata B1 and cycloviolacin O1 (25) provided definitive evidence for the disulfide connectivity as originally proposed, it remains the case that so far there have been no direct chemical determination of the disulfide connectivities in the Möbius cyclotides. To address this question, we have developed an MS-based approach for disulfide mapping that overcomes the notorious difficulties associated with disulfide analysis of small cystine-rich peptides in general and for peptides with a low or clustered content of positively charged residues, such as the cyclotides, in particular. As illustrated in Fig. 2, the method involves the introduction of charges, as well as cleavage sites, directly at Cys residues involved in disulfides, which then can be determined easily by mass spectrometric analyses.

The results show that partial reduction and selective labeling via alkylation of individual disulfide bonds allows an unambiguous determination of the disulfide connectivity and unequivocally confirms the presence of a cystine knot in the prototypic cyclotide kalata B1. Further, we have undertaken a detailed topological analysis of disulfide bonding networks in tridisulfide proteins and show that the presence of the I-IV, II-V, III-VI connectivity is sufficient to result in a cystine knot topology provided the backbone is cyclic.

MATERIALS AND METHODS

General—Kalata B1 was isolated and purified as described (4) from the fresh aerial parts of *Oldenlandia affinis* DC (Rubiaceae). An Agilent 1100 HPLC (Agilent, Palo Alto, CA) and a Grom-Sil ODS-4 HE (125 × 2 (inner diameter) mm, 3 μm, 200 Å) column were used for RP-HPLC, eluted with linear gradients from A (0.05% trifluoroacetic acid in water (v/v)) to B (90% AcN in 0.045% trifluoroacetic acid) at a flow rate of 0.3 ml/min. Monoisotopic masses (M) are used throughout the text.

Partial Reduction and Stepwise Alkylation—Partial reduction was done as described (26, 27). In short, peptides (typically ~0.5 mg in 0.5 ml of 0.05% trifluoroacetic acid and ~30% AcN as eluted from HPLC) were incubated with an equivalent volume of 20 mM TCEP, 0.2 M citrate buffer, pH 3.0, at 64 °C for 3 min. The peptide solution and the reducing buffer were pre-equilibrated at the same temperature before mixing. The reaction was stopped by immediate injection on a Phenomenex Jupiter C18 column (250 × 4.6 (inner diameter) mm, 5 μm, 300 Å), eluted with a shallow gradient at a flow rate of 1 ml/min (isocratic at 20% B for 10 min; 20–34% B in 3 min; 34–40% B in 24 min; 40–100% B in 1 min; 100% for 2 min; and finally re-equilibration at 20% B for 11 min, giving a total time of 51 min). Partially reduced species were collected manually.

For NEM alkylation, the same volume of 60 mM NEM in 0.2 M citrate buffer, pH 3, was added directly to collected fractions containing (partially) reduced peptides. Typically, such a fraction contained ~1 nmol peptide in 100 μl of solvent; the addition of 100 μl of NEM (6 μmol) then

corresponds to 1000 equivalents of the maximum number of alkylation sites, that is the six cysteines (6 nmol). This mixture was incubated for 60 min at room temperature before direct injection and separation on RP-HPLC. Collected peptides were then dried on using a Speedvac (Savant Instruments, Holbrook, NY).

For complete reduction and aminoethylation NEM-alkylated peptides were dissolved and fully reduced in 100 μl of 40 mM dithiothreitol in 0.5 ml of 0.25 M Tris-HCl, pH 8.5, containing 1 mM EDTA and 6 M guanidine HCl. After 1 h of incubation (37 °C) 2 mg of 2-bromoethylamine dissolved in 10 μl of the same Tris buffer was added, after which the reaction was incubated overnight (37 °C). The reaction was then terminated by injection on RP-HPLC.

Enzymatic Cleavage—Collected fractions were dried on a Speedvac and cleaved with modified trypsin (0.2 μg in 50 μl of 50 mM NH_4HCO_3, pH 7.8, 37 °C, 4 h; Promega, Madison, WI). The cleavage buffer was removed by lyophilization before MS analyses.

Mass Spectrometry—To monitor the reactions in the stepwise alkylation scheme MALDI-TOF MS analyses were done on a Perseptive Biosystems Voyager DE operated in the positive reflectron mode using α-cyano-4-hydroxy cinnamic acid as matrix. Nanospray MS and MS/MS analyses were done in the positive ion mode on a Qstar Pulsar (Applied Biosystems, Foster City, CA) equipped with a Protana nanospray ion source (Protana, Odense, Denmark). Typically, a <1-nmol sample was dissolved in 10 μl of 60% aqueous methanol containing 0.1% formic acid, of which ~1 μl was transferred to the nanospray needle. Molecular weight determinations were then done using full scan mode *m/z* 200–1000 using a potential of 900 V applied to the nanospray needle. MS/MS studies were done on the same sample and potential as above, with a collision energy of between 25 and 45 eV.

RESULTS

The strategy to definitively determine the disulfide connectivity of kalata B1 involved first determining conditions under which only one or two of the disulfide bonds could be reduced, labeling the reduced Cys residues using an alkylating reagent and then reducing and labeling the remaining Cys residues for subsequent MS analysis. The choice of reducing agent and alkylators was a key feature in this strategy; rearrangement of disulfide bonds was avoided by performing partial reduction and the first alkylation at acidic pH, whereas the second labeling was used to introduce site-specific enzymatic cleavage labels at the remaining disulfides. This method allowed an unambiguous determination of the I-IV, II-V, III-VI disulfide pattern, thus providing definite chemical proof of the disulfide bond connectivity that characterizes the CCK motif.

Initially, the cyanylation strategy described by Wu and coworkers (28–30) was attempted, but we found a number of problems associated with the chemistry of the reactions that had undesired effects when applied to cyclotides. In this method the amino acid on the N-terminal side of the cyanylated Cys affects the yield of the chemical cleavage reaction, with bulky or rigid side chains acting as inhibitors, resulting in incomplete cleavage. Moreover, cleavage sites are also lost by β-elimination of the cyanylated cysteines, which further complicates the analysis (29). When this method was applied to kalata B1 interpretation of the MS spectra was greatly complicated as the spectra were dominated by fragments containing the only positively charged residue in the sequence (kalata B1 contains one Arg residue).

