A common structural motif incorporating a cystine knot and a triple-stranded {beta}-sheet in toxic and inhibitory polypeptides

P. K. PALLAGHY, K. J. NIELSEN, D. J. CRAIK and R. S. NORTON

Protein Sci. 1994 3: 1833-1839

Protein Science (1994), *3*:1833–1839. Cambridge University Press. Printed in the USA.

A common structural motif incorporating a cystine knot and a triple-stranded β-sheet in toxic and inhibitory polypeptides

PAUL K. PALLAGHY,[1] KATHERINE J. NIELSEN,[2] DAVID J. CRAIK,[2] AND RAYMOND S. NORTON[1]

[1] NMR Laboratory, Biomolecular Research Institute, 381 Royal Parade, Parkville 3052, Australia
[2] School of Pharmaceutical Chemistry, Victorian College of Pharmacy, 381 Royal Parade, Parkville 3052, Australia

(RECEIVED June 2, 1994; ACCEPTED August 5, 1994)

Abstract

A common structural motif consisting of a cystine knot and a small triple-stranded β-sheet has been defined from comparison of the 3-dimensional structures of the polypeptides ω-conotoxin GVIA (*Conus geographus*), kalata BI (*Oldenlandia affinis* DC), and CMTI-I (*Curcurbita maxima*). These 3 polypeptides have diverse biological activities and negligible amino acid sequence identity, but each contains 3 disulfide bonds that give rise to a cystine knot. This knot consists of a ring formed by the first 2 bonds (1–4 and 2–5) and the intervening polypeptide backbone, through which the third disulfide (3–6) passes. The other component of this motif is a triple-stranded, antiparallel β-sheet containing a minimum of 10 residues, XXC_2, XC_5X, XXC_6X (where the numbers on the half-cystine residues refer to their positions in the disulfide pattern). The presence in these polypeptides of both the cystine knot and antiparallel β-sheet suggests that both structural features are required for the stability of the motif. This structural motif is also present in other protease inhibitors and a spider toxin. It appears to be one of the smallest stable globular domains found in proteins and is commonly used in toxins and inhibitors that act by blocking the function of larger protein receptors such as ion channels or proteases.

Keywords: ω-conotoxin GVIA; cystine knot; kalata; NMR; protease inhibitor; protein structure

With increasing numbers of polypeptide and protein structures being solved in recent years, it has become apparent that there is a limited number of topologically different protein folds employed in nature (Chothia, 1992; Orengo et al., 1993). Proteins with little or no sequence identity can have remarkably similar secondary structure patterns and tertiary folds. In this paper we analyze the occurrence of a cystine knot and triple-stranded β-sheet motif in 3 polypeptides, which, although similar in size, are unrelated functionally and originate from biologically diverse sources, as follows: (1) ω-CgTx, a 27-residue polypeptide found in venom of the cone shell *Conus geographus* (Olivera et al., 1985), which is neurotoxic as a consequence of its high-affinity binding to N-type voltage-gated Ca^{2+} channels; (2) kalata B1, a 29-residue cyclic polypeptide that causes uterine contractions and is the principal active component from the tropical plant *Oldenlandia affinis* DC (Gran, 1973); and (3) CMTI-I, a 29-residue trypsin inhibitor found in seeds of the pumpkin *Curcurbita maxima* (Wilusz et al., 1983; Polanowski et al., 1987). CMTI-I is one of a number of similar polypeptides belonging to the squash family of serine protease inhibitors (Laskowski & Kato, 1980; Richardson, 1991).

The 3D structures in aqueous solution of ω-CgTx (Davis et al., 1993; Pallaghy et al., 1993; Sevilla et al., 1993; Skalicky et al., 1993), kalata (O. Saether, D.J. Craik, I.D. Campbell, K. Sletten, J. Juul, & D. Norman, ms. in prep.), CMTI-I (Holak et al., 1989, 1991; Nilges et al., 1991), and related members of the squash family (Heitz et al., 1989; Nielsen et al., 1994) have been determined from NMR data, and it is apparent that these molecules display similarities at both the secondary and tertiary structure levels. In particular, all 3 contain a triple-stranded, antiparallel β-sheet and a cystine knot,[3] the latter consisting of a ring formed by 2 disulfide bridges (1–4 and 2–5) and the inter-

Reprint requests to: Raymond S. Norton, NMR Laboratory, Biomolecular Research Institute, 381 Royal Parade, Parkville 3052, Australia; e-mail: ray@mel.dbe.csiro.au.

Abbreviations: ω-CgTx, ω-conotoxin GVIA; kalata, kalata BI; CMTI-I, *Cucurbita maxima* trypsin inhibitor I; EETI-II, *Ecballium elaterium* trypsin inhibitor II; CPI, potato carboxypeptidase inhibitor; ω-Aga IVB, ω-agatoxin IVB; 3D, 3-dimensional.

[3] Topologically, the cystine knots in ω-CgTx and CMTI-I are "pseudolinks" because they can be continuously deformed into unknotted structures (Benham & Saleet, 1993). In kalata, however, the knot is genuine due to the cyclic nature of the backbone, which prevents its unraveling. In this work the term "knot" is used to describe both types of structures.

connecting backbone through which the third disulfide (3–6) passes. In this paper we describe the key features of this motif, designated the "inhibitor cystine knot motif," and suggest that it will be found in an increasing number of polypeptides that act as inhibitors of larger target proteins.

