Article No. mb982057 *J. Mol. Biol.* (1998) **283**, 205–219

JMB

A Structural Homologue of Colipase in Black Mamba Venom Revealed by NMR Floating Disulphide Bridge Analysis

Jérôme Boisbouvier[1], Jean-Pierre Albrand[1]†, Martin Blackledge[1] Michel Jaquinod[1], Hugues Schweitz[2], Michel Lazdunski[2] and Dominique Marion[1]*

[1]*Institut de Biologie Structurale Jean-Pierre Ebel (CEA-CNRS) 41 Av. des Martyrs, 38027 Grenoble, France*

[2]*Institut de Pharmacologie Moléculaire et Cellulaire (CNRS-UPR 411) Sophia-Antipolis, 06560 Valbonne, France*

**Corresponding author*

The solution structure of mamba intestinal toxin 1 (MIT1), isolated from *Dendroaspis polylepis polylepis* venom, has been determined. This molecule is a cysteine-rich polypeptide exhibiting no recognised family membership. Resistance to MIT1 to classical specific endoproteases produced contradictory NMR and biochemical information concerning disulphide-bridge topology. We have used distance restraints allowing ambiguous partners between S atoms in combination with NMR-derived structural information, to correctly determine the disulphide-bridge topology. The resultant solution structure of MIT1, determined to a resolution of 0.5 Å, reveals an unexpectedly similar global fold with respect to colipase, a protein involved in fatty acid digestion. Colipase exhibits an analogous resistance to endoprotease activity, indicating for the first time the possible topological origins of this biochemical property. The biochemical and structural homology permitted us to propose a mechanically related digestive function for MIT1 and provides novel information concerning snake venom protein evolution.

Keywords: snake toxin; colipase; nuclear magnetic resonance; disulphide connectivities; resistance to endoproteases

Introduction

Among snake venom proteins, many compounds have been isolated with pronounced pharmacological function (Harvey, 1991). These toxins act diversely to immobilise prey; either by blocking the central and peripheral nervous systems, affecting various types of muscles, or by perturbing blood coagulation (Ménez *et al*., 1992). It has also been noted that many snake venoms possess enzymatic activity closely related to that of digestive proteins (Kochva, 1987; Housset & Fontecilla-Camps, 1996). These enzymes cause tissue damage, thereby facilitating toxin diffusion and preparing prey for digestion. In addition to these recognised toxin and enzyme families, another type of protein has been isolated from *Dendroaspis polylepis polylepis* venom: MIT1 (mamba intestinal toxin 1),

†This was the last work of J.P.A., to whom this article is dedicated.

Abbreviations used: MIT1, mamba intestinal toxin 1; NMR, nuclear magnetic resonance; 2D, two-dimensional; 3D, three-dimensional; HSQC, heteronuclear single quanta correlation spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser spectroscopy; NOE, nuclear Overhauser effect; SAI, simulated annealing protocol starting from random coordinates; SAII, simulated annealing protocol starting from SAI determined folds; SSI, SAI with ambiguous restraints for disulphide connectivities; SSII, SAII with ambiguous restraints for disulphide connectivities; rMD, restrained molecular dynamics; E_{exp}, experimental energy; rmsd, root-mean-square deviation; BB, backbone; VDW, van der Waals; ESIMS, electrospray ionisation mass spectrometry; MALDIMS, matrix assisted laser desorption ionisation mass spectrometry; rp-HPLC, reverse phase high-performance liquid chromatography; 4-VP, 4-vinylpyridine; TCEP, tris(2-carboxyethyl)-phosphine; TFA, trifluoroacetic acid; PTH, phenylthiohydantoin; V8, *Staphylococcus aureus* V8 proteinase; $sPLA_2$, secretory phospholipase A_2; BPTI, bovine pancreatic trypsin inhibitor; ppm, parts per million.

