Protein Engineering vol.14 no.9 pp.639–646, 2001

Structural analysis of free and enzyme-bound amaranth α-amylase inhibitor: classification within the knottin fold superfamily and analysis of its functional flexibility

Oliviero Carugo[1,2], Shanyun Lu[3], Jingchu Luo[3], Xiaocheng Gu[3], Songping Liang[4], Stefan Strobl[5] and Sándor Pongor[1,6]

[1]International Centre for Genetic Engineering and Biotechnology, Padriciano 99, Trieste [2]Department of General Chemistry, University of Pavia, Pavia, Italy, [3]Peking University, Beijing, China, [4]Hunan Normal University, Hunan, China and [5]Proteros Biostructures GmbH, Planegg-Martinsried, Germany

[6]To whom correspondence should be addressed.
Email: pongor@icgeb.trieste.it

The three-dimensional structure of the amaranth α-amylase inhibitor (AAI) adopts a knottin fold of *abcabc* topology. Upon binding to α-amylase, it adopts a more compact conformation characterized by an increased number of intramolecular hydrogen bonds, a decreased volume and in addition a *trans* to *cis* isomerization of Pro20. A systematic analysis of the 3-D structural databanks revealed that similar proteins and domains share with AAI the characteristic presence of proline residues, many of which are in a *cis* backbone conformation. As these proteins fulfil a variety of functional roles and are expressed in very different organisms, we conclude that the structure of the knottin fold, including the propensity of the *cis* bond, are the result of convergent evolution.

Keywords: amylase inhibitors/*cis*-prolines/disulfide bridges/knottins/protein structure

Introduction

Knottins or cystine knots are an intensively studied group of proteins characterized by a large number of intramolecular S–S bridges within a chain of typically 35–45 residues (McDonald and Hendrickson, 1993; Murray-Rust *et al.*, 1993; Isaacs, 1995; Sun and Davies, 1995; Bode and Renatus, 1997). The correct formation of the S–S bridges is believed to convey stability to this fold (Hunter and Komives, 1995), which has also been successfully used for pharmaceutical purposes (Miljanich and Ramachandran, 1995), as well as a structural scaffold for the generation of peptide libraries (Christmann *et al.*, 1999).

A major source of interest in disulfide-rich small proteins is their widespread distribution in Nature (e.g. plants, insects, mammals) as well as their extreme functional variability, ranging from enzymes to metal ion channel inhibitors. The knottin fold occurs both in monoglobular proteins (e.g. protease inhibitors or ion channel regulators) and in multidomain proteins where sometimes one finds several subsequent knot motifs within the same chain, such as in thrombomodulin, transforming growth factor or E-selectin (Harvey *et al.*, 1991; Graves *et al.*, 1994; Meininger *et al.*, 1995).

Recently, the three-dimensional solution structure of a knottin protein, the α-amylase inhibitor AAI, has been determined in solution by NMR spectroscopy (Lu *et al.*, 1999) along with its crystal structure in complex with an insect α-amylase (Barbosa Pereira *et al.*, 1999). This amylase inhibitor adopts a typical knottin fold (Figure 1) with an *abcabc* disulfide topology, which means that the first cysteine in the protein sequence forms a disulfide bridge with the fourth, the second with the fifth and the third with the sixth. Several other enzyme inhibitors, mainly protease inhibitors, are known to assume such a compact structure: hirustasin (Mittl *et al.*, 1997; Uson *et al.*, 1999), antistasin (Lopatto *et al.*, 1997), carboxypeptidase inhibitor (Rees and Lipscomb, 1982) and some trypsin inhibitors (Bode *et al.*, 1989; Chiche *et al.*, 1989; Huang *et al.*, 1993). The latter have been shown to resemble structurally several other proteins of different biological functions (Pallaghy *et al.*, 1994).

In this work, we compared the free and complexed structures of the AAI. We found that enzyme binding results in a *cis–trans* isomerization of a peptide bond at Pro20 of AAI. In an attempt to interpret this structural rearrangement, we carried out a systematic structural analysis of structurally related proteins and found that among the proteins most related to AAI, *cis* peptide bonds often appear in similar positions.

Methods

All three-dimensional structural data were taken from the Protein Data Bank (Bernstein *et al.*, 1977). In the case of NMR data, the one ranked first by the authors was selected. Superpositions were performed with the algorithm of Kabsch (Kabsch, 1978) and McLachan (McLachan, 1979) or with the software SUPERIMPOSE (Diederichs, 1995). Cluster analyzes were performed with the method reported by Carugo's group (Carugo, 1995; Carugo and Argos, 1997). Briefly, the proximity matrix, whose elements indicate the similarity between a pair of structures, is subjected to cluster analysis based on a hierarchical agglomerative algorithm and a single linkage similarity criterion between two clusters. Secondary structure definitions were taken from the literature (preferred source), from the PDB files (Bernstein *et al.*, 1977) or were calculated with the DSSP program (Kabsch and Sander, 1983) and validated by visual inspection. Residues forming the protein core were selected as the residues with a fractional solvent-accessible area $<40\%$. Fractional solvent-accessible area values were computed as described by Heringa *et al.* (Heringa *et al.*, 1995), by using 1027 non-homologous protein three-dimensional structures (maximum sequence identity 25%) taken from the PDB_SELECT database (Hobohm and Sander, 1994). For these computations, solvent accessibilities and secondary structures were computed with DSSP (Kabsch and Sander, 1983) and secondary structures were simplified as helical, strand and others.

Results and discussion

Comparison of the free and enzyme-bound α-amylase inhibitor

The backbone conformation of the α-amylase inhibitor in solution and in complex with the *Tenebrio molitor* α-amylase

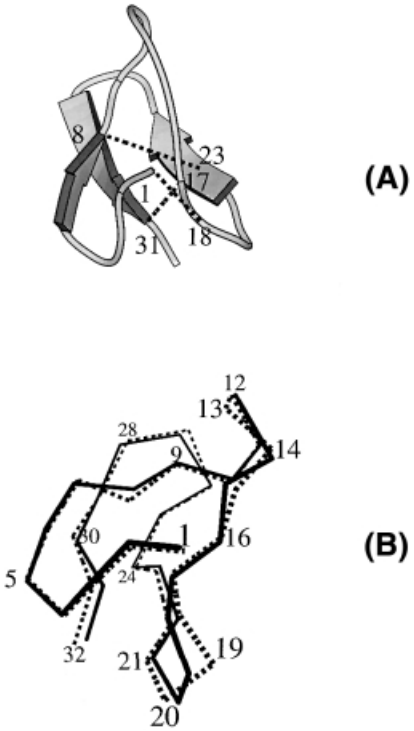

Fig. 1. (**a**) The structure of the amaranth α-amylase inhibitor (AAI). The positions of the disulfide bonds are indicated with dashed lines. (**b**) Superposition of the Cα atoms of the free (1qfd, continuous line) and complexed (1clv, dashed line) amaranth α-amylase inhibitor. The largest discrepancies are observed in loops 18–22 and 11–14 and in the polypeptide fragment in between them. Figure prepared with MOLSCRIPT (Kraulis, 1991).