Partial Reduction by TCEP and Isolation of Intermediates—Partial reduction of kalata B1 was done with TCEP in 0.1 M citrate buffer at pH 3.0 and was optimized with respect to incubation time to give the highest ratio of partially reduced disulfide species relative to fully reduced peptide. The optimal time for sampling was determined to be 3 min; the consumption of native material was then kept to a minimum, and a rapid turnover time for the total reaction-separation procedure was achieved. Two main intermediates were collected and analyzed by MALDI-TOF MS. Their molecular weights were determined to be 2894 and 2892, corresponding to one species containing one remaining disulfide bond (referred to as a 1S species) and one species that contains two disulfide bonds (referred to as a 2S species).

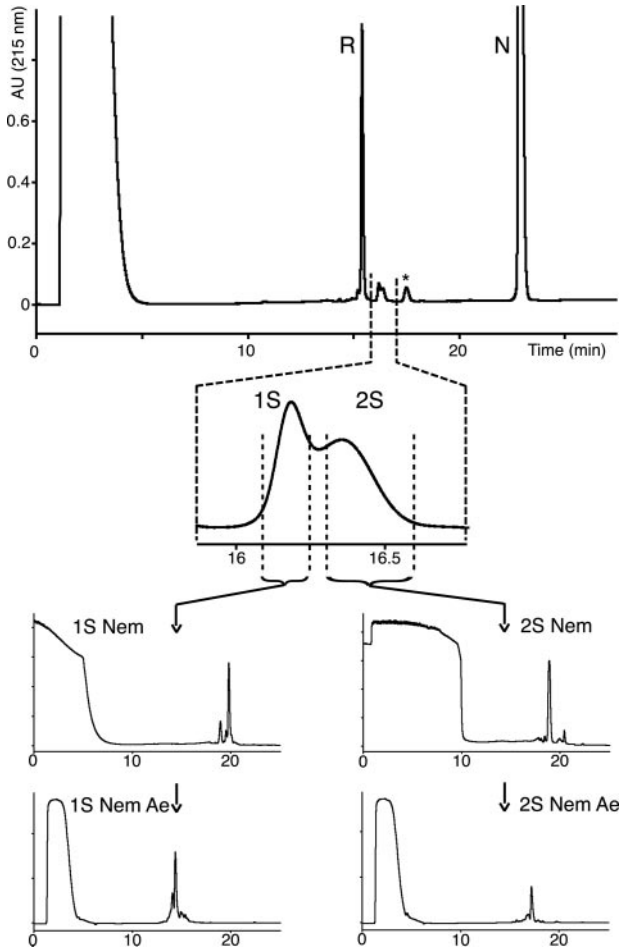

FIG. 3. **Partial reduction and stepwise alkylation of kalata B1 monitored by RP-HPLC.** At the *top*, a typical RP-HPLC trace after partial reduction of kalata B1 is shown. The 1S and 2S species elute near the fully reduced peptide (*R*), which dominates the trace, together with the native peptide (*N*). The *inset* shows the chromatographic cuts for the collection of the 1S and 2S species, followed by the respective isolation of NEM and NEM/Ae-alkylated derivatives. Note the changes in retention time caused by the overall change in hydrophobicity because of the introduction of the different alkylators. The *peak* marked with an asterisk (*) in the *top* chromatogram was also present in the blank run. All chromatograms shown were obtained using a Grom-Sil ODS-4 HE column.

As shown in Fig. 3, the reduced peptide(s) eluted before the native one on RP chromatography. This behavior is opposite to what is normally observed for most peptides and proteins but for cyclotides, and for some other disulfide-rich small peptides (31), is explained by the fact that the interior of the molecule is occupied by the disulfides, thus forcing hydrophobic amino acids to be surface-exposed (32). Removal of the disulfide bonds reduces the conformational constraints and favors burial of the exposed hydrophobic residues.

Stepwise Alkylation of Partially Reduced Disulfide Species—The first alkylation was carried out immediately after collection of the 1S and 2S species by adding NEM (60 mM in 0.2 M citrate buffer, pH 3.0) in a volume equivalent to the collected fraction. The pH of the fractions was measured and confirmed to be 3.0 at which the alkylation reaction was carried out for 1 h. NEM-alkylated peptides were then re-isolated by RP-HPLC, and their molecular weights were determined to 3396 and 3142 by MALDI-TOF MS, corresponding to reduction and NEM alkylation of four cysteines for the 1S species and two cysteines for the 2S species.