Recently, McDonald and Hendrickson (1993) and Murray-Rust et al. (1993) have identified the presence of a common cystine knot plus β-sheet topology in the crystal structures of transforming growth factor-β2, nerve growth factor, and platelet-derived growth factor-BB. Although there are some similarities between this motif and the inhibitor cystine knot motif described in this paper, in that 3 disulfides and antiparallel β-sheet are present, the 2 motifs are topologically distinct and cannot be superimposed. Furthermore, the growth factors have acquired significant loops during the course of evolution that presumably enable them to trigger cooperative conformational changes in their cognate receptors as part of the process of signal transduction. By contrast, polypeptides containing the inhibitor cystine knot motif are of low molecular mass and may lack the conformational versatility necessary to participate in cooperative conformational changes upon interaction with their target proteins, behaving therefore mainly as inhibitors.

Results and discussion

Comparison of primary, secondary, and tertiary structure

Figure 1 shows the sequences and disulfide pairings of the 3 molecules central to this work, ω-CgTx, kalata, and CMTI-I, as well as 3 others, EETI-II, CPI, and ω-Aga IVB (a P-type Ca^{2+} channel blocker from the American funnel-web spider *Agelenopsis aperta*). With the exception of ω-Aga IVB, which has a fourth disulfide bridge, the pattern of disulfide pairings is the same for each of these molecules. ω-Aga IVB, the structure of which was determined recently (Yu et al., 1993), is included because its fourth disulfide bridge can be considered independent of the other 3 bridges, which form a cystine knot with the same topology as that of the other molecules in Figure 1 (although this was not identified by Yu et al., 1993). EETI-II and CPI are included because broad topological and cystine knot similarities between these polypeptides and ω-CgTx and CMTI-I have been noted previously (Bode et al., 1989; Le-Nguyen et al., 1990; Chiche et al., 1993; Pallaghy et al., 1993). The squash protease inhibitors CMTI-I and EETI-II share 70% sequence identity and a comparison of their 3D structures (Nielsen et al., 1994) shows that their structures are very similar both globally and locally. For most purposes in this paper, therefore, the EETI-II and CMTI-I structures can be considered as identical. The other molecules shown in Figure 1 have little sequence identity apart from the 6 half-cystine residues with disulfide pairings 1–4, 2–5, 3–6. The only exception to this is the common sequence Gly-Gly-Thr in kalata and ω-Aga IVB. In ω-CgTx and ω-Aga IVB, the third and fourth half-cystine residues are adjacent in the sequence.

Broad structural similarities between individual proteins in this group have been noted previously. Thus, in the 3D structure of ω-CgTx a small triple-stranded β-sheet was identified and its similarity to the corresponding secondary structure element in EETI-II noted (Pallaghy et al., 1993; Skalicky et al., 1993). Recently, a structural alignment between EETI-II and CPI (Chiche et al., 1993) revealed that the cystine knot and β-strands of EETI-II superimposed well in 3 dimensions with the cystine knot and extended segments (probably β-strands) of CPI. In the present work, we have found that the cystine knot and β-strands of ω-CgTx, kalata, and CMTI-I also superimpose very well. Their global folds, shown in Figure 2, are remarkably similar, with each displaying 4 common structural segments in sequence, as follows: a peripheral β-strand (strand I), a connecting region containing turns or 3_{10}-helix, the other peripheral β-strand (II), and the central β-strand (III), the latter two being joined by a turn to give a β-hairpin structure. These 4 segments are separated by 3 topologically important chain reversals, the positions of which are maintained in all 3 molecules.

Figure 3 shows a superposition of ω-CgTx, kalata, and CMTI-I based on the common β-sheet residues and the C^{β} and S^{γ} atoms of the half-cystine residues in the sheet. The β-sheet regions (shown in pink) superimpose very well, even to the extent of