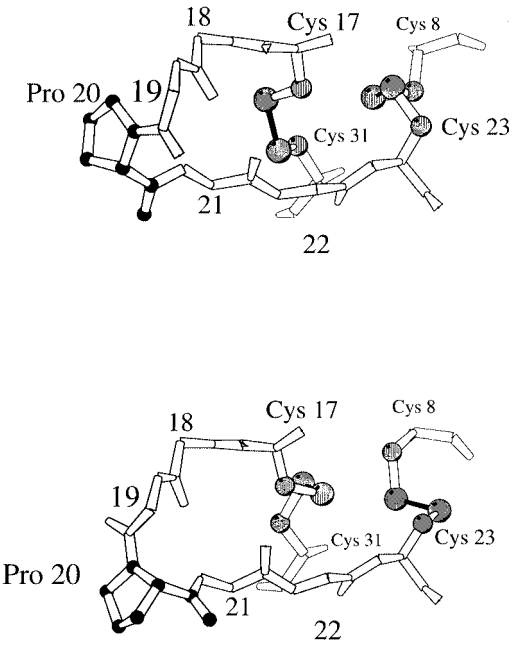

Fig. 2. Comparison of the conformations of the loop centered on Pro20 in the free (up) and complexed (down) amaranth α-amylase inhibitor. Also the conformations of the Cys17–Cys31 and Cys8–Cys 23 S–S bridges change. Figure prepared with MOLSCRIPT (Kraulis, 1991).

Table I. Comparison of the conformations of the disulfide bridges Cys17–Cys31 and Cys8–Cys23 in the free and complexed forms of the amaranth α-amylase inhibitor[a]

	Cys17–Cys31(′)		Cys8–Cys23(′)	
	Free	Complexed	Free	Complexed
N–CA–CB–SG	–165	–72	–122	56
CA–CB–SG–SG′	178	–72	84	83
CB–SG–SG′–CB′	–107	–82	–71	97
SG–SG′–CB′–CA′	162	–69	–50	66
SG′–CB′–CA′–N′	–166	–65	–75	–160
CA–CA′	6.5	5.7	5.3	5.9

[a]The atoms labeled with primes refer to the residue analogously labeled.

is shown in Figure 1. Upon binding to the amylase, the structure of the α-amylase inhibitor becomes more compact: (i) the solvent-accessible area decreases from 2648 to 2406 Å^2, (ii) the volume of the inhibitor decreases from 2909 to 2806 Å^3 and (iii) the number of hydrogen bonds, identified with the WHAT IF program (Vriend, 1990), increases from 12 to 21. Nevertheless, the two structures superpose fairly well (r.m.s.d. between all equivalent Cα = 0.78 Å).

The major difference between complexed and free α-amylase inhibitor is the isomerization of Pro20, which is *trans* in the free and *cis* in the complexed form. As the *cis* peptide bond is more constrained than the *trans* (Weiss *et al.*, 1998; Jabs *et al.*, 1999; Pal and Chakrabarty, 1999), this feature is likely to confer a higher rigidity to the complexed form. As a consequence of the *trans–cis* isomerization, the polypeptide segment 13–22 located between the first and the second β-strands (1–10 and 24–30, respectively) undergoes a major rearrangement. This is shown quantitatively by the fact that if only the Cα of residues of 1–10 and 24–30 are superposed (r.m.s.d. = 0.48 Å), the segment 13–22 shows a remarkable deviation (r.m.s.d. = 1.24 Å) (Figure 1). Furthermore, the two disulfide bridges close to Pro20 undergo substantial conformational rearrangements (Figure 2 and Table I), although this could also reflect the lower accuracy of ^{1}H NMR experiments in determining the S–S stereochemistry (Fletcher *et al.*, 1997). Interestingly, *trans–cis*-proline isomerization upon enzyme binding has been observed in another knottin-like protein, the protease inhibitor hirustasin (Mittl *et al.*, 1997; Uson *et al.*, 1999).

The *trans–cis* isomerization of the backbone is accompanied by a large number of side-chain reorientations that account for most of the observed differences between complexed and free α-amylase inhibitor. About 80% of the decrease of the solvent-accessible area is due to the residues which are known to interact with α-amylase (Barbosa Pereira *et al.*, 1999). In the complexed α-amylase inhibitor, the Trp5 sidechain rotates by about 30° around both χ1 and χ2 and approaches the Lys4 side chain, which also adopts a different conformation so that it finds itself sandwiched in between the aromatic rings of Tyr21 and Trp5 (Figure 3). Notably, the Tyr21 side-chain moves as a consequence of the *cis–trans* isomerization of Pro20. Similarily, rotations of about 90° around χ2 of Lys11 and of about 30° around χ1 of Tyr27 put these two side chains in closer contact in the complexed form (Figure 4). Upon binding to α-amylase, hydrogen bonds become possible between the ε-amino groups of Lys4 and Lys11 on the one hand and the OH atoms of Tyr21 and Tyr27 on the other. A similar phenomenon, close stacking of Lys on aromatic groups and electrostatic interactions of the ε-amino group, has been observed in the recognition of NADP by various proteins (Carugo and Argos, 1997), so we believe that this feature is

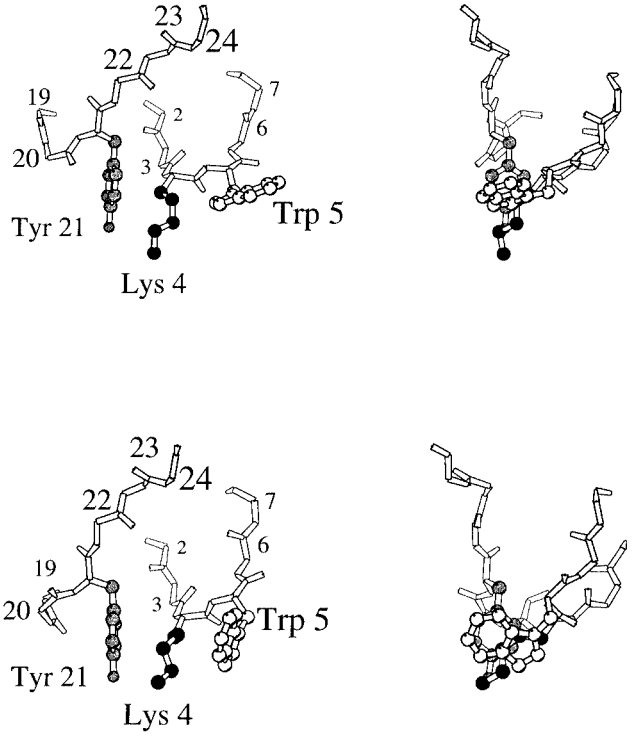

Fig. 3. Comparison of the conformations of Lys4, Trp5 and Tyr21 in the free (up) and complexed (down) amaranth α-amylase inhibitor. Trp5 moves down when the inhibitor interacts with α-amylase and Lys4 results sandwiched between Tyr21 and Trp5. In each case, two views perpendicular to each other are given. Figure prepared with MOLSCRIPT (Kraulis, 1991).