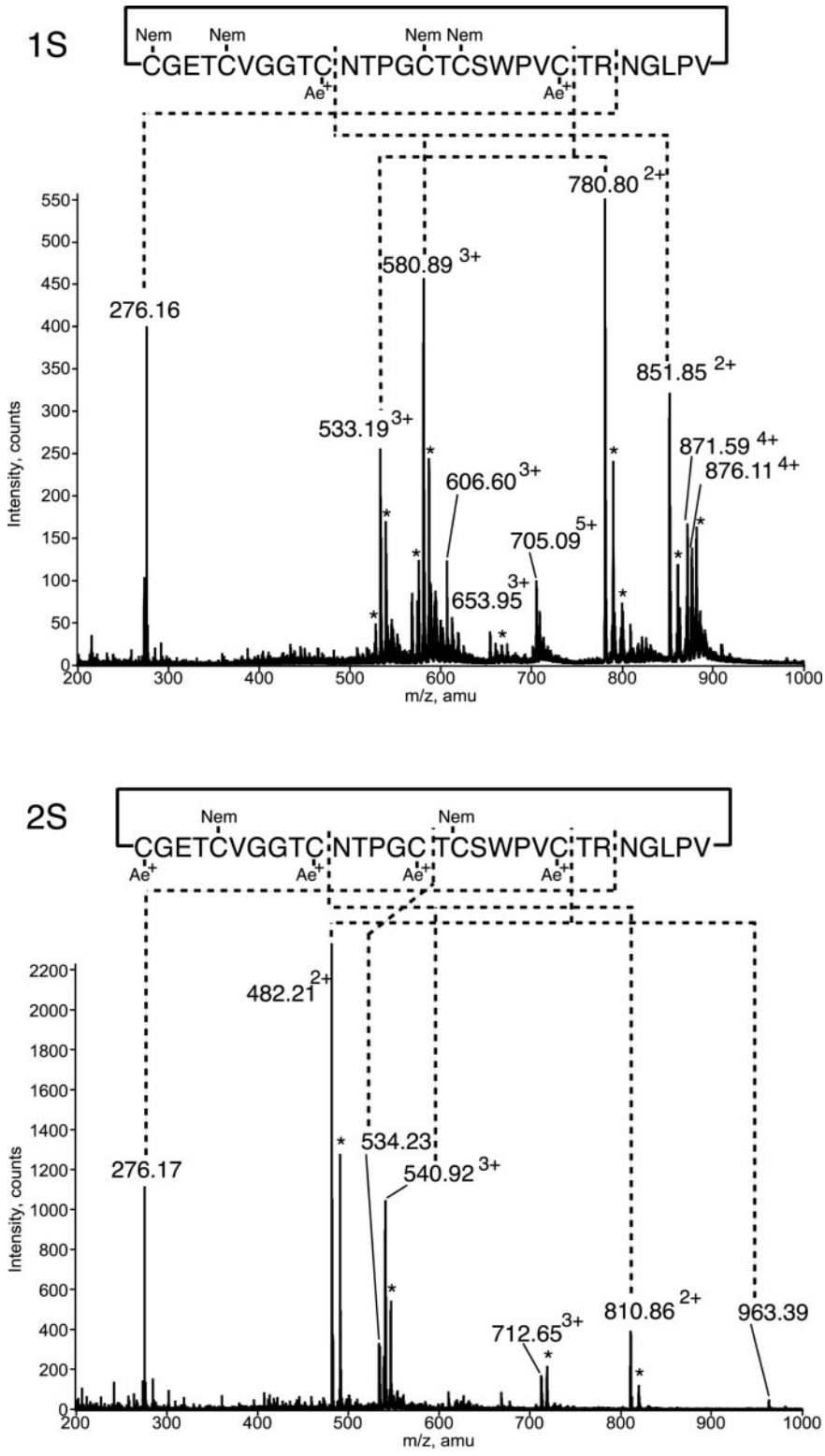

FIG. 4. **MS analyses after tryptic digestion of the 1S and 2S species.** Major ions that directly identify the site of the different alkylators (*NEM/Ae*$^+$) are assigned as shown by the sequences in this figure and in Table I. For the 1S species (*upper spectrum*), ions corresponding to three main fragments dominate the spectra as can be expected with the introduction of two extra cleavage sites (2×Cys(Ae)); the singly charged ion at *m/z* 276.16, and the doubly charged ions at 780.80 and 851.85 including their respective +18 NEM adducts (marked with an asterisk (*)). The latter two also show a cluster of ions originating from triply charged ions, with their respective $[M+2H+K]^{3+}$ ions as the most intense ones (*m/z* 533.19 and 580.89). In addition, ions corresponding to one missed cleavage site are seen (606.60^{3+} and 653.95^{3+}; see assignments in Table I), as well as the modified but non-cleaved peptide (871.59^{4+}) and corresponding adducts (876.11^{4+}, +18 because of either an NEM adduct or a single cleavage; 705.09^{5+}, K^+). All of these fragments are consistent with the assignment of NEM/Ae as deduced from the three main fragments. For the 2S species the *m/z* 276.17, 534.23, and 482.21^{2+}/963.39 *peaks* correspond to three of the five fragments expected for the introduction of Ae on four Cys residues. The two remaining were detected as one fragment at *m/z* 540.92^{3+}/810.86^{2+}. The assignment of the missed cleavage site (Cys(Ae)) and the NEM-alkylated Cys residue in this fragment was done by MS/MS as shown in Fig. 5.

The alkylated peptides were then lyophilized and redissolved at a pH of 8.5, at which the remaining disulfides were reduced with dithiothreitol and alkylated with 2-bromoethylamine. After isolation on RP-HPLC, the molecular weights of these aminoethylated peptides were determined to be 3482 and 3318 (by MALDI-TOF MS), congruent with aminoethylation (Ae) of the two remaining cysteines of the 1S species and the four remaining cysteines of the 2S species.

No reshuffling of alkylators between non-labeled/labeled or differently labeled cysteines was observed under the experimental conditions used. The latter was confirmed by subjecting fully NEM-alkylated peptide to attempted further reduction and aminoethylation and the identical molecular weights and retention times (RP-HPLC) that were observed for this peptide before and after the experiment.

It is interesting to note the changes in retention times during the stepwise alkylation as shown in Fig. 3. Although the partially reduced species eluted just after the fully reduced peptide, retention times were increased after NEM alkylation and decreased again after aminoethylation. This is consistent with

TABLE I
Fragments after tryptic cleavage for assignment of disulfide connectives

Disulfide species	Ions[a,b]	$Mw_{Exp.}$[a,b]	$Mw_{Calc.}$[a]	Identified fragments
	m/z			
1S	276.16	275.16	275.16	TR
	780.80^{2+} (789.80^{2+})	1559.60 (1577.60)	1559.62	NTPGC(NEM)TC(NEM)SWPVC(Ae)
	851.85^{2+} (860.85^{2+})	1701.70 (1719.70)	1701.72	NGLPVC(NEM)GETC(NEM)VGGTC(Ae)
	606.60^{3+}	1816.80	1816.77	NTPGC(NEM)TC(NEM)SWPVC(Ae)TR
	653.95^{3+}	1958.85	1958.86	TRNGLPVC(NEM)GETC(NEM)VGGTC(Ae)
2S	276.17	275.17	275.16	TR
	534.23	533.23	533.23	NTPGC(Ae)
	482.21^{2+}, 963.39 (491.21^{2+})	962.39 (980.42)	962.40	TC(NEM)SWPVC(Ae)
	540.92^{3+}, 810.86^{2+} (546.58^{3+}, 819.86^{2+})	1619.72 (1637.72)	1619.71	NGLPVC(Ae)GETC(NEM)VGGTC(Ae)
	534.75^{4+}, 712.65^{3+} (539.24^{4+}, 718.66^{3+})	2134.95 (2152.98)	2134.93	NGLPVC(Ae)GETC(NEM)VGGTC(Ae)NTPGC(Ae)

[a] Monoisotopic masses are used, and $Mw_{Exp.}$ are calculated through deconvolution of the doubly or triply charged ions (Qstar MS) and given as (M).
[b] Masses in parentheses contains one Cys(NEM) adduct (+18 atomic mass units).

the expected increase in hydrophobicity by NEM-alkylated cysteine side chains and the introduction of the hydrophilic, positively charged aminoethyl groups.