E-mail address of the corresponding author: marion@ibs.fr

0022–2836/98/410205–15 $30.00/0

containing 80 residues and five disulphide-bridges. On the basis of sequence similarity, this new protein does not belong to any class of protein so far isolated from snake venom (H.S. *et al.*, unpublished results). MIT1 is relatively abundant (3.2% of the dry venom) and has been shown to contract intestinal; smooth muscle (Schweitz *et al.*, 1990). Despite this common particularity with other toxic components of snake venom, the protein is not toxic and its function and contribution to envenomation are not known. In order to clarify the role of MIT1, we have investigated its solution structure using 2D homonuclear NMR. Indeed, the relatively high proportion of disulphide-bridges present in this protein confers a particular importance on this topological information for accurate 3D structure determination. For this reason, we have concentrated our efforts on the unambiguous determination of the disulphide-bridge pattern (945 are theoretically available), using a recently proposed protocol exploiting floating distance constraints between cysteine S atoms and based only on NMR-derived data (Nilges, 1995).

Results

NMR data analysis and secondary structure determination

Sequence-specific resonance assignments were performed according to standard procedures (Wüthrich, 1986) using the set of spectra acquired at 32°C (Figure 1(a)). In the case of ambiguities caused by spectral overlap or coincidence with the water signal, spectra acquired at 25 and 40°C were used for assignment. Previous sequence-specific resonance assignments were verified with 2D heteronuclear scalar correlation spectra (^{13}C-HSQC and ^{13}C-HSQC-TOCSY, Figure 1(b)) by comparison with characteristic ^{13}C chemical shift ranges (Wishart & Sykes, 1994). Virtually complete resonance assignments were obtained for ^{1}H, aromatic and aliphatic ^{13}C, except for residues 1 and 57, where fast exchange rates with water precluded amide proton assignment. Cysteine $^{13}C^{\beta}$ chemical shift between 37.3 and 48.6 ppm indicated the presence of the five disulphide-bridges. Apart from Pro49, which occurred in *cis* conformation, the five other Pro residues were found to be in a *trans* conformation, as deduced from the presence of strong characteristic sequential NOEs.

Elements of secondary structure were identified from the intensity of the sequential NOEs, the $^{3}J_{HNH^{\alpha}}$ coupling constants, slowly exchanging amide protons and chemical shift index (Wishart & Sykes, 1994; Figure 2(a)). According to these criteria, six extended strands were identified which, from the pattern of long-range NOEs (Figure 2(b)) and slowly exchanging amide protons, form two triple-stranded antiparallel β-sheets. Analysis of the $^{3}J_{HNH^{\alpha}}$ coupling constants and the slowly exchanging amide protons indicated that the short two-residue strand (Ala6-Cys7) seems to be less stable than the five others. The $^{3}10$ helix-typical $\{i\}H^{\alpha}$-$\{i+2\}H_N$ NOEs were observed for residues 10/12, 11/13 and a slowly exchanging Cys13 amide proton, suggesting the presence of a single helical turn (spectral overlap preclude the observation of $\{10\}H^{\alpha}$-$\{13\}H_N$).

Classical biochemical studies of the disulphide bridge arrangement

MIT1 measured by ESIMS yielded an average mass of 8506.4(±1) Da. The measured mass was