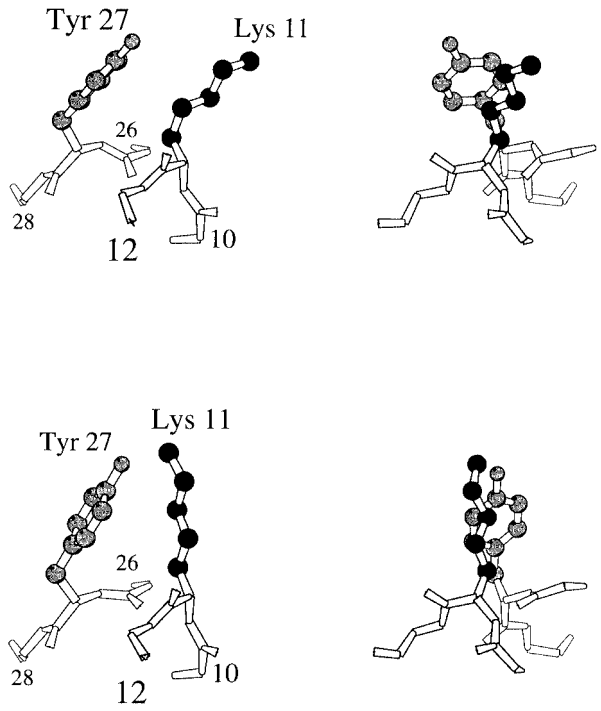

Fig. 4. Comparison of the conformations of Tyr27 and Lys11 in free (up) and complexed (down) amaranth α-amylase inhibitor. In each case, two views perpendicular to each other are given. Figure prepared with MOLSCRIPT (Kraulis, 1991).

likely to contribute to the stability of the complexed α-amylase inhibitor. Residues with backbone or side chains reoriented by more than 30° are listed in Table II.

Structural homologues of α-amylase inhibitor

In a search for similar structures, we analyzed all proteins of the Protein Data Bank (Bernstein *et al.*, 1977) that have, at least three disulfide bridges in an *abcabc* topology which is identical with the amaranth α-amylase inhibitor. In structures with more than three disulfide bonds, the topology of all possible subsets of three S–S bridges was examined. The retrieved data were then inspected visually in order to eliminate redundancies. From the structures of identical chain(s) and identical crystal space groups, the one with the best crystallographic resolution was retained. Single point mutants were rejected, while both the crystallographic and the NMR structures of the same protein were considered, where appropriate, since it is not obvious that structures in different physicochemical phases are identical. This analysis yielded a data set of 84 structures of domains with *abcabc* topology.

The similarity between two structures within this data set was then estimated as the root-mean-square distance between the aligned cysteine atoms computed after their optimal superposition. An alternative distance measure based on the superposition of all the Cα atoms, was also computed. Since the two approaches resulted in analogous results, those given by the first strategy are reported and commented on here. A cluster analysis was performed on a proximity matrix in which each element x_{ij} is an r.m.s.d. value between the structures i and j calculated in the above-indicated manner. The dendrogram resulting from the cluster analysis is shown in Figure 5. The r.m.s.d. value corresponding to the optimal number of partitions is about 1.6–1.8 Å since the plot of the number of clusters versus the threshold of similarity, under which two clusters merge, was found to show a clear edge around these values (data not shown) (Malinowski, 1991)

Schemes of the secondary structural elements together with the location of the disulfide bonds are shown in Figure 6 for the clusters that include at least two structures. Notably, the secondary structural assignments are sometimes rather ambiguous and C- or N-terminal extensions may be present. The fold of the underlying proteins can be roughly pictured as two large sequentially adjacent loops. Each loop consists of either a pair of antiparallel β-sheets or by a β-strand followed by an antiparallel helix; in some cases the loops have no regular secondary structure. The antiparallel three-stranded β-sheet seen in the structure of the AAI protein is a frequent motif (clusters 2, 5 and 6). In the structures containing these motifs the first large loop consists of a β-strand followed by an antiparallel polypeptide segment (helical in clusters 2 and 6). The second large loop is constituted by a pair of antiparallel β-strands, the second of which contacts the β-strand of the first large loop. The most relevant difference between the members of clusters 2, 5 and 6 is thus the three-dimensional location of the half cystines. The structures of cluster 3 adopt a fold similar to the four-helix bundle whereas those of cluster 4 basically lack any regular secondary structure. The members of clusters 1 and 8 partially lack the second large loop. Different from all the other structures are the members of cluster 7, which are made by two domains, one of which, constituted by three polypeptide segments, folds in an irregular four-stranded antiparallel β-sheet stabilized by three disulfide bonds.

The amaranth α-amylase inhibitor results are classified in cluster 5 with 15 other disulfide-rich small proteins (Table III).