Identification of Fragments and Disulfide Connectivities after Enzymatic Cleavage—Fig. 4 shows the MS spectra for the NEM- and Ae-alkylated 1S and 2S species after tryptic digestion. The ions observed in the spectra, obtained by nanospray-MS, all correspond to peptide fragments congruent with the introduction of two or four introduced cleavage sites by aminoethylation of the 1S or 2S species, respectively. As shown in Fig. 4 and Table I, the molecular weights of the fragments alone identify the sites of the two different alkylators for each of the two disulfide species. Thus the two aminoethylated cysteines in the 1S species revealed the CysIII-CysVI bond. For the 2S species the sites of NEM-alkylated residues directly established its identity as des(CysII-CysV). The identification of all fragments and the sites of the different alkylators were confirmed by MS/MS peptide sequencing, as shown by the example in Fig. 5.

The NEM alkylation results demonstrated the advantage of this reagent as an efficient alkylator at low pH but showed some initially puzzling results. First NEM alkylation introduces a chiral center on the modified cysteine side chain, resulting in 2^n stereoisomers for n alkylated cysteines and, possibly, the same number of peaks on RP-HPLC. No indication of the latter could be seen until after tryptic cleavage when the peaks became apparent, especially for the 1S fragments after extended digestion (*i.e.* 20-h incubation). These fragments each carried two NEM-modified cysteines and thus showed peaks split in four on RP-HPLC. However, these isomers all have the same mass, which does not affect analysis by MS alone (*i.e.* by infusion or nanospray as used here). Second, mass adducts of +18 and +36 were observed after digestion for fragments containing one or two NEM-alkylated cysteine/cysteines, respectively. These adducts were also shown by MS/MS fragmentation to originate from the NEM-modified cysteines as shown in Fig. 5. Most probably, they result from the uptake of water by hydrolysis (+18) of the maleimide ring structure. This is a known reaction for this type of compounds; maleimides have been shown to undergo hydrolysis at slightly basic pH (>7.5) (33). Such a reaction would also be favored at the conditions we normally use for tryptic digestion (overnight incubation in 50 mM NH_4HCO_3, pH 7.8). With this in mind it was possible to use the formation of adducts to our advantage in the identification of NEM-containing fragments. As shown in Fig. 6, the combined information on the two disulfide bonds identified from the isolated 1S and 2S species, CysIII-CysVI and CysII-CysV, respectively, gives the third and remaining one by deduction: CysI-CysIV.

DISCUSSION

In the current study we have selectively reduced one or two of the native disulfide bonds in kalata B1 and then alkylated the reduced Cys residues to specifically tag them. Using mass spectrometric methods we have unequivocally determined the connectivity of disulfide bonds and demonstrated that they have a I-IV, II-V, III-VI pattern, as originally proposed by Saether *et al.* (22). When this information is combined with NMR data on internal distances and dihedral angles (25) it is clear that the disulfide bonds form a tightly intertwined network that defines the CCK motif.

The cyclic backbone and the extremely compact structure of the cyclotides are complicating factors for chemical mapping of disulfides. In addition the cyclotides have an extremely low content of suitable cleavage sites between the cysteines, and they show a high resistance to enzymes. Together these facts make existing methods for determination of disulfide connectivities inadequate, and to address this problem a new method was developed. This approach is a development of the method presented by Gray (27, 34) that used partial reduction in combination with disulfide-specific tags for the first time. Here this method is complemented with the introduction of charges, as well as enzymatic cleavage sites, to circumvent the inherent challenges of disulfide mapping in the cyclotides.

The key to Gray's original method was the use of TCEP and its ability to act as a reducing agent under acidic conditions, where the major problem of reshuffling of disulfides in partially reduced peptides is avoided (27, 35). However, during the alkylation process in this original protocol the pH is raised to 8, which dramatically increases the risk of disulfide reshuffling (34). The use of a rapid alkylation procedure with an excess of alkylator (iodoacetamide) partly solved this problem, but to unambiguously identify the correct disulfide species, parallel experiments favoring reshuffling (*i.e.* with ten times lower con-