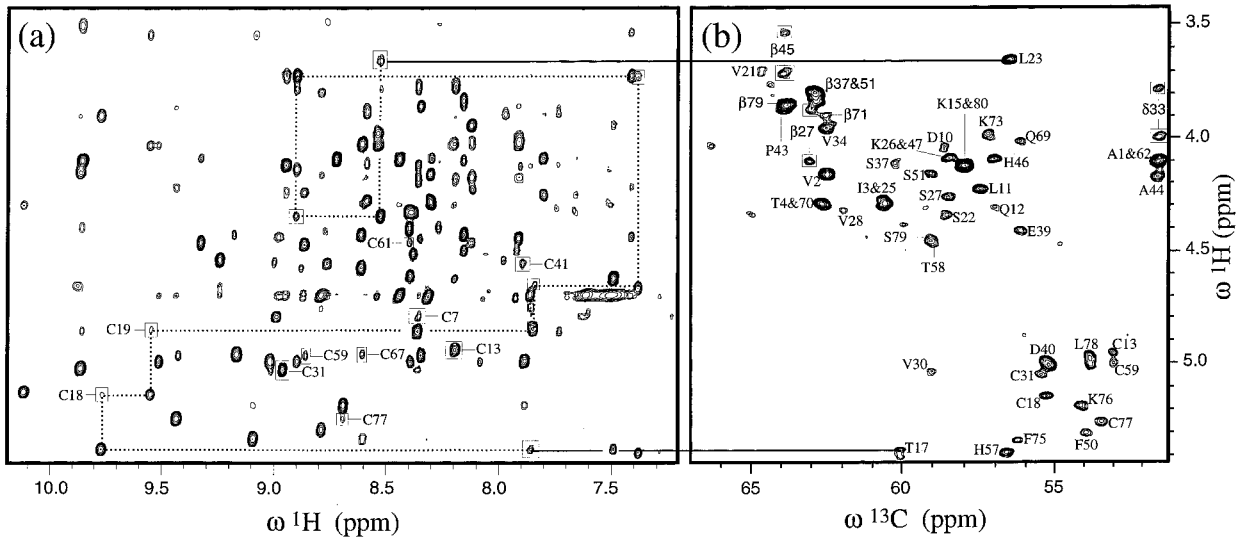

Figure 1. NMR spectra of MIT1. (a) Fingerprint region of a NOESY spectrum with a mixing time of 100 ms. HNH^{α} dipolar correlations are labelled for the ten cysteine residues. The assignment pathway from residues 17 to 23 is displayed. (b) ^{13}C-^{1}H correlation spectroscopy. Each intense C^{α}-H^{α} correlation is labelled with its residue name; serine methylene protons are labelled with β and proline δ methylene with δ. Spectra were acquired at 600 MHz with a 2 mM MIT1 sample and recorded at 32°C in 25 mM sodium acetate buffer (pH 4.5).