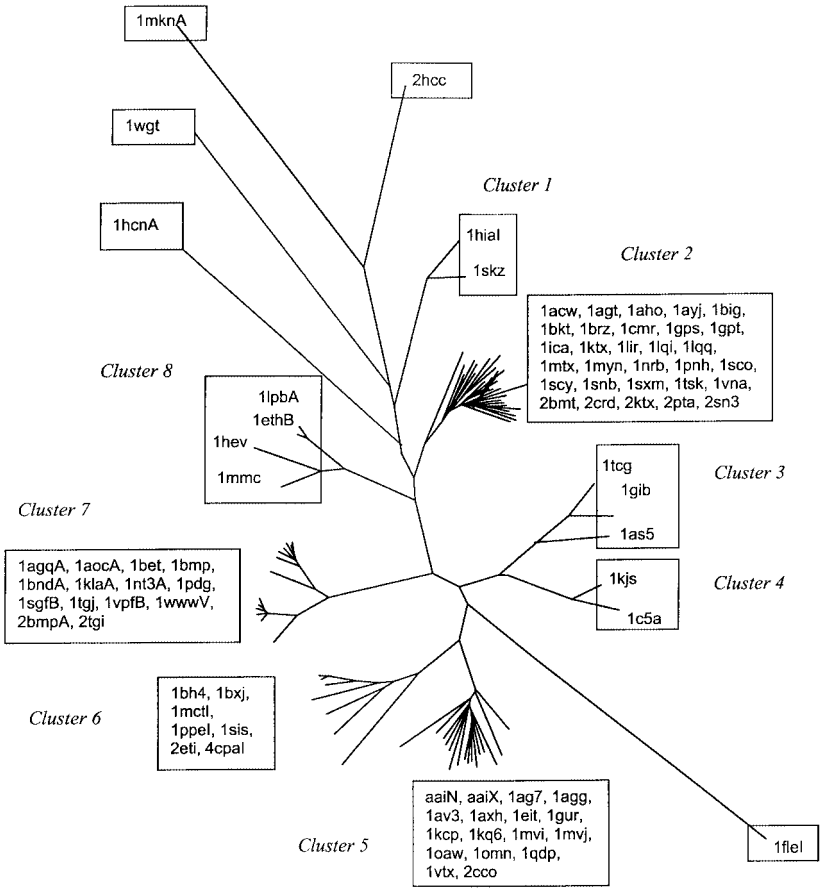

Fig. 5. Dendrogram showing the results of the cluster analysis of the geometrical properties of the knottins. For each entry, the lower-case four-letter Protein Data Bank (Bernstein *et al.*, 1977) identification code is given together with an optional upper-case chain identifier. The same protein was considered separately if complexed with different partners and if structurally characterized in different physicochemical phases. The protein structures are as follows (biological source in parentheses): aaiN, α-amylase inhibitor (*Amaranthus hypochondriacus*); aaiX, α-amylase inhibitor (*Amaranthus hypochondriacus*); 1acw, peptide specific for apamin-sensitive potassium channel (*Androctonus mauretanicus*); 1ag7, conotoxin GS (*Conus geographus*); 1agg, ω-agatoxin IVB (*Agelenopsis aperta*); 1agqA, neurotrophic factor (*Rattus norvegicus*); 1agt, agitoxin 2 (*Leiurus quinquestriatus hebraeus*); 1aho, toxin II (*Androctonus australis* Hector); 1aocA, coagulogen (*Tachpleus tridentatus*); 1as5, conotoxin Y-PIIIE (*Conus purpurascens*); 1ayj, antifungal protein 1 (*Raphanus sativus*); 1av3, κ-conotoxin PVIIA (*Conus purpurascens*); 1axh, atracotoxin HVI (*Hadronyche versuta*); 1bet, β-nerve growth factor (*Mus musculus*); 1bh4, circulin A (*Chassalia parviflora*); 1big, toxin BMTX1 (*Buthus martensii* Karsch); 1bkt, toxin BMKTX (*Buthus martensii* Karsch); 1bmp, bone morphogenetic protein 7 (*Homo sapiens*); 1bndA, brain derived neurotrophic factor (*Homo sapiens*); 1brz, brazein (*Pentadiplandra brazzeana*); 1bxj, trypsin inhibitor (*Cucurbita maxima*); 1c5a, des-Arg-complement factor (*Sus scrofa*); 1cmr, chymeric charibdotoxin (*Leiurus quinquestriatus hebraeus*); 1eit, μ-agatoxin I (*Agelenopsis aperta*); 1eth, colipase (*Sus scrofa*); 1fleI, elafin (*Sus scrofa*); 1gib, μ-conotoxin GIIIB (*Conus geographus*); 1gps, γ-1-P-thionein (*Buthus martensii*); 1gpt, γ-1-H-thionein (*Horedum vulgare*); 1gur, gurmarin (*Gymnema sylvestrae*); 1hcn, chorionic gonadotropin (*Homo sapiens*); 1hev, hevein (*Hevea brasiliensis*); 1hiaI, hirustasin (*Sus scrofa*); 1ica, insect defensin A (*Phormia terranovae*); 1kcp, κ-conotoxin PVIIA (*Conus purpurascens*); 1kjs, cell adesion protein 5 (*Homo sapiens*); 1klaA, transforming growth factor β1 (*Homo sapiens*); 1qk6, huwentoxin-I (*Selenocosmia huwena*); 1ktx, potassium channel inhibitor (*Androctonus mauretanicus*); 1lir, toxin LQ2 (*Leiurus quinquestriatus hebraeus*); 1lpb, colipase (*Homo sapiens*); 1lqi, insct toxin α (*Leiurus quinquestriatus hebraeus*); 1lqq, insect toxin LQQIII (*Leiurus quinquestriatus*); 1mctI, trypsin inhibitor (*Momorica charantia*); 1mkn, N-terminal half of midkin (*Homo sapiens*); 1mmc, antimicrobial peptide 2 (*Amaranthus causatus*); 1mtx, margatoxin (*Centruroides margaritatus*); 1mvi, ω-conotoxin MVIIA (*Conus magus*); 1mvj, ω-conotoxin SVIB (*Conus striatus*); 1myn, drosomycin (*Drosophila melanogaster*); 1nrb, neurotoxin V (*Centruroides sculpturatus* Ewing); 1nt3A, neurotrophin-3 (*Homo sapiens*); 1oaw, ω-agatoxin IVA (*Agelenopsis aperta*); 1omn, ω-conotoxin MVIIC (*Conus magus*); 1pdg, platelet-derived growth factor BB (*Saccharomyces cervisiae*); 1pnh, toxin for apamin-sensitive potassium channels (*Androctonus mauretanicus*); 1ppeI, trypsin inhibitor (*Cucurbita maxima*); 1qdp, robustoxin (*Atrax robustus*); 1sco, toxin OSK1 (*Orthochirus scrobiculosus*); 1scy, scyllatoxin (*Leiurus quinquestriatus hebraeus*); 1sgfB, β-nerve growth factor (*Mus musculus*); 1sis, insectotoxin (*Buthus eupeus*); 1skz, antistasin (*Heamenteria officinalis*); 1snb, neurotoxin BMKM8 (*Buthus martensii* Karsch); 1sxm, nexiustoxin (*Centruroides noxius* Hofmann); 1tcg, m-conotoxin GIIIA (*Conus geographus*); 1tgj, transforming growth factor β 3 (*Homo sapiens*); 1tsk, toxin active on small conductance potassium channels (*Tityus serrulatus*); 1vna, neurotoxin (*Centruroides sculpturatus* Ewing); 1vpfB, endothelial growth factor (*Homo sapiens*); 1vtx, δ-atracotoxin HVI (*Hadronyche versuta*); 1wgt, agglutinin (*Triticum vulgais*); 1wwwV, nerve growth factor (*Homo sapiens*); 2bmpA, bone morphogenetic protein s (*Homo sapiens*); 2bmt, toxin BMTX2 (*Buthus martensii*); 2cco, ω-conotoxin GVIA (*Conus geographus*); 2crd, charybdotoxin (*Leiurus quinquestriatus hebraeus*); 2eti, trypsin inhibitor (*Ecballium elaterium*); 2hcc, chemokine HCC-s (*Homo sapiens*); 2ktx, kaliotoxin (*Androctonus mauretanicus*); 2pta, toxin K-A (*Pandius imperator*); 2sn3, neurotoxin (*Centruroides sculpturatus* Ewing); 2tgi, transforming growth factor β 2 (*Homo sapiens*); 4cpaI, carboxypeptidase A inhibitor (Russet–Burbank potatoes).

Table II. List of the residues of the amaranth α-amylase inhibitor which experience rotations larger than 30° of their dihedral angles after complexation of the substrate

Main-chain reorientation only	Pro20
Side-chain reorientation only	Ile2, Lys4, Trp5, Asn6, Arg7, Pro10, Lys11, Met12, Cys17, Cys23, Thr24, Ser25, Asp26, Tyr27, Asn30
Main- and side-chain reorientation	Cys8, Asp13, Val15, Glu19, Tyr21, Cys31

Table III. List of the structures classified together with the amaranth α-amylase inhibitor by cluster analysis (cluster 5 in Figure 5)[a]

PDB code	Protein name	Organism name	Biological role	Ref.[b]
aaiX	α-Amylase inhibitor	*Amaranthus hypochondriacus*	Enzyme inhibitor	1
1ag7	Conotoxin GS	*Conus geographus*	Ion channel inhibitor	2
1agg	ω-Agatoxin IVB	*Agelenopsis aperta*	Ion channel inhibitor	3
1av3	κ-Conotoxin PVIIA	*Conus purpurascens*	Ion channel inhibitor	4
1axh	Atracotoxin HVI	*Hadronyche versuta*	Ion channel inhibitor	5
1eit	μ-Agatoxin	*Agelenopsis aperta*	Ion channel inhibitor	6
1gur	Gurmarin	*Gymnema sylvestrae*	Unknown	7
1kcp	κ-Conotoxin PVIIA	*Conus purpurascens*	Ion channel inhibitor	8
1mvi	ω-Conotoxin MVIIA	*Conus magus*	Ion channel inhibitor	9
1mvj	ω-Conotoxin SVIB	*Conus striatus*	Ion channel inhibitor	9
1oaw	ω-Agatoxin IVA	*Agelenopsis aperta*	Ion channel inhibitor	10
1omn	ω-Conotoxin MVIIC	*Conus magus*	Ion channel inhibitor	11
1qdp	Rubustoxin	*Atrax robustus*	Ion channel inhibitor	12
1qk6	Huwentoxin-I	*Selenocosmia huwena*	Acetylcholine receptor inhibitor	13
1vtx	δ-Atracotoxin HVI	*Hadronyche versuta*	Ion channel inhibitor	14
2cco	ω-Conotoxin GVIA	*Conus geographus*	Ion channel inhibitor	15