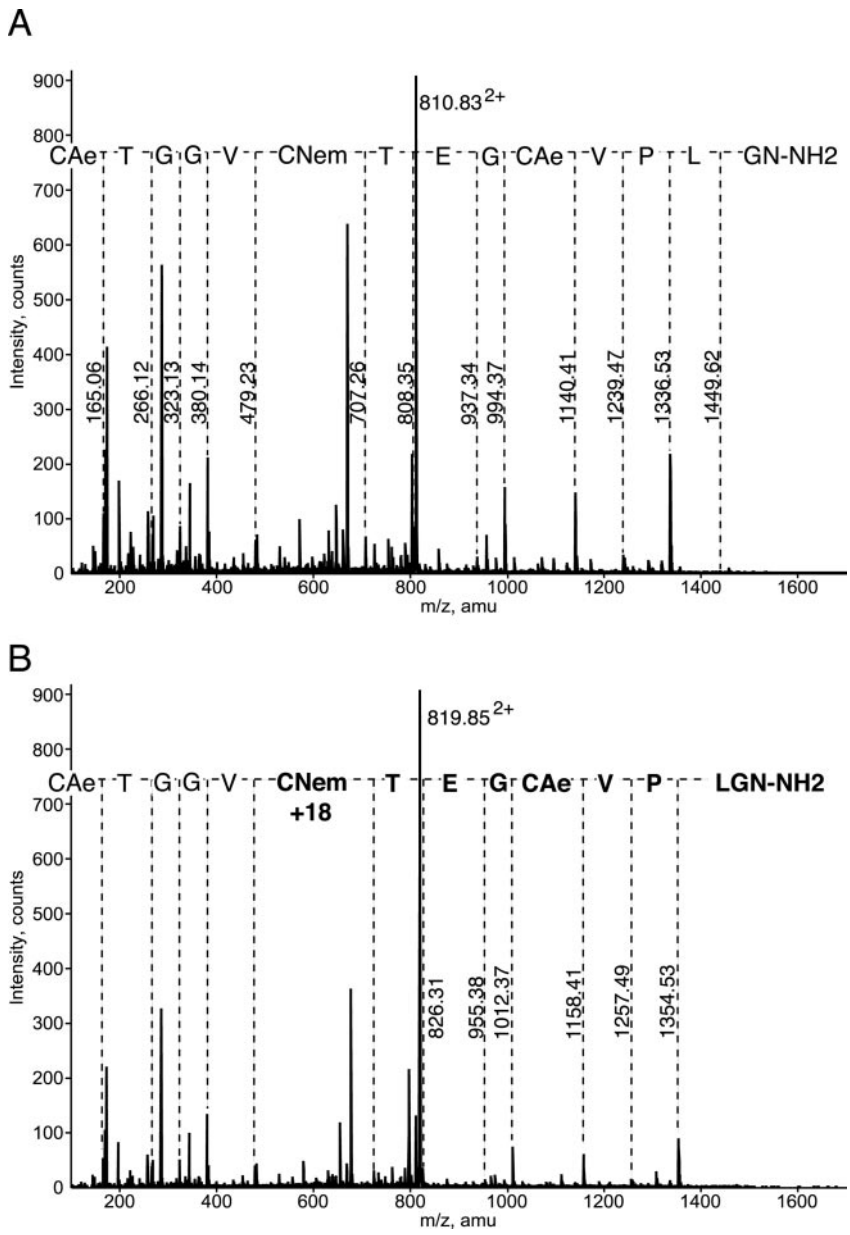

FIG. 5. **MS/MS sequencing.** The spectra show the fragmentation of the doubly charged ion *m/z* 810.86 (*A*) and its +18 adduct (*B*) obtained from the digestion of the derivatized 2S species. The C-terminal fragments (y-series) are marked in the spectra, which confirm both the missed cleavage site Cys(Ae) in this particular fragment and Cys(NEM) as the site of the +18 adduct (*B*).

centration of the alkylating agent) have to be run (26). Comparing the HPLC results from these two experiments then helps in the identification of the correct species. A complete reduction followed by a second alkylation (4-vinylpyridine) is then carried out, after which the positions of two alkylators are determined by Edman sequencing, and thus the disulfide connectivities may be deduced.

Ideally, however, alkylation would be carried out under conditions totally preventing disulfide scrambling. The methodology introduced by Wu and co-workers based on cyanylation of partially reduced disulfide species at pH 3, followed by chemically induced cleavage on the N-terminal side of cysteines, thus seems very appealing (28–30, 36). However as this method did not prove suitable for the cyclotides, the use of NEM as the first alkylator in a two-step alkylating scheme was attempted. This alkylator has been used successfully before at acidic pH, including pH values as low as 3 for the purpose of disulfide analysis (37). For kalata B1 this approach proved to be successful. Alkylation was quantitative, and disulfide reshuffling was kept to a minimum, and thus the identification of the correct species after this first alkylation step was straightforward. After reduction of remaining disulfides, we used bromoethylamine for the second alkylation, thus introducing positively charged Ae groups at the cysteines, which serve as pseudolysine substrates during the enzymatic digestion. The fragments of this digest correspond directly to the numbers and positions of aminoethylated cysteines (15, 38). The aminoethylation also proves extremely valuable during the subsequent analysis as all resulting peptide fragments acquire a charge amenable to MS analyses (15). In comparison to previous methods our approach offers several advantages: it avoids reshuffling, it offers the clean chemistry of enzymatic reactions, and it supplements charged native amino acid residues with introduced ones, thus capitalizing on the advantages of MS analyses (*i.e.* speed and sensitivity).

Using this method, we successfully identified the two main intermediates during the reductive unfolding of kalata B1. The 1S and 2S disulfide species were unambiguously determined to involve CysIII-CysVI and des(CysII-CysV), respectively. This was done by matching the experimentally determined masses to those calculated for all possible fragments with aminoethylated cysteines as cleavage sites, a result that was confirmed by MS/MS sequencing. The identification of these two disulfide bonds gives the third, CysI-CysIV, by deduction. This disulfide

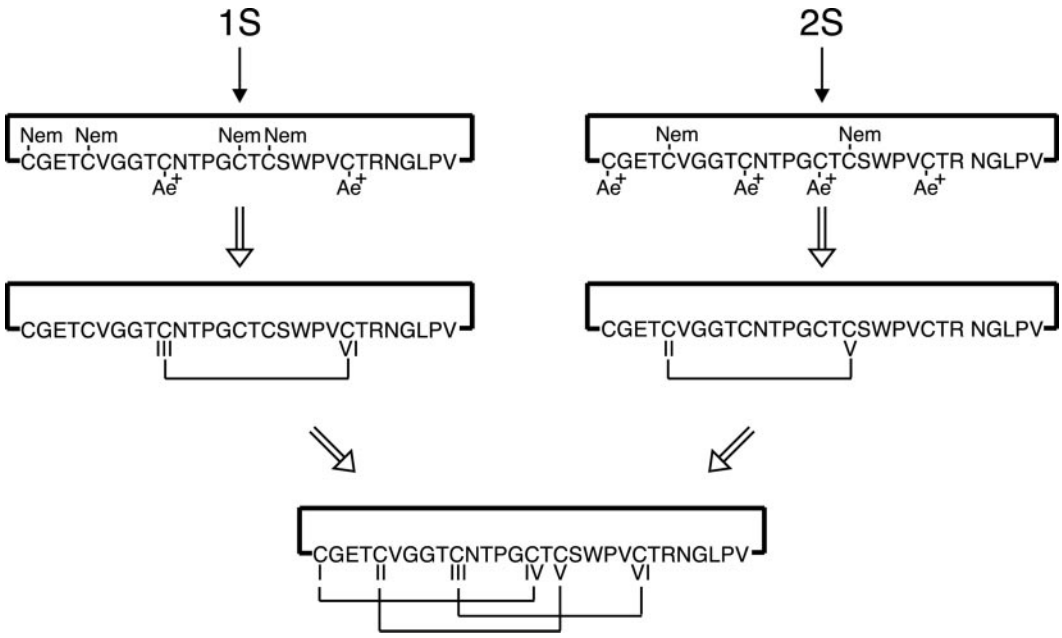

FIG. 6. **Overview of the labeling results that led to the identification of the disulfide connectivity of kalata B1.** After determination of the sites of the different alkylators by MS, the CysIII-CysVI bond could directly be deduced for the 1S species and the CysII-CysV bond for the 2S species, as shown by the *bold arrows*. A combination of these data led to the deduction of the complete set of disulfides as CysI-CysIV, CysII-CysV, CysIII-CysVI.

pattern, the cystine knotted I-IV, II-V, III-VI configuration, is identical with the one proposed in the first structural study by NMR of kalata B1 by Saether *et al.* (22). Compared with the alternative configuration suggested recently by Skjeldal *et al.* (24) only the I-IV disulfide is in common. The laddered disulfide arrangement that would be associated with the additional connectivities II-V and III-IV in the latter proposal is here shown to be incorrect.