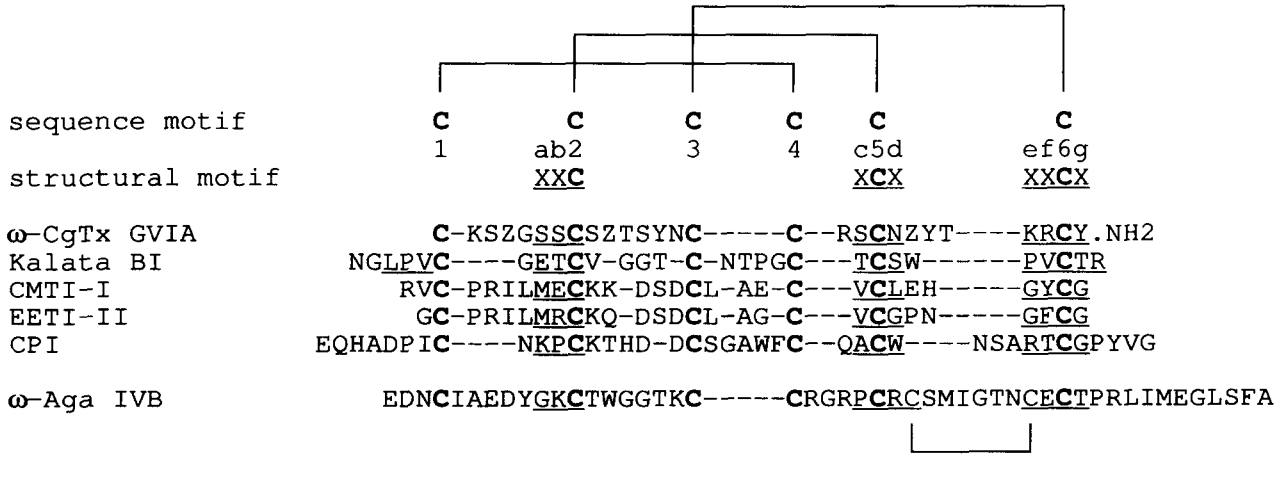

Fig. 1. Polypeptide sequences aligned on the basis of half-cystine positions and β-strand (underlined) hydrogen bonding patterns. The half-cystines are highlighted in bold. The minimal β-strand motif common to all molecules is XXC (strand I), XCX (strand II), and XXCX (strand III). X represents any residue other than cysteine (C), and Z represents hydroxyproline. Only ω-Aga IVB has a fourth disulfide bridge.

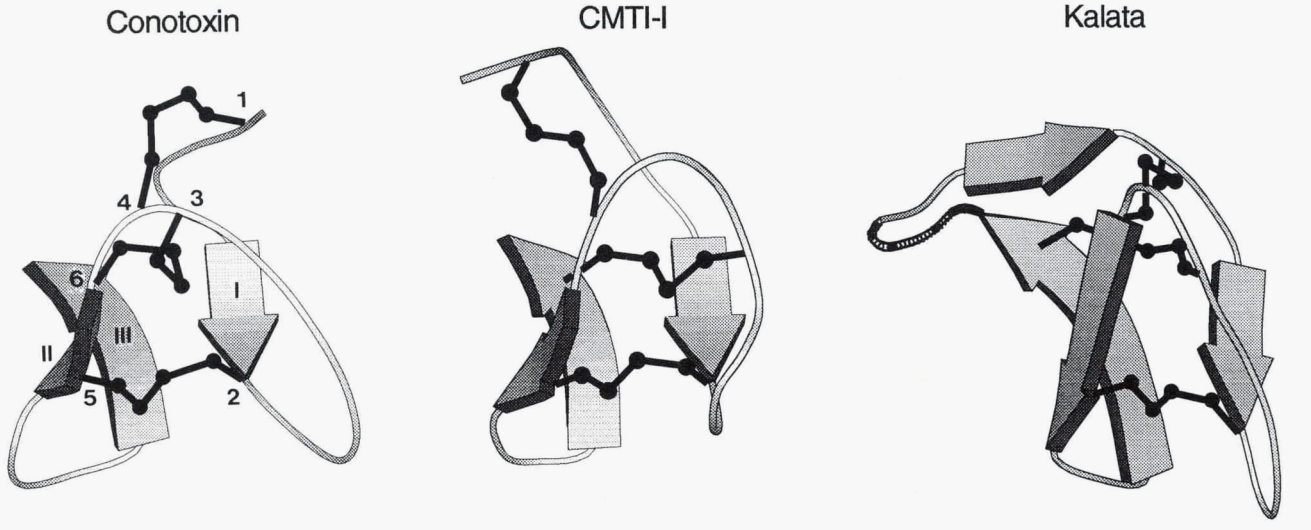

Fig. 2. Richardson-type diagrams of ω-CgTx (Pallaghy et al., 1993), CMTI-I (Holak et al., 1989; Nilges et al., 1991), and kalata (O. Saether et al., ms. in prep.), with the β-strands represented by arrows and the disulfide bridges in black. The half-cystines in the disulfide bridges are numbered as in Figure 1. Note that 2 of the topologically important chain reversals occur at bottom left and bottom right and the third topologically important chain reversal occurs at top center (front). Kalata, due to its cyclic nature, has a fourth topologically important chain reversal at top left. The furthest right β-strand (strand I) in kalata is represented by 2 arrows due to disruption by an S3-type β-bulge. The diagrams were generated by MOLSCRIPT (Kraulis, 1991).

hydrogen-bonding donors and acceptors in most cases, despite the fact that the sheets are quite distorted. The 2–5 and 3–6 disulfide bridges also superimpose well, which is not surprising because half-cystines 2, 5, and 6 are part of the sheet. Pairwise RMS differences over the β-strand residues and half-cystine C^{β} and S^{γ} atoms in the motif are in the range 0.87–1.14 Å. By contrast, the 1–4 disulfide bridge shows greater structural variation (Fig. 3). This is probably due to the proximity of the first half-cystine residue to the N-terminus in ω-CgTx and CMTI-I, allowing this bond greater structural flexibility.

The β-sheet residues and cystine knot of ω-CgTx, kalata, and CMTI-I also superimpose well with the extended segments and cystine knot of CPI (pairwise RMS differences over the motif residues of 0.73, 0.96, and 0.67 Å, respectively). This is consistent with the results of Chiche et al. (1993) and the previously mentioned structural similarity between CMTI-I and EETI-II. On the basis of these superpositions and the similar hydrogen bonding patterns in all 4 molecules, the extended segments of CPI can be described as a small, triple-stranded β-sheet as in EETI-II and ω-CgTx, as was suggested by Chiche et al. (1993) but not commented on in the original X-ray crystallographic structure determination (Rees & Lipscomb, 1982).