[a]All the 3-D structures were determined by NMR spectroscopy.
[b](1) Chagolla-Lopez *et al.*, 1994; (2) Hill *et al.*, 1997; (3) Regan, 1991 (4) Scanlon *et al.*, 1997; (5) Atkinson *et al.*; 1993; Fletcher *et al.*, 1997; (6) Skinner *et al.*, 1989; (7) Imoto *et al.*, 1991; Fletcher *et al.*, 1999; Katsukawa *et al.*, 1999; (8) Savarin *et al.*, 1997; Scanlon *et al.*, 1997; (9) Nielsen *et al.*, 1996; (10) Mintz *et al.*, 1992; (11) Farr-Jones *et al.*, 1995; (12) Mylecharane *et al.*, 1989; (13) Zhou *et al.*, 1997; (14) Mylecharane *et al.*, 1989; Nicholson *et al.*, 1994; (15) Kim *et al.*, 1994; Nadasdi *et al.*, 1995; Lew *et al.*, 1997.

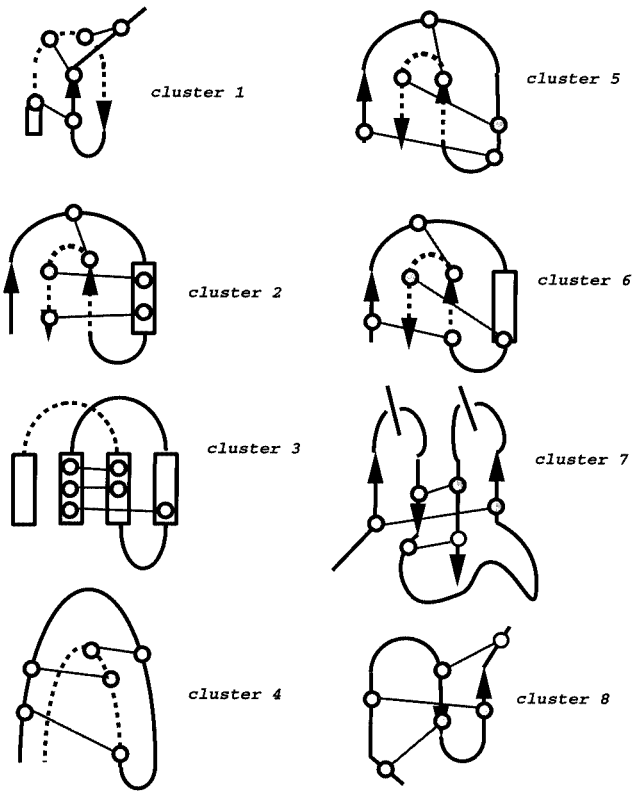

Fig. 6. Schematic representations of the knottins grouped in clusters 1–8. Helices are shown by rectangles, strands by arrows, half cystine positions by spheres and disulfide bonds by lines. Dashed and continous lines are used to indicated objects closer to (continuous lines) or far from (dashed lines) the observer.

We considered these structures as structural homologues of AAI and the following analysis is restricted mainly to these proteins.

Functional features and the molecular surface of α-amylase inhibitor and its structural homologues (cluster 5)

Members of cluster 5 have a variety of origins and biological roles, but all of them seem to be the ligands, most often

Table IV. Percentage of three-dimensional conservation of the position of residues thought to be important for activity[a]

Residue	Main	Side
Cys1	27	27
Lys4	20	13
Trp5	20	33
Asn6	7	7
Arg7	13	13
Asp13	13	13
Thr24	0	7
Ser25	7	13
Asp26	13	13
Tyr27	7	27
Tyr28	27	27
Asn30	13	13
Ser32	0	0

[a]Main and side indicate the position conservation of the main and the side chain, respectively. Only the residues known to be important for the recognition of α-amylase by its inhibitor are considered. The percentage of conservation indicates the fraction of cases in which a functionally relevant residue of an other protein occupies the same position of an AAI residue after optimal superposition of the Cα atoms.

inhibitors, of other proteins, such as voltage-sensitive ion channels and enzymes. They are found in venom of snails and spiders and in plant tissues. The most typical representatives of cluster 5 are various conotoxins, toxins present in the venom of piscivorous marine snails which interact selectively with the various types of voltage-sensitive calcium, sodium and potassium channels, either in neurons or in skeletal muscles (Catterall, 1988;Gray *et al.*, 1988; McCleskey *et al.*, 1988; Olivera *et al.*, 1988; Yanagawa *et al.*, 1988; Terlau *et al.*, 1996). They have pharmacological applications (Miljanich and Ramachandran, 1995) and have been used to characterize membrane channel assemblies (Tsien *et al.*, 1991; Olivera *et al.*, 1994). Cluster 5 also contains several neurotoxins present in the venom of spiders. Huwentoxin was shown to interact with the nicotinic acetylcholine receptor (Zhou *et al.*, 1997). ω-Atracotoxin-HV1 (Atkinson *et al.*, 1993), ω-agatoxin IVA and ω-agatoxin IVB interact with voltage-sensitive