Before the present study there had been only one study aimed at chemical disulfide mapping of native cyclotides. Partial acid hydrolysis of circulins A and B, which belong to the bracelet subfamily of cyclotides, was combined with a statistical analysis of the molecular weights of the extensive hydrolytic products as determined by fast atom bombardment to conclude that the circulins contain a cystine knot motif (23). Additional support has been provided by disulfide directed synthesis of these circulins and cyclopsychotride A, where cystine knotted synthetic peptides were shown to coelute with native peptides on RP-HPLC (8). These studies that provide chemical support for the cystine knot arrangement were performed on cyclotides that belong to the bracelet subfamily, and until the present study no proof to complement existing NMR structures has been reported for Möbius cyclotides such as kalata B1. According to Skjeldal *et al.* (24), this lack of chemical support, together with potential structural divergence between the subfamilies, could introduce the possibility of different arrangements in the two subfamilies: laddered for the Möbius cyclotides and cystine knotted for the bracelet cyclotides (24).

To clarify this, we recently reported (25) a structural analyses by high resolution NMR of prototypic members of the two subfamilies, kalata B1 (Möbius) and cycloviolacin O1 (bracelet). A determination of the dihedral angles of cysteine side chains in both peptides clearly suggested that they both contain the cystine knot motif. That study also showed the limitations of previously used NMR techniques for disulfide mapping of cyclotides; the global fold approach used in both previous studies does not discriminate between the two arrangements, and cysteines are too closely packed together for assignment based on individual side chain NOEs (as done by Skjeldal *et al.* (24)). Together with the chemical proof in the present study, the presence of the I-IV, II-V, III-VI disulfide connectivity in kalata B1, and thus in both subfamilies, is unambiguous.

The experimental results from this paper provide unambiguous evidence for the disulfide connectivity and clearly distinguish between a knotted and a laddered topology. However from a chemical determination of disulfide connectivities alone, it is impossible to distinguish between the three seemingly different topological isomers (A, B, C) drawn in Fig. 7. It is tempting to assume that the threaded III-VI disulfide bond shown in Fig. 7*A* is a prerequisite for defining the cystine knot motif, and correspondingly, when the III-VI bond is drawn to the left (Fig. 7*B*) or to the right (Fig. 7*C*) of the embedded ring the structures might not be regarded as cystine knotted structures. However, a closer examination surprisingly reveals that the latter two in fact are topologically identical; the III-VI bond in Fig. 7*B* can be theoretically stretched and pulled over the backbone to give that in Fig. 7*C*. These movements, illustrated in the *boxed inset* in Fig. 7, are perfectly acceptable in the field of topology. However, by stretching and bending alone it is impossible to obtain that in Fig. 7*A* from either *B* or *C* and *vice versa*; this would require that one of the disulfides is broken and formed again at the other side of either the disulfide to the left or the right. That it is impossible to obtain identical topological isomers although the disulfide connectivities are identical is explained by the fact that the CCK motif is topologically chiral and has two possible stereoisomers (39) as shown in Fig. 7*D*. Topologically, *A* is the mirror image of *B* and *C*.

Following this discussion it is clear that the two stereoisomers, although of different topological chirality, both involve a cyclic cystine knot motif. The presence of the cystine knot motif is visualized in Fig. 7 (*A′*, *B′*, and *C′*) via the differential shading of the penetrating disulfide bond and the ring formed by the other two disulfides and their connecting backbone. Note, for example, that although the knot is not immediately apparent in Fig. 7*B*, the topological penetration associated with the cystine knot is clearly illustrated via the shading in Fig. 7*B′* of the same isomer. In summary, the important point is that

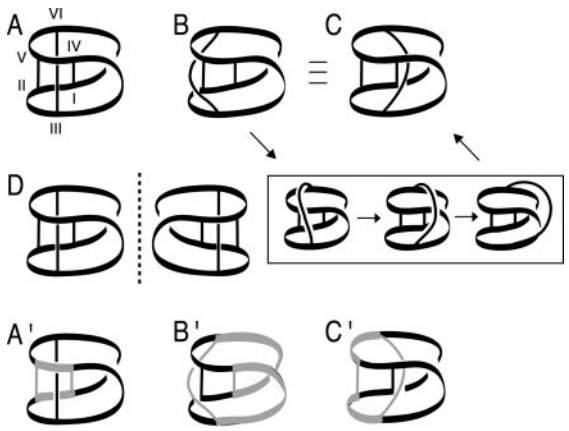

FIG. 7. **Topology of the CCK motif.** The I-IV, II-V, III-VI Cys connectivities may be represented by all of the three seemingly different topological isomers shown in *A*, *B*, and *C*. Topologically, however, *B* can be transformed to *C* via stretching the III-VI disulfide bond over the backbone as illustrated in the *boxed inset*, and they are thus topologically identical. However, it is impossible to obtain *A* from such a continuous deformation of *B* or *C* and *vice versa*. In fact, *A* is the stereoisomer to *B* and *C* as shown in *D*. Hence the same topological motif is present in all isomers. The fact that they all are cystine knots is emphasized in structures *A′*, *B′*, and *C′*, which highlight the disulfide bond (in *black*) penetrating the ring (in *gray*) formed by the other two disulfides and their connecting backbones that define the cystine knot motif in all of them.