The inhibitor cystine knot motif

The key features of the inhibitor cystine knot motif are a triple-stranded, antiparallel β-sheet and a cystine knot, as shown schematically in Figure 4. The 3 strands of the β-sheet are comprised

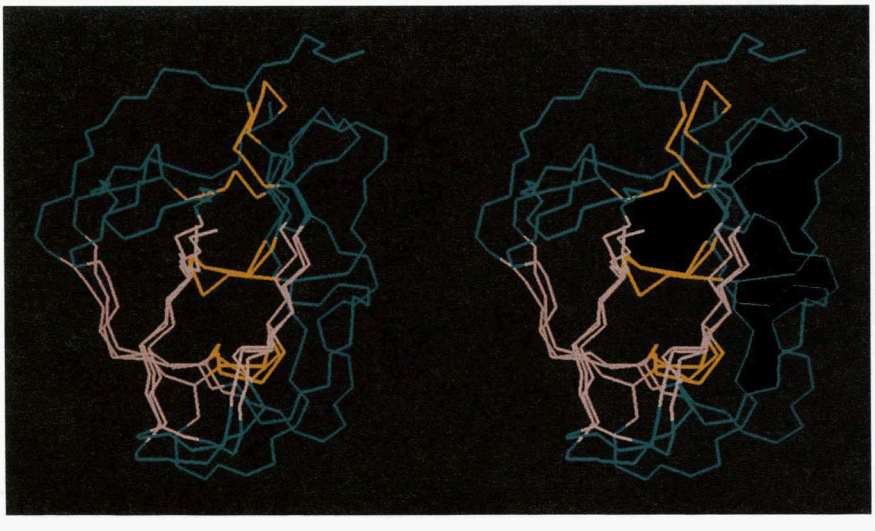

Fig. 3. Stereo view of the superposition of kalata and CMTI-I onto ω-CgTx, based on the backbone atoms of the β-strand motif residues XXC, XCX, XXCX (pink), and the motif cysteine C^{β} and S^{γ} atoms (yellow). The RMS differences from ω-CgTx over these atoms were 1.14 Å (kalata) and 0.87 Å (CMTI-I). The RMS difference of kalata from CMTI-I was 0.97 Å.

of a minimum of 10 residues, as follows: $X_aX_bC_2$, $X_cC_5X_d$, $X_eX_fC_6X_g$, where C represents a half-cystine (numbered to indicate its position in the disulfide pairing pattern) and X represents any other residue. The locations of these strands in the amino acid sequences of the 6 polypeptides considered here are shown in Figure 1. Of the 3 disulfide bonds that also define this motif, one crosslinks the first and second strands of the sheet (C_2 to C_5), one extends from the third strand to the loop linking the first and second strands (giving rise to the knot), and one links C_1, which is N-terminal to the β-sheet, with C_4, which is also in the loop linking the first and second strands. In addition to the position of the half-cystines in the β-strands (relative to the hydrogen bonding pattern), the packing of the strands in the sheet is also maintained, with X_a adjacent to C_6, which is in turn adjacent to C_5 (Fig. 4). This motif is also characterized by the hydrogen bonding pattern in the β-sheet, the minimal pattern involving hydrogen bonds between NH(C_6) and CO(X_a), NH(C_2) and CO(X_e), NH(X_g) and CO(X_c), and NH(X_d) and CO(X_f). Residues X_b and C_5 are not involved in hydrogen bonding due to their positions as the central residues of the 2 peripheral strands.[4] These β-sheet hydrogen bonds represent the minimal common pattern found; additional β-sheet hydrogen bonds are present in all members of this structural family, but they are either not present or not well defined in all molecules. The cystine knot consists of a disulfide bridge, C_3 to C_6, that passes through a closed ring formed by the disulfide bridges C_1 to C_4 and C_2 to C_5 and the 2 segments of backbone $C_1\cdots X_aX_bC_2$ and $C_4\cdots X_cC_5$. The size of the knot can be quantified according to the number of residues comprising the closed ring, which differs for each of the molecules under consideration. In ω-CgTx, 12 residues form this ring, whereas in the squash polypeptides and kalata, 11 and 8 residues, respectively, are involved.

In the inhibitor cystine knot motif identified here, the positions of the half-cystines in the β-strands (relative to the hydrogen bonding pattern) and the packing of the β-strands in the sheet are maintained. This motif is qualitatively different from the cystine knot motif identified in a family of growth factors (McDonald & Hendrickson, 1993; Murray-Rust et al., 1993), which emphasizes the approximately conserved gap size between half-cystines in the ring, $CXX_{1-7}XC$, CXC, where X_{1-7} represents between 1 and 7 non-cysteine residues. Using similar nomenclature, the inhibitor cystine knot motif would be described as $CXX_{1-5}XC$, $CX_{1-4}C$, but this is not sufficient to define the motif or distinguish it from the growth factor motif. In both motifs, the 2 backbone segments of the ring are part of β-strands that are similarly directed and not linked by hydrogen bonds, but in the inhibitor motif, the disulfide bond passing through the ring emanates from the central β-strand, which is antiparallel to the 2 peripheral strands, whereas in the growth factor motif, the half-cystines of this disulfide join flanking β-strands that are antiparallel with, but not central to, the first 2 strands.