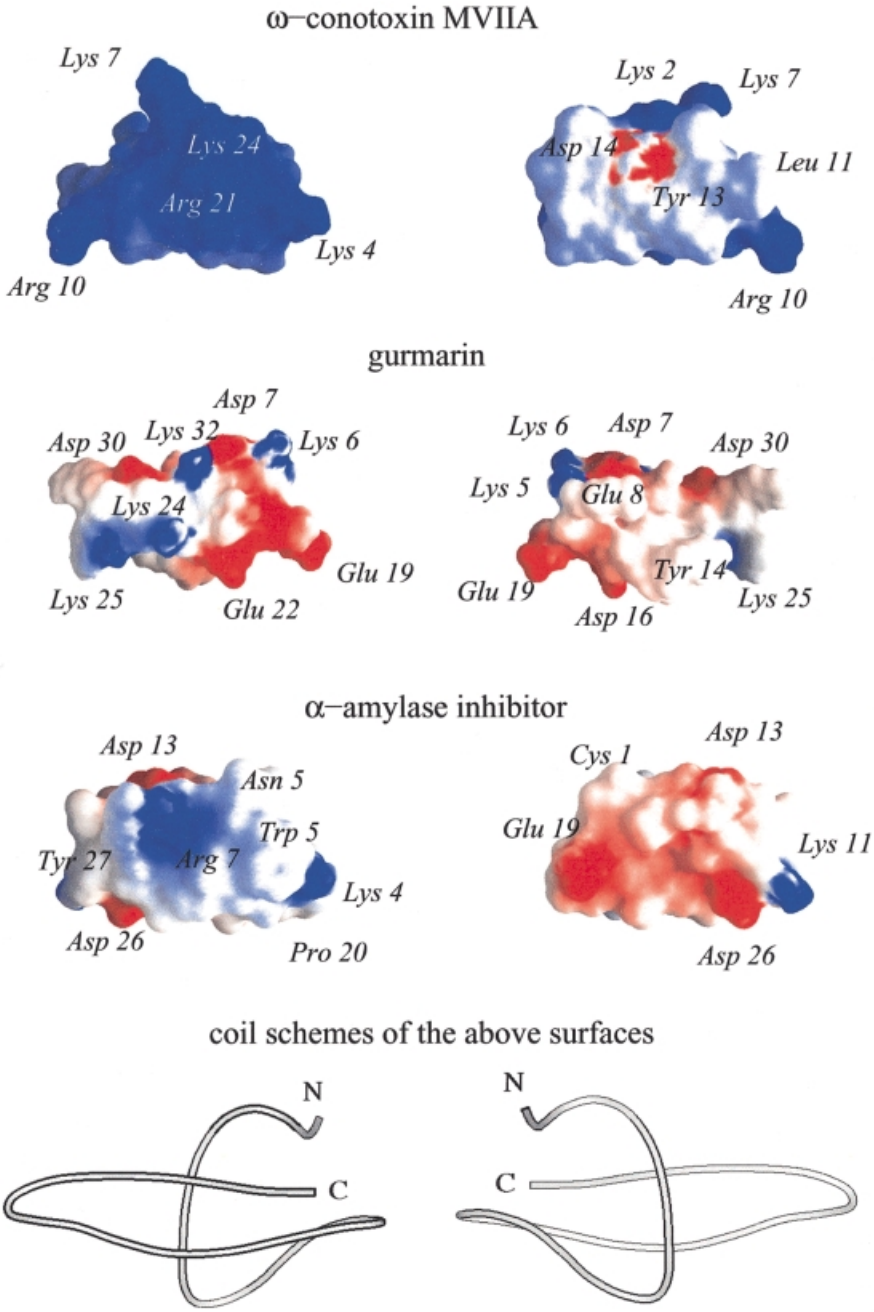

Fig. 7. Comparison of the surface electric potentials of some proteins similar to the amaranth α-amylase inhibitor. Positively and negatively charged surface patches are colored in blue and red, respectively, while neutral regions are white. Figure prepared with GRASP (Nicholls *et al.*, 1991).

calcium-channels (Regan, 1991; Mintz *et al.*, 1992). Voltage-sensitive sodium channels interact with δ-atracotoxin-HV1 (Nicholson *et al.*, 1994), robustoxin (Mylecharane *et al.*, 1989) and μ-agatoxin-I (Skinner *et al.*, 1989). Interestingly, some of these toxic polypeptides have been shown to possess remarkable specificity. For example, ω-atracotoxin-HV1 acts on insects but not on mammalian neuronal calcium channels (Fletcher *et al.*, 1997). δ-Atracotoxin-HV1 is severely toxic towards newborn mice and primates but not other vertebrates (Mylecharane *et al.*, 1989; Nicholson *et al.*, 1994).

A further member of cluster 5 is gurmarin, a protein extracted from leaves of *Gymnema sylvestre*, an indian plant whose leaves were chewed as a folkloric treatment for diabetes mellitus (Imoto *et al.*, 1991). Gurmarin is known to suppress specifically the sweet taste sensation in rats (but not in humans) (Katsukawa *et al.*, 1999). It is thermostable and supports both high pH and urea concentrations, but its biological role is unknown at a molecular level. It has recently been reported that gurmarin has no effect on several voltage-sensitive ion channels (Fletcher *et al.*, 1999).

In the proteins of cluster 5, the residues important for partner recognition and activity have been convincingly identified only in very few cases. Residues important for activity have been identified by site-directed mutagenesis studies on ω-conotoxin GVIA (Kim *et al.*,1994; Nadasdi *et al.*, 1995; Lew *et al.*, 1997). Nevertheless, in most cases, the relative importance of various residues is merely speculative. The action mechanism of gurmarin is unknown. At least two hypotheses have been

Table V. Total length of the protein (L) and lengths of the three β-strands (β1, β2 and β3) and of the polypeptide segments preceding the first β-strand (l1), following the third β-strand (l4), intercalated between the first and second β-strands (l2) and intercalated between the second and third β-strands (l3)

PDB	L	l1	β1	l2	β2	l3	β3	l4	
2cco	27	5	3	9	3	3	4	0	
1omn	26	5	3	11	3	2	3	0	
1mvi	25	5	3	11	2	1	3	0	
1mvj	26	5	3	11	2	1	3	0	
1kcp	27	5	3	10	3	2	4	0	
1ag7	34	6	3	7	4	5	4	5	
1qk6	33	6	3	11	3	3	4	3	
1axh	37	9	1	11	5	5	5	1	
1oaw	48	9	1	13	4	6	4	11	
1av3	27	5	3	10	3	2	4	0	(also 1kcp)
1vtx	42	5	2	10	4	8	4	9	
1qdp	42	5	3	10	3	7	4	10	
1agg	48	10	1	13	3	6	3	12	
1eit	36	5	1	13	4	7	4	2	
1gur	35	7	4	10	5	3	5	1	
aaiX	32	5	4	11	3	5	3	1	

Table VI. Positions of the six cysteines involved in the three disulfide bridges[a]

PDB	Positions
2cco	Y2CX7CX6CX0CX2CX6CY1
1omn	Y0CX6CX6CX0CX3CX5CY0
1mvi	Y0CX6CX6CX0CX3CX4CY0
1mvj	Y0CX6CX6CX0CX3CX5CY0
1kcp	Y0CX6CX6CX0CX3CX5CY1
1ag7	Y1CX6CX3CX0CX5CX7CY7
1qk6	Y1CX6CX6CX0CX4CX6CY4
1axh	Y3CX6CX5CX0CX3CX13CY1
1oaw	Y3CX7CX6CX0CX4CX10CY12
1av3	Y0CX6CX6CX0CX3CX5CY1 (also 1kcp)
1vtx	Y0CX6CX6CX0CX4CX10CY10
1qdp	Y0CX6CX5CX0CX4CX10CY10
1agg	Y3CX7CX6CX0CX4CX10CY12
1eit	Y1CX6CX6CX0CX4CX9CY4
1gur	Y2CX7CX6CX0CX5CX9CY2
aaiX	Y0CX6CX8CX0CX4CX7CY1

[a]The Y*n*CX*n*CX*n*CX*n*CX*n*CY*n* nomenclature is adopted (Pallaghy *et al.*, 1994). Y*n* indicates that *n* residues precede or follow the first and sixth cysteine and X*n* indicates that *n* residues separate two sequentially subsequent cysteines.

proposed for the action mechanism of κ-conotoxin PVIIA (Savarin *et al.*, 1997; Scanlon *et al.*, 1997).