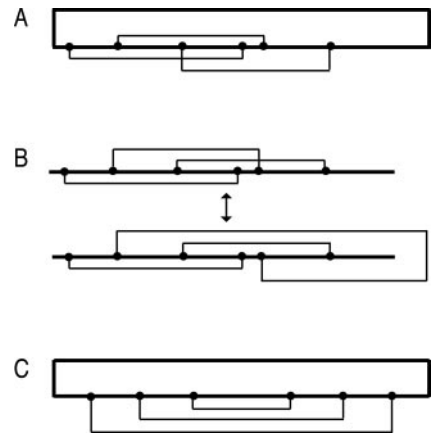

FIG. 8. **The CCK motif is topologically unique.** As shown in *A*, the cyclic backbone (*bold lines*) and the disulfide bonds can not be drawn without any lines crossing. Topologically, this pattern may thus be described as graph theoretically non-planar. This is a unique feature for the CCK motif and is not found in acyclic three disulfide cystine knotted analogues as they may always be re-drawn as planar graphs (as shown in *B*), demonstrating that they are not knotted. However three disulfide bonds in a circular backbone is not a sufficient condition for knotting, as demonstrated by the peptide RTD-1 (42), which has a planar graph representation (as shown in *C*).

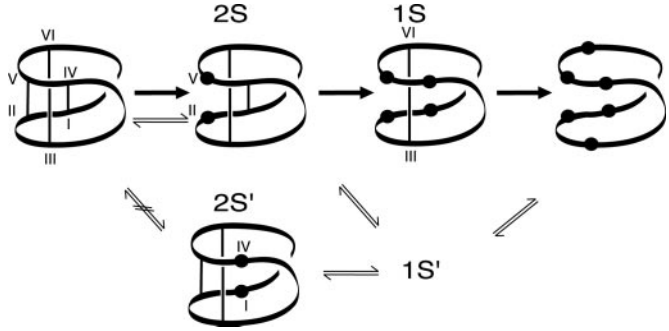

FIG. 9. **Unfolding of kalata B1.** Determination of disulfide connectivities by partial reduction also discloses information about the unfolding pathway, as shown in the *upper row* by the *bold arrows* and the disulfide species identified in this study (*2S* and *1S*). Under the experimental conditions used, the II-V bond is the first to be reduced (*2S*), followed by the I-IV bond (*1S*). In addition, the major folding intermediate isolated and characterized by NMR by Daly *et al.* (32) is shown (marked *2S′*), together with the proposed pathway for the oxidative folding (*thin arrows*) (32). The 2S species identified in the current study in reduction experiments is identical with the 2S species hypothesized to be the precursor to the native conformation in oxidation experiments (32).

the I-IV, II-V, III-VI connectivities as determined here are alone sufficient to establish the topology of the cyclotides as cystine knotted.

Even though movements like shuffling the order of disulfide bonds by pulling and stretching as described above and in Fig. 7 yields topologically similar molecules, they would have a dramatic effect on the topography of the molecule. Combining the data presented here with the information known about the overall fold of kalata B1 derived from NOE data, for which there is no dispute (22, 24, 25), the native topology is clearly established to be that shown schematically in Fig. 7*A*.

Although in this study we unambiguously have established the presence of a cystine knotted motif in the cyclotides using the commonly accepted biochemical definition of a knot (1, 2), it is interesting to note that the CCK motif is not to be regarded as a true knot from a mathematical point of view (39). Mathematically, a knot is a closed curve in three-dimensional space that does not intersect itself and cannot be embedded in the plane without intersections (40). However, the CCK motif does involve an additional level of complexity compared with non-cyclic cystine knotted structures containing three disulfide bonds, which are found in a large number of other peptides and proteins (1, 2). Although these structures may easily be unknotted and drawn in a single plane without any lines crossing each other, the CCK motif cannot (2, 4). In mathematical terms, the former may be described as graph theoretically planar molecules and the CCK motif as graph theoretically non-planar (41). This is illustrated in Fig. 8, where the topology of the CCK motif is compared with a linear analogue with identical cystine connectivity. Fig. 8 also shows that a cyclic backbone alone does not necessarily result in a graph theoretically non-planar molecule, as exemplified by the rhesus theta defensin 1, RTD-1 (42), which contains a laddered arrangement of the three disulfides (43). In fact, the here established I-IV, II-V, III-VI connectivity for kalata B1 is the only possible pattern of a cyclic, three-disulfide peptide that is graph theoretically non-planar (41). This makes the CCK motif topologically unique.

In addition to confirming the cystine connectivities of the prototypic cyclotide kalata B1 the results obtained here also disclose the susceptibility and order of reduction of individual disulfide bonds during reductive unfolding of kalata B1, as shown in Fig. 9. Not surprisingly the penetrating disulfide, CysIII-CysVI, appears to be the most stable one, that is the last one to be broken, and conversely the most readily reduced bond is the one broken to form the 2S species identified in this study, des(CysII-CysV). The identification of the latter confirms the recent proposal by our group (32) that according to structure calculations this disulfide bond is the most surface-exposed of the three and thus is cleaved most readily. In addition, it was also suggested that this intermediate is a possible immediate precursor during oxidative refolding of native kalata B1 (32). Further studies are needed to confirm this hypothesis, but it has been clearly established that the other native-like 2S species that might be considered a possible precursor (*2S′* in Fig. 9) is in fact a kinetic trap that does not directly lead to the fully oxidized peptide in the absence of disulfide shuffling agents (32).

In conclusion, in the present study we have provided definite

chemical proof of the I-IV, II-V, III-VI disulfide connectivity of kalata B1. This has been achieved by the development of a new approach for disulfide analysis, which overcomes the limitations of previously described methods such as a lack of cleavage sites between cysteines and a low or clustered content of basic amino acids. Hence the presence of the cyclic cystine knot motif in this prototypic cyclotide has been confirmed unambiguously.

Acknowledgments—We thank Alun Jones for helpful discussions and technical advice regarding MS and Dan Barry for helpful comments and discussions.