The core of the inhibitor motif identified here, XXC, XCX, XXCX, is not recognizable from sequence alone. Considering the 6 polypeptides in Figure 1, however, the variations in the spacings between successive half-cystine residues may be summarized as $CX_{3-7}CX_{4-6}CX_{0-5}CX_{1-4}CX_{4-10}C$. A search for amino acid sequences having this disposition of half-cystines identified several hundred sequences, representing about 20 nonhomologous types of polypeptides. In addition to ω-CgTx, CMTI-I/EETI-II, and CPI, 3 of these are polypeptides of similar size to those considered in this work, viz., huwentoxin-I from a Chinese bird spider (Liang et al., 1993), gurmarin from an Indian plant (Kamei et al., 1992), and the race-specific eliciter A9 from a fungus (Van Kan et al., 1991). It is possible that these molecules also contain the inhibitor cystine knot motif, but confirmation of this must await determination of their 3D structures and demonstration that they satisfy the structural criteria outlined above. Not all polypeptides with the 1–4, 2–5, 3–6 disulfide pairing and a similar triple-stranded β-sheet contain this motif. Charybdotoxin (Bontems et al., 1991) is one that does not, although neither does it have the same gap sizes as the sequences in Figure 1.

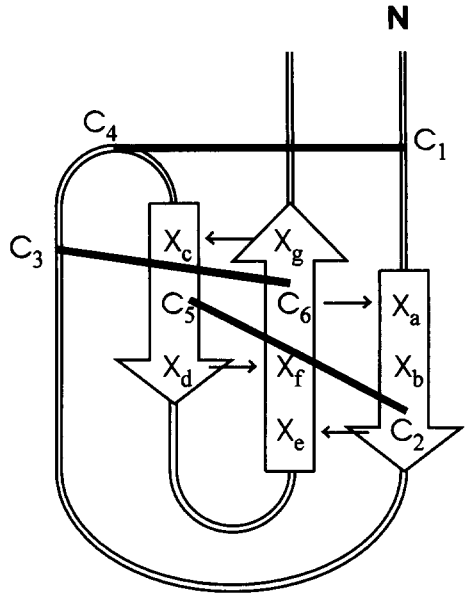

Fig. 4. Schematic diagram of the motif identified, XXC, XCX, XXCX, where C represents a half-cystine residue and X represents any other residue. The common hydrogen bonds are indicated by arrows directed from the donor to the acceptor. Because the lengths of the strands vary slightly from one molecule to another, the positions of the half-cystines in the strands relative to the common hydrogen bonds are significant.

Polypeptides with 4 disulfide bridges

ω-Aga IVB, as mentioned above, has a fourth disulfide bridge (Cys 27–Cys 34) that can be considered independent of the other 3 disulfide bridges, which form a cystine knot topologically identical to the inhibitor motif knot. Three of the knot half-cystine residues originate in the same positions of the triple-stranded β-sheet as in the inhibitor motif. In terms of Figure 4, residue X_e and the residue after X_d form the fourth disulfide bridge, spanning strands II and III. The loop between these 2 strands is much larger in ω-Aga IVB than in the other molecules (Fig. 1), perhaps in order to accommodate this disulfide bridge. Although other polypeptides have 4 disulfide bridges and a similar triple-stranded β-sheet, for example, the plant thionins and the scor-

[4] Prolines are substituted in the sheet only for residues that act as hydrogen bond acceptors or do not participate in hydrogen bonds at all. Thus, X_b in CPI, X_e in kalata, and X_c in ω-Aga IVB are prolines, with the CO groups of X_e and X_c participating in hydrogen bonds.

pion toxin CsE v3 (Bruix et al., 1993), they have neither the same disulfide pairing as ω-Aga IVB nor the cystine knot.

Comparison of local structure

In addition to having common global folds (Figs. 2, 3), kalata, CMTI-I, and ω-CgTx are also similar at the level of their local polypeptide backbone structures. Alignment of these polypeptides on the basis of local backbone structure rather than sequence identity, as summarized in Table 1, shows that there are several similar structural features, apart from the triple-stranded β-sheet described above. In particular, the chain reversals have high positional identity. For example, β-strand I is followed immediately by a chain reversal in each case, which is in turn followed by the similar ϕ/ψ combinations $\alpha\alpha\gamma$, $\alpha\alpha\alpha$, or $\alpha\alpha\gamma/\alpha\gamma\alpha$ (Table 1) in kalata, CMTI-I, and ω-CgTx, respectively (although ω-CgTx is not well defined in this region). This region of CMTI-I has been described as being a distorted 3_{10}-helix (Holak et al., 1989). Preceding β-strand II is a turn, but the nature of this turn differs in each molecule. In all 3 polypeptides, the β-strands II and III are linked by a turn to form a β-hairpin. Although this turn consists of 5 residues for ω-CgTx, the central residues lie in the α-region of a Ramachandran plot, as do the central residues of the corresponding 4-residue turns in kalata and CMTI-I. Interestingly, both ω-CgTx and CMTI-I have a G1-type β-bulge at the end of this turn, the residue with a positive ϕ angle (γ configuration) being Gly 26 in CMTI-I and, somewhat unexpectedly, Lys 24 in ω-CgTx.