In an attempt to detect possible similarities in the three-dimensional arrangement of residues important for activity, all structures of this group were superposed on that of the complexed amaranth α-amylase inhibitor, by considering the Cα atoms. Care was taken to treat separately the main and side chains, since it is not necessarily true that the main and side chains of a given residue of the α-amylase inhibitor best superpose the main and side chains of the same residue in another structure. The results (see Table IV) clearly indicate that there is little three-dimensional conservation among the members of this group. For example, some important residues of the α-amylase inhibitor never superpose on residues important for activity in other proteins of cluster 5. The mean three-dimensional position conservation is only 11 and 14% for the main and side chains, respectively.

Also, there is little similarity between the electrostatic potential of the proteins which was computed with the GRASP program (Nicholls *et al.*, 1991). For example, ω-conotoxin MVIIA has one face positively charged and the opposite face substantially neutral, and gurmarin and the α-amylase inhibitor have spread surface patches of opposite charge on both sides (Figure 7).

Structural features shared by α-amylase inhibitor and its structural homologues (cluster 5)

The members of cluster 5 fold with a triple-stranded β-sheet stabilized by three disulfide bonds (Figure 1). The lengths of the chains and of the three β strands and also the length of the polypeptide segments separating the six cysteines are variable (Tables V and VI). The protein core is formed by only 6–20% of the residues. The mean fractional solvent accessibility is very high, ranging from 1.22 to 1.56. For comparison, in proteins of 100–200 amino acids, the core comprises about 40% of the residues and the mean fractional solvent accessibility is usually around 0.70. It is therefore not surprising that these small proteins, despite the presence of three disulfide bridges, experience a high mobility. In ω-conotoxin MVIIC, for example, there are only three slowly exchanging amide protons (Farr-Jones *et al.*, 1995). The flexibility is not restricted to the loops. In some cases, such as in κ-conotoxin PVIIA (Savarin *et al.*, 1997) and conotoxin GS (Hill *et al.*, 1997), even the disulfide bridges have been observed in multiple conformations. We mention that that some of these findings could reflect the difficulty of defining disulfide bond geometries from ^{1}H NMR data (Fletcher *et al.*, 1997).

Other common features of the members of cluster 5 are (i) the presence of other post-translational modifications different from S–S bonds, (ii) the large number of prolines and (iii) the high frequency of *cis*-prolines:

(i) Post-translationally modified hydroxyprolines are observed, for example, in ω-conotoxin GVIA, κ-conotoxin PVIIA and conotoxin GS; in the last structure, also γ-carboxyglutamic acid substitutes for Glu.

(ii) Only two members of cluster 5, ω-conotoxins MVIIA and SVIB, lack Pro residues, whereas in three cases, ω-conotoxin GVIA, ω-atracotoxin-HV1 and α-amylase inhibitor, prolines represent >10% of the residues. This contrasts with the fact that prolines account for only about 5% of the residues in a set of 1027 non-homologous protein structures taken from the PDB_SELECT database (Hobohm and Sander, 1994).

(iii) Nearly half of the structures of cluster 5 contains at least one *cis*-proline, all the prolines of conotoxin GS are *cis* and about one-third of all the prolines present in the structures of cluster 5 are *cis*, a much higher fraction than usually observed. It was recently reported (Weiss *et al.*, 1998) that about 5% of the X–Pro peptide bonds are *cis* in protein three-dimensional structures, contrary to the 30% that would be expected on the basis of the different stabilities of the *cis* and *trans* isomers.

Since there is no reason to suppose that the structures of cluster 5 are more accurate than other protein three-dimensional structures, it must be concluded that the high frequency of *cis*-prolines together with the high proline content is an essential feature of these proteins. It has been shown that *cis* peptide bonds restrict the conformational space available (Jabs

et al., 1999) and it therefore reasonable that structures of cluster 5 employ *cis*-prolines to stabilize their fold.

All the three features described above discriminate the members of cluster 5 from the other knottin domains. Post-translational modifications are less frequently observed within the other knottins, only the domains of clusters 3 and 7 have an unusual high frequency of prolines and only in cluster 3 these residues quite often adopt a *cis* backbone conformation.

Conclusions

Upon binding to α-amylase, the AAI molecule adopts a compact conformation which is characterized by, among others, a *trans* to *cis* isomerization of Pro20. This isomerization is likely to imply subtle but relevant consequences: (i) the complexed inhibitor becomes conformationally more constrained; (ii) two of the three disulfide bridges are forced to adopt different conformations relative to the uncomplexed inhibitor; and (iii) the consequent displacement of some side chains is likely to favor the reorientation of other, adjacent side chains, which thus become optimally oriented towards the amylase sites that they must recognize.

A systematic analysis of the 3-D structure databank revealed several structural clusters among proteins and domains that share the *abcabc* disulfide topology of AAI. Interestingly, the proteins that cluster together with AAI have a variety of evolutionary origins, but the same as AAI they have a relatively high proline content and many of them contain *cis*-proline residues. We therefore conclude that the *cis*-Pro may be a structurally important feature of this group of proteins.

The structural comparison of the knottin proteins revealed large variations among the members of this group. Little conservation is seen in terms of surface electrostatics and among the functionally important residues. As a consequence, it appears that a common evolutionary origin cannot be suggested from these data. In other terms, the knottin fold in general may have emerged as the result of convergent evolution.

References

Atkinson,R.K., Howden,M.E.H., Tyler,M.I. and Vonarx,E.J. (1993) *International Patent Application*, ID 34443/93, 1993.

Barbosa Pereira,P.J., Lozanov,V., Patthy,A., Huber,R., Bode,W., Pongor,S. and Strobl,S. (1999) *Structure*, **7**, 1079–1088.

Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.E.,Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Bode,W. and Renatus,M. (1997). *Curr. Opin. Struct. Biol*., **7**, 865–872.

Bode,W., Greyling,H.J., Huber,R., Otlewski,J. and Wilusz,T. (1989) *FEBS Lett.*, **242**, 285–290.

Carugo,O. (1995) *Acta Crystallogr*., **B51**, 314–328.

Carugo,O. and Argos,P. (1997) *Proteins*, **28**, 29–40.

Catterall,W.A. (1988) *Science*, **242**, 50–61.

Chagolla-Lopez,A., Blanco-Labra,A., Patthy,A., Sanchez,R. and Pongor,S. (1994) *J. Biol. Chem.*, **269**, 23675–23680.

Chiche,L., Gaboriaud,C., Heitz,A., Mornon,J.-P., Castro,B. and Kollman,P.A. (1989) *Proteins*, **6**, 405–412.

Christmann,A., Walter,K., Wentzel,A., Kratzner,R. and Kolmar,H. (1999) *Protein Eng*., **12**, 797–806.

Diederichs,K. (1995) *Proteins*, **23**, 187–195.

Farr-Jones,S., Miljanich,G.P., Nadasdi,L., Ramachandran,J. and Basus,V.J. (1995) *J. Mol. Biol.*, **248**, 106–124.

Fletcher,J.I., Smith,R., O'Donoghue,S.I., Nilges,M., Connor,M., Howden,M.E.H., Christie,M.J. and King,G.F. (1997) *Nature Struct. Biol*., **4**, 559–566.

Fletcher,J.I., Dingley,A.J., Smith,R., Connor,M.J. and King,G.F. (1999) *Eur. J. Biochem*., **264**, 525–533.

Graves,B.J., Crowther,R.L., Chandran,C., Rumberger,J.M., Li,S., Huang,K.S., Presky,D.H., Familletti,P.C., Wolitsky,B.A. and Burns,D.K. (1994) *Nature*, **367**, 532–538.