REFERENCES

1. Pallaghy, P. K., Nielsen, K. J., Craik, D. J., and Norton, R. S. (1994) *Protein Sci.* **3,** 1833–1839
2. Craik, D. J., Daly, N. L., and Waine, C. (2001) *Toxicon* **39,** 43–60
3. Craik, D. J. (2001) *Toxicon* **39,** 1809–1813
4. Craik, D. J., Daly, N. L., Bond, T., and Waine, C. (1999) *J. Mol. Biol.* **294,** 1327–1336
5. Gran, L. (1973) *Acta Pharmacol. Toxicol.* **33,** 400–408
6. Gran, L. (1973) *Lloydia* (*Cinci.*) **36,** 174–178
7. Lindholm, P., Göransson, U., Johansson, S., Claeson, P., Gullbo, J., Larsson, R., Bohlin, L., and Backlund, A. (2002) *Mol. Cancer Ther.* **1,** 365–369
8. Tam, J. P., Lu, Y. A., Yang, J. L., and Chiu, K. W. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8913–8918
9. Felizmenio-Quimio, M. E., Daly, N. L., and Craik, D. J. (2001) *J. Biol. Chem.* **276,** 22875–22882
10. Hernandez, J. F., Gagnon, J., Chiche, L., Nguyen, T. M., Andrieu, J. P., Heitz, A., Hong, T. T., Pham, T. T. C., and Nguyen, D. L. (2000) *Biochemistry* **39,** 5722–5730
11. 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., Pannell, L. K., and Boyd, M. R. (1994) *J. Am. Chem. Soc.* **116,** 9337–9338
12. Gustafson, K. R., Walton, L. K., Sowder, R. C., Johnson, D. G., Pannell, L. K., Cardellina, J. H., and Boyd, M. R. (2000) *J. Nat. Prod.* **63,** 176–178
13. Witherup, K. M., Bogusky, M. J., Anderson, P. S., Ramjit, H., Ransom, R. W., Wood, T., and Sardana, M. (1994) *J. Nat. Prod.* **57,** 1619–1625
14. Göransson, U., Luijendijk, T., Johansson, S., Bohlin, L., and Claeson, P. (1999) *J. Nat. Prod.* **62,** 283–286
15. Göransson, U., Broussalis, A. M., and Claeson, P. (2003) *Anal. Biochem.* **318,** 107–117
16. Broussalis, A. M., Göransson, U., Coussio, J. D., Ferraro, G., Martino, V., and Claeson, P. (2001) *Phytochemistry* **58,** 47–51
17. Hallock, Y. F., Sowder, R. C., Pannell, L. K., Hughes, C. B., Johnson, D. G., Gulakowski, R., Cardellina, J. H., and Boyd, M. R. (2000) *J. Org. Chem.* **65,** 124–128
18. Bokesch, H. R., Pannell, L. K., Cochran, P. K., Sowder, R. C., McKee, T. C., and Boyd, M. R. (2001) *J. Nat. Prod.* **64,** 249–250
19. Claeson, P., Göransson, U., Johansson, S., Luijendijk, T., and Bohlin, L. (1998) *J. Nat. Prod.* **61,** 77–81
20. Jennings, C., West, J., Waine, C., Craik, D. J., and Anderson, M. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10614–10619
21. Craik, D. J., Simonsen, S., and Daly, N. L. (2002) *Curr. Opin. Drug Discovery Dev.* **5,** 251–260
22. Saether, O., Craik, D. J., Campbell, I. D., Sletten, K., Juul, J., and Norman, D. G. (1995) *Biochemistry* **34,** 4147–4158
23. Derua, R., Gustafson, K. R., and Pannell, L. K. (1996) *Biochem. Biophys. Res. Comm.* **228,** 632–638
24. Skjeldal, L., Gran, L., Sletten, K., and Volkman, B. F. (2002) *Arch. Biochem. Biophys.* **399,** 142–148
25. Rosengren, K. J., Daly, N. L., Plan, M. R., Waine, C., and Craik, D. J. (2003) *J. Biol. Chem.* **278,** 8606–8616
26. Gray, W. R. (1997) in *Protein Structure* (Creighton, T. E., ed) 2nd Ed., pp. 165–186, Oxford University Press, New York
27. Gray, W. R. (1993) *Protein Sci.* **2,** 1732–1748
28. Wu, J., Gage, D. A., and Watson, J. T. (1996) *Anal. Biochem.* **235,** 161–174
29. Wu, J., and Watson, J. T. (1997) *Protein Sci.* **6,** 391–398
30. Wu, J., and Watson, J. T. (1998) *Anal. Biochem.* **258,** 268–276
31. Shon, K. J., Hasson, A., Spira, M. E., Cruz, L. J., Gray, W. R., and Olivera, B. M. (1994) *Biochemistry* **33,** 11420–11425
32. Daly, N. L., Clark, R. J., and Craik, D. J. (2003) *J. Biol. Chem.* **278,** 6314–6322
33. Ishii, Y., and Lehrer, S. S. (1986) *Biophys. J.* **50,** 75–89
34. Gray, W. R. (1993) *Protein Sci.* **2,** 1749–1755
35. Burns, J. A., Butler, J. C., Moran, J., and Whitesides, G. M. (1991) *J. Org. Chem.* **56,** 2648–2650
36. Wu, J., Yang, Y., and Watson, J. T. (1998) *Protein Sci.* **7,** 1017–1028
37. van den Hooven, H. W., van den Burg, H. A., Vossen, P., Boeren, S., de Wit, P. J., and Vervoort, J. (2001) *Biochemistry* **40,** 3458–3466
38. Lindley, H. (1956) *Nature* **178,** 647–648
39. Liang, C. Z., and Mislow, K. (1995) *J. Am. Chem. Soc.* **117,** 4201–4213
40. Mislow, K., and Liang, C. Z. (1996) *Croat. Chem. Acta* **69,** 1385–1403
41. Rücker, C., and Meringer, M. (2002) *MATCH Commun. Math. Comput. Chem.* **45,** 153–172
42. Tang, Y. Q., Yuan, J., Osapay, G., Osapay, K., Tran, D., Miller, C. J., Ouellette, A. J., and Selsted, M. E. (1999) *Science* **286,** 498–502
43. Trabi, M., Schirra, H. J., and Craik, D. J. (2001) *Biochemistry* **40,** 4211–4221
44. Schneider, T. D., and Stephens, R. M. (1990) *Nucleic Acids Res.* **18,** 6097–6100

Disulfide Mapping of the Cyclotide Kalata B1: CHEMICAL PROOF OF THE CYCLIC CYSTINE KNOT MOTIF

Ulf Göransson and David J. Craik

J. Biol. Chem. 2003, 278:48188-48196.
doi: 10.1074/jbc.M308771200 originally published online September 5, 2003