The regions preceding the first β-strands in CMTI-I, ω-CgTx, and kalata display the greatest differences among the structures. In CMTI-I, the region encompassing residues 1–6, which incorporates the active site, is the most flexible part of the molecule in solution and is relatively extended (Holak et al., 1989). The corresponding region in kalata is part of a β-sheet, which is in-

Table 1. *Comparison of local ϕ–ψ conformations for kalata, CMTI-I, and ω-CgTx*

	Kalata		CMTI-I		ω-CgTx		
Position	kesidue	ϕ–ψ[a]	Residue	ϕ–ψ[a]	Residue	ϕ–ψ[a]	Secondary structure[b]
1	N8	γ	R1				
2	G9	γ	V2	β_P			
3	L10	β_E	C3	β_E	C1		
4	P11	β_P	P4	β_P	K2	β_P	
5	V12	α/γ	R5	β_P	S3	β_P	
6	C13	α	I6	β_E	X4	β_P	
7			L7	β_E	G5	γ	
8	G14	γ	M8	β_E	S6	β_P	β-Sheet
9	E15	β_E	E9	β_P	S7	β_P	β-Sheet
10	T16	β_E	C10	β_E	C8	β_P	β-Sheet
11					S9	β_E	
12	C17	α[c]	K11	α	X10	α	CR
13			K12	β_E			
14	V18	α	D13	α	T11	α	+
15	G19	α	S14	α	S12	α/γ	+
16	G20	γ	D15	α	Y13	γ/α	+
17	T21	β_E			N14	α	
18	C22	[c]	C16	β	C15	β_P	
19	N23	α/γ					
20	T24	β	L17	β			
21	P25	α	A18	ε	C16	α	CR
22	G26	γ	E19	α	R17	α	CR
23	C27	β	C20	β_P			
24	T28	β_E	V21	β_E	S18	β_E	β-Sheet
25	C29	β_E	C22	β	C19	β_P	β-Sheet
26	S1	β	L23	β	N20	β_P	β-Sheet
27					X21	α	
28	W2	α	E24	α	Y22	γ/α	CR
29	P3	α	H25	α	T23	α/γ	CR
30	V4	β_E	G26	γ	K24	γ	β-Sheet
31	C5	β	Y27	β_E	R25	β_E	β-Sheet
32	T6	β_E	C28	β_P	C26	β_P	β-Sheet
33	R7	β_E	G29		Y27		β-Sheet

[a] Regions in Ramachandran plot, based on the definitions of Adzhubei et al. (1987), as follows: β_E, idealized β-strand conformation; β_P, polyproline conformation; α, right-handed α-helical conformation; γ, region near left-handed α-helical conformation; ε, region with positive ϕ angle and large negative ψ angle.

[b] Secondary structure features common to all 3 proteins. CR, chain reversal; +, region may contain loop or 3_{10}-helix, but chain reversal does not occur.

[c] Residue scattered in Ramachandran plot.

terrupted by an S3 β-bulge (Chan et al., 1993). In ω-CgTx, the N-terminal region is extended, except for a type II β-turn that leads into strand I and is part of a G1T β-bulge (Chan et al., 1993). The precise locations of the receptor binding domains in kalata and ω-CgTx are not yet known, although Lys 2 is important for the activity of ω-CgTx (Lampe et al., 1993; Sato et al., 1993).

Conclusions

The occurrence of the inhibitor cystine knot motif, consisting of a cystine knot and triple-stranded antiparallel β-sheet, in polypeptides from diverse biological sources and with unrelated amino acid sequences, suggests that it is an energetically favorable and stable structural element. The triple-stranded β-sheet, although small, is likely to be important in stabilizing the structure, with further stability being conferred by the disulfide bridge C_2 to C_5, which spans the 2 peripheral β-strands. This bridge may also be responsible for the fact that all molecules share a similar, nonstandard distortion of the sheet. It is significant that bridges C_2 to C_5 and C_3 to C_6 superimpose well in all 3 sets of structures because the analogous disulfide bridges in EETI-II and CPI have also been found to be structurally conserved and it has been proposed that the region stabilized by these bridges forms a 2-disulfide stable motif serving as a basic scaffold on which active sites are anchored (Chiche et al., 1993). Furthermore, the Cys 2–Cys 19 disulfide bridge in EETI-II, which is analogous to the poorly superimposable disulfide bridge C_1 to C_4 in CMTI-I, kalata, and ω-CgTx, is considered to play a minor role in the folding of EETI-II (Le-Nguyen et al., 1993).

The biological functions of these molecules are diverse, but at the protein interaction level, at least ω-CgTx, CMTI-I/EETI-II, ω-Aga IVB, and CPI are similar in that they act as antagonists of their respective targets. Indeed, the primary role of the inhibitor cystine knot motif may be to provide a compact and stable framework for the presentation of active residues for specific binding interactions, and it might not be used in cases where a cooperative conformational change leading to activation of a target protein is required.