Gray,W.R., Olivera,B.M. and Cruz,L.J. (1988) *Annu. Rev. Biochem.*, **57**, 665–700.

Harvey,T.S., Wilkinson,A.J., Tappin,M.J., Cooke,R.M. and Campbell,I.D. (1991). *Eur. J. Biochem.*, **198**, 555–562.

Heringa,J., Argos,P., Egmon,M.R. and de Vlieg,J. (1995) *Protein Eng*., **8**, 21–30.

Hill,J.M., Alewood,P.F. and Craik,D.J. (1997) *Structure*, **5**, 571–583.

Hobohm,U. and Sander,C. (1994) *Protein Sci*., **3**, 522–524.

Huang,Q., Liu,S. and Tang,Y. (1993) *J. Mol. Biol.*, **229**, 1022–1033.

Hunter,M.J. and Komives,E.A. (1995) *Protein Sci.*, **4**, 2129–2137.

Imoto,T., Mihasaka,A., Ishima,R. and Akasaka,K. (1991) *Comp. Biochem. Physiol. A*, **100**, 309–314.

Isaacs,N.W. (1995) *Curr. Opin. Struct. Biol.*, **5**, 391–395.

Jabs,A., Weiss,M.S. and Hilgenfedl,R. (1999) *J. Mol. Biol*., **286**, 291–304.

Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.

Kabsch,W. (1978) *Acta Crystallogr*., **A34**: 827–828.

Katsukawa,H., Imoto,T. and Ninomiya,Y. (1999) *Chem Senses*, **24**, 387–392.

Kim,J.I., Takashashi,M., Ogura,A., Kohno,T., Kudo,Y. and Sato K. (1994) *J. Biol. Chem*., **269**, 23876–23878.

Kraulis,P.J. (1991) *J. Appl. Cryst*., **24**, 946–950.

Lew,M.J., Flinn,J.P., Pallaghy,P.K., Murphy,R., Whorlow,S.L., Wright,C.E., Norton,R.S. and Angus,J.A. (1997) *J. Biol. Chem*., **272**, 12014–12023.

Lopatto *et al*. (1997) *EMBO J*. **16**, 5151–5161.

Lu,S. *et al.* (1999) *J. Biol. Chem.*, **274**, 20473–20478.

Malinowski,E.R. (1991). *Factor Analysis in Chemistry*. John Wiley, New York.

McCleskey,E.W., Fox,A.P., Feldman,D.H., Cruz,L.J., Olivera,B.M., Tsien,R.W. and Yoshikami,D. (1988) *Proc. Natl Acad. Sci. USA*, **84**, 4327–4331.

McDonald,N.Q. and Hendrickson,W.A. (1993) *Cell*, **73**, 421–424.

McLachan,A.D. (1979) *J. Mol. Biol*., **128**, 48–67.

Meininger,D.P., Hunter,M.J. and Komives,E.A. (1995) *Protein Sci.*, **4**, 1683–1695.

Miljanich,G.P. and Ramachandran,J. (1995) *Annu. Rev. Pharmacol. Toxicol.*, **35**, 707–734.

Mintz,I.M., Venema,V.J., Swiderek,K.M., Lee,T.D., Bean,B.P. and Adams,M.E. (1992) *Nature*, **355**, 827–829.

Mittl,P.R., Di Marco,S., Fendrich,G., Pohlig,G., Heim,J.,Sommerhoff,C., Fritz,H., Priestle,J.P. and Gruetter,M.G. (1997) *Structure*, **5**, 253–264.

Murray-Rust,J., McDonald,N.Q., Blundell,T.L., Hosang,M., Oefner,C., Winkler,F. and Bradshaw,R.A. (1993). *Structure*, **1**, 153–159.

Mylecharane,E.J., Spence,I., Sceumack,D.D., Claassens,R. and Howden,M.E.H. (1989) *Toxicon*, **27**, 481–492.

Nadasdi,L., Yamashiro,D., Chung,D., Tarczyhornoch,K., Adrjanssen,P. and Ramachandran,J. (1995) *Biochemistry*, **34**, 8076–8081.

Nicholls,A., Sharp,K. and Honig,B. (1991) *Proteins*, **11**, 281–296.

Nicholson,G.M., Willow,M., Howden,M.E.H. and Narahashi,T. (1994) *Pfluegers Arch.*, **428**, 400–409.

Nielsen,K.J., Thomas,L., Lewis,R.J., Alewood,P.F. and Craik,D.J. (1996) *J. Mol. Biol*., **263**, 297–312.

Olivera,B.M., Gray,W.R., Zeikus,R., McIntosh,J.M. and Varga,J. (1988) *Science*, **230**, 1338–1343.

Olivera,B.M., Miljanich,G.P., Ramachandran,J. and Adams,M.E. (1994) *Annu. Rev. Biochem.*, **63**, 823–867.

Pal,D. and Chakrabarti,P. (1999) *J. Mol. Biol.*, **294**, 271–288.

Pallaghy,P.K., Nielsen,K.J., Craik,D.J. and Norton,R.S. (1994) *Protein Sci.*, **3**, 1833–1839.

Regan,L.J. (1991) *J. Neurosci.*, **11**, 2259–2269.

Rees,D.C. and Lipscomb, (1982) *J. Mol. Biol.*, **160**, 475–498.

Savarin,P., Guenneugues,M., Gilquin,B., Lamthanah,H., Gasparini,S., Zinn-Justin,S. and Menez,A. (1997) *Biochemistry*, **37**, 5407–5416.

Scanlon,M.J., Naranjo,D., Thomas,L. Alewood,P.F., Lewis,R.J. and Craik,D.J. (1997) *Structure*, **5**, 1585–1597.

Skinner,W.S., Adams,M.E., Quistad,G.B., Kataoka,H., Cesarin,B.J., Enderlin,F.E. and Schooley,D.A. (1989) *J. Biol. Chem.*, **264**, 2150–2155.

Sun,P.D. and Davies,D.R. (1995) *Annu. Rev. Biophys. Biomol. Struct.*, 24, 269–291.

Terlau,H., Shon,K.-J., Grilley,M., Stocker,M., Stuehmer,W. and Olivera,B.M. (1996) *Nature*, **381**, 148–151.

Tsien,R.W., Ellinor,P.T. and Horne,W.A. (1991) *Trends Pharm. Sci*., **12**, 349–354.

Uson,I., Sheldrick,G.M., de la Fortelle,E., Bricogne,G., Di Marco,S., Priestle,J.P., Gruetter,M.G. and Mittl,P.R. (1999) *Structure*, **7**, 55–63.

Vriend,G. (1990) *J. Mol. Graphics*, **8**, 52–56.

Weiss,M.S., Jabs,A. and Hilgenfeld,R. (1998) *Nature Struct. Biol.*, **5**, 676–676.

Yanagawa,Y., Abe,T., Satake,M., Odani,S., Suzuki,J. and Ishikawa,K. (1988) *Biochemistry*, **27**, 6256–6262.

Zhou,P.A., Xie,X.J., Li,M., Yang,D.M., Xie,Z.P., Zong,X. and Kiang,S.P. (1997) *Toxicon*, **35**, 39–45.

Received May 9, 2000; revised February 26, 2001; accepted June 18, 2001