The site of action of kalata has not yet been identified. It was isolated originally on the basis of its oxytocin-like activity (Gran, 1973), but its function in the plant was not investigated. Examination of its primary sequence (Fig. 1) reveals that it contains a possible trypsin inhibitory site, with the P1 site (using the nomenclature of Schechter & Berger, 1967) at Arg 7 and the P1′ site at Asn 8. The putative P3-P1′ residues are identical in character to the corresponding regions in the potato type II proteinase inhibitors isolated from *Nicotiana alata* (Atkinson et al., 1993). Thus, it is plausible that kalata and the squash proteinase inhibitors share a common function.

There is insufficient information available on the members of this family of polypeptides to judge whether they are related in an evolutionary sense and, if so, whether convergently or divergently. Proteins with similar amino acid sequences or similar biological functions are expected to adopt similar folds, but it has become clear that proteins that are unrelated biologically can also do so. It is emerging that the concept of a finite number of energetically favored α and β secondary structures stabilized by hydrogen bonding (Pauling et al., 1951a, 1951b) can be extended to that of a larger but finite number of energetically favored structural motifs governed by more diverse physical principles.

Materials and methods

In the case of ω-CgTx and kalata, structures calculated by the present authors were used (Pallaghy et al., 1993; O. Saether et al., ms. in prep.). The ω-CgTx structure is available from the Protein Data Bank (Bernstein et al., 1977), deposition number 1CCO. The structure of CMTI-I was obtained from Protein Data Bank deposition 2CTI (Holak et al., 1989; Nilges et al., 1991). The first structure of the NMR family was used for comparison, except in the case of ω-CgTx, where the second structure proved to be more representative of the family as a whole. The molecular graphics program Insight version 2.2.0 (Biosym Technologies, San Diego, California) was used to visualize and superimpose the molecules.

Acknowledgments

We thank Mark Hinds for assistance with the amino acid sequence database searches.

References

Adzhubei AA, Eisenmenger F, Tumanyan VG, Zinke M, Brodzinski S, Esipova NG. 1987. Approaching a complete classification of protein secondary structure. *J Biomol Struct & Dyn 5*:689–704.

Atkinson AH, Heath RL, Simpson RJ, Clarke AC, Anderson MA. 1993. Proteinase inhibitors in *Nicotiana alata* are derived from a precursor protein which is processed into five homologous inhibitors. *Plant Cell 5*: 203–213.

Benham CJ, Saleet J. 1993. Disulfide bonding patterns and protein topologies. *Protein Sci 2*:41–54.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Bode W, Greyling HJ, Huber R, Otlewski O, Wilusz T. 1989. The refined 2.0 Å X-ray structure of the complex between bovine β-trypsin and CMTI-I, a trypsin inhibitor from squash seeds (*Cucurbita maxima*). *FEBS Lett 242*:285–292.

Bontems F, Roumestand C, Gilquin B, Menez A, Toma F. 1991. Refined structure of charybdotoxin: Common motifs in scorpion toxins and insect defensins. *Science 254*:1521–1523.

Bruix M, Jimenez MA, Santoro J, Gonzalez C, Colilla FJ, Mendez E, Rico M. 1993. Solution structures of 1-H and 1-P thionins from barley and wheat endosperm determined by ^{1}H-NMR: A structural motif common to toxic arthropod proteins. *Biochemistry 32*:715–724.

Chan AWE, Hutchinson EG, Harris D, Thornton JM. 1993. Identification, classification, and analysis of beta-bulges in proteins. *Protein Sci 2*:1574–1590.

Chiche L, Heitz A, Padilla A, Le-Nguyen D, Castro B. 1993. Solution conformation of a synthetic bis-headed inhibitor of trypsin and carboxypeptidase A: New structural alignment between the squash inhibitors and the potato carboxypeptidase inhibitor. *Protein Eng 6*:675–682.

Chothia C. 1992. One thousand families for the molecular biologist. *Nature 357*:543–544.

Davis JH, Bradley EK, Miljanich GP, Nadasdi L, Ramachandran J, Basus VJ. 1993. Solution structure of ω-conotoxin GVIA using 2-D NMR spectroscopy and relaxation matrix analysis. *Biochemistry 32*:7396–7405.

Gran L. 1973. An oxytocic principle found in *Oldenlandia affinis* DC. *Meddelelser 32*:173–180.

Heitz A, Chiche L, Le-Nguyen D, Castro B. 1989. ^{1}H 2D NMR and distance geometry study of the folding of *Ecballium elaterium* trypsin inhibitor, a member of the squash inhibitors family. *Biochemistry 28*: 2392–2398.

Holak TA, Gondol D, Otlewski J, Wilusz T. 1989. Determination of the complete three-dimensional structure of the trypsin inhibitor from squash seeds in aqueous solution by nuclear magnetic resonance and a combination of distance geometry and simulated annealing. *J Mol Biol 210*:635–648.

Holak TA, Habazettl J, Oschkinat H, Otlewski J. 1991. Structures of proteins in solution derived from homonuclear three-dimensional NOE–NOE nuclear magnetic resonance spectroscopy. High-resolution structure of squash trypsin inhibitor. *J Am Chem Soc 113*:3196–3198.

Kamei K, Takano R, Miyasaka A, Imoto T, Hara S. 1992. Amino acid sequence of sweet-taste-suppressing peptide (gurmarin) from the leaves of *Gymnema sylvestre*. *J Biochem (Tokyo) 111*:109–112.

Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Lampe RA, Lo MMS, Keith RA, Horn BM, McLane MW, Herman JL, Spreen RC. 1993. Effects of site-specific acetylation on ω-conotoxin GVIA binding and function. *Biochemistry 32*:3255–3260.

Laskowski M Jr, Kato I. 1980. Protein inhibitors of proteinases. *Annu Rev Biochem 49*:593–626.

Le-Nguyen D, Heitz A, Chiche L, Castro B, Boigegrain RA, Favel A, Coletti-Previero MA. 1990. Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. *Biochimie 72*: 431–435.

Le-Nguyen D, Heitz A, Chiche L, El Hajji M, Castro B. 1993. Characterization and 2D NMR study of the stable [9–21, 15–27] 2 disulfide intermediate in the folding of the 3 disulfide trypsin inhibitor EETI II. *Protein Sci 2*:165–174.

Liang SP, Zhang DY, Pan X, Chen Q, Zhou PA. 1993. Properties and amino acid sequence of huwentoxin-I, a neurotoxin purified from the venom of the Chinese bird spider *Selenocosmia huwena*. *Toxicon 31*:969–978.

McDonald NQ, Hendrickson WA. 1993. A structural superfamily of growth factors containing a cystine knot motif. *Cell 73*:421–424.

Murray-Rust J, McDonald NQ, Blundell TL, Hosang M, Oefner C, Winkler F, Bradshaw RA. 1993. Topological similarities in TGF-β2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. *Structure 1*:153–159.

Nielsen KJ, Alewood D, Andrews J, Kent SBH, Craik DJ. 1994. An ^{1}H NMR determination of the three-dimensional structures of mirror image forms of a Leu-5 variant of the trypsin inhibitor *Ecballium elaterium* (EETI-II). *Protein Sci 3*:291–302.

Nilges M, Habazettl J, Brünger AT, Holak TA. 1991. Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. *J Mol Biol 219*:499–510.

Olivera BM, Gray WR, Zeikus R, McIntosh JM, Varga J, Rivier J, de Santos V, Cruz LJ. 1985. Peptide neurotoxins from fish-hunting cone snails. *Science 230*:1338–1343.

Orengo CA, Flores TP, Taylor WR, Thornton JM. 1993. Identification and classification of protein fold families. *Protein Eng 6*:485–500.

Pallaghy PK, Duggan BM, Pennington MW, Norton RS. 1993. Three-dimensional structure in solution of the calcium channel blocker ω-conotoxin. *J Mol Biol 234*:405–420.

Pauling L, Corey RB, Branson HR. 1951a. Configurations of polypeptide chains with favoured orientations around single bonds: Two new pleated sheets. *Proc Natl Acad Sci USA 37*:729–740.

Pauling L, Corey RB, Branson HR. 1951b. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA 37*:205–211.

Polanowski A, Cieslar E, Otlewski J, Nienartowicz B, Wilimowska-Pelc A, Wilusz T. 1987. Protein inhibitors from the seeds of Cucurbitaceae plants. *Acta Biochim Pol 34*:395–406.

Rees DC, Lipscomb WN. 1982. Structure of the potato inhibitor complex of carboxypeptidase A at 2.5-Å resolution. *Proc Natl Acad Sci USA 77*:4633–4637.

Richardson M. 1991. Seed storage proteins in plants: The enzyme inhibitors. In: Rogers LJ, ed. *Methods in plant biochemistry, vol 5*. London: Academic Press. pp 259–305.

Sato K, Park NG, Kohno T, Maeda T, Kim JI, Kato R, Takahashi M. 1993. Role of basic residues for the binding of ω-conotoxin GVIA to N-type calcium channels. *Biochem Biophys Res Commun 194*:1292–1296.

Schechter I, Berger A. 1967. On the size of the active site in proteinases. 1. Papain. *Biochem Biophys Res Commun 27*:157–162.

Sevilla P, Bruix M, Santoro J, Gago F, Garcia AG, Rico M. 1993. Three-dimensional structure of ω-conotoxin GVIA determined by ^{1}H NMR. *Biochem Biophys Res Commun 192*:1238–1244.

Skalicky JJ, Metzler WJ, Ciesla DJ, Galdes A, Pardi A. 1993. Solution structure of the calcium channel antagonist ω-conotoxin GVIA. *Protein Sci 2*:1591–1603.

Van Kan JAL, Van den Ackerveken GFJM, De Wit PJGM. 1991. Cloning and characterization of cDNA of avirulence gene *avr9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. *Mol Plant Microb Interact 4*:52–59.

Wilusz T, Wieczorek M, Polanowski A, Denton A, Cook J, Laskowski M Jr. 1983. Amino-acid sequence of two trypsin isoinhibitors (*Cucurbita maxima*). *Hoppe-Seyler's Z Physiol Chem 364*:93–95.

Yu H, Rosen MK, Saccomano NA, Phillips D, Volkmann RA, Schreiber SL. 1993. Sequential assignment and structure determination of spider toxin ω-Aga-IVB. *Biochemistry 32*:13123–13129.