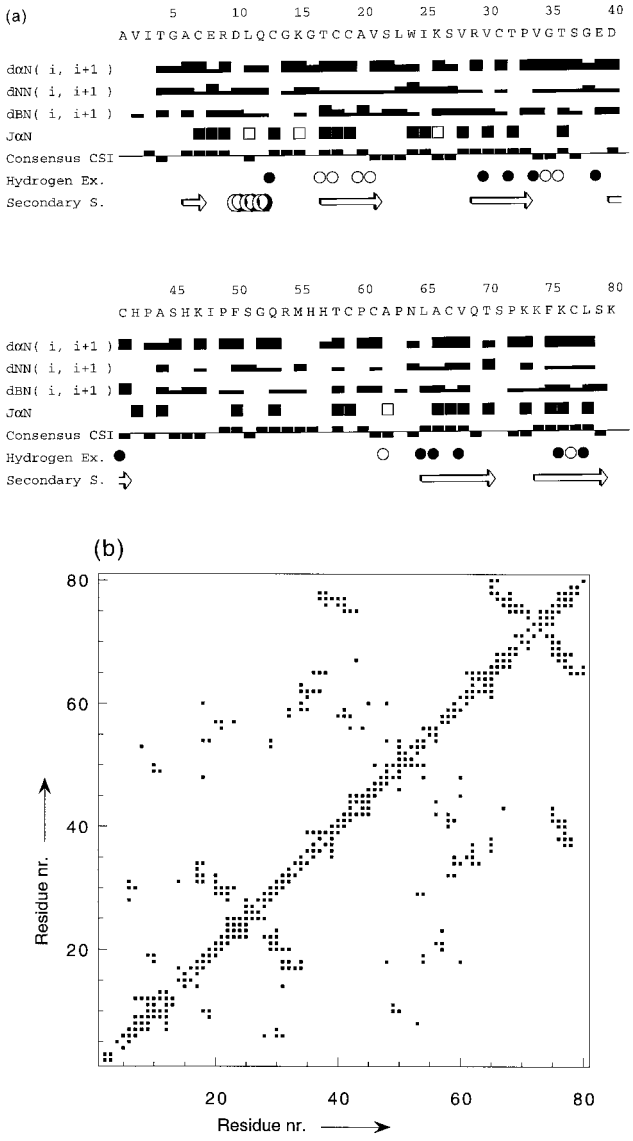

Figure 2. Summary of NMR restraints. (a) Secondary structure of MIT1: summary of sequential NOEs, slowly exchanging amide protons, vicinal backbone HNH^{α} coupling constants ($^{3}J_{\alpha N}$) and consensus chemical shifts indices (Wishart & Sykes, 1994). The relative strengths of the NOEs , categorised as strong, medium or weak, by cross-peak intensity, are indicated by the width of the horizontal bars. Values of $^{3}J_{\alpha N} > 8$ Hz or <5.5 Hz are denoted by filled and open squares, respectively. Backbone amide protons that were observable more than two or 12 hours after transfer of the protein sample from H_2O to $^{2}H_2O$ are marked by open or filled circles, respectively. Positive or negative consensus chemical shift indices for $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{1}H^{\alpha}$ are denoted by filled rectangles above or below the axis, respectively. Corresponding secondary structure is depicted below. (b) Matrix representation of the NOE restraints *versus* the MIT1 sequence.

10 Da less than the mass deduced from the sequence (8517.2 Da), indicating that the ten cysteine residues are probably involved in disulphide-bridges. No incorporation of 4-vinylpyridine (4-VP) was observed after alkylation of the native protein, indicating that all S-S bridges are formed. To locate these disulphide-bridges, MIT1 was hydrolysed by trypsin. Peptides were separated by rp-HPLC, collected, and their masses were measured by ESIMS and MALDIMS. The masses of peptides 16-26 and 30-54 measured by ESIMS were 1179.3 and 2608.6 Da, but the masses calculated from the sequence were 1181.5 Da and 2611.9 Da, suggesting that for each peptide the two cysteine residues are linked by a disulphide-bridge, and implying the presence of bridges 18-19 and 31-41. This was definitely demonstrated by on-probe reduction with tris(2-carboxyethyl)-phosphine (TCEP) since a shift of 2 Da was observed. By sequence comparison, a third peak at $m/z = 1368.5$ Da was assigned to peptide (77-80) linked by an S-S bridge to peptide 1-9. Upon reduction with TCEP, the initial peak at 1368.5 Da was converted into two new ion signals $m/z = 450.6$ and 920.5 Da and, therefore, the initial dipeptides were linked by Cys7 and Cys77. However, all UV chromatograms (peptic, tryptic as well as V8) showed major peaks corresponding approximately to the retention time of the native protein. These components were analysed by MALDIMS and their masses indicate that the major part of the initial sample was not degraded even after 48 hours exposure and consequently presented a resistance to proteolysis. Thus, despite endoprotease resistance, a low proportion of fragments purified after trypsin proteolysis permitted unambiguous identification of three of the five disulphide bridges (18-19, 31-41 and 7-77).

Ambiguous constraints for determination of S-S bridging pattern

Initial structure calculations performed using NMR data (NOE and dihedrals restraints are summarised in Table 1) in combination with inter-sulphur constraints identified by mass spectrometry resulted in highly strained structures, revealing a clear contradiction between measured dipole interactions and these three disulphide bonds. The most probably explanation of this apparent contradiction between information derived from two classical methods is that the experimental results did not derive from the same form of the protein. We therefore attempted to determine the global fold using only NMR-derived data without any disulphide-bridge restraints, in order to obtain the disulphide-bridge pattern independently. Analysis of the position of the cysteine side-chain strongly suggested the existence of a disulphide bond between Cys18 and Cys59. Other S atoms are arranged in two distinct clusters. The spatial proximity of the four cysteine residues in each cluster, and the steric hindrance of the protonated S atoms, renders the direct determination of the cysteine connectivity virtually impossible without proposing explicit partners in the calculation.

Table 1. Summary of constraints and statistics for Discover structure calculations on MIT1 in solution

Constraints for structure calculations	
Total NOE restraints	687
Intra-residue	154
Sequential ($\|i-j\|=1$)	217
Short range ($1<\|i-j\|\leqslant 4$)	77
Long range ($\|i-j\|>4$)	239
Hydrogen bond constraints	17
Dihedral angle constraints	
ϕ	30
χ_1	22
Experimental statistics[a]	
Number of experimental violations	
Distances >0.1 Å	13.4 ± 2.6
Distances >0.3 Å	0.46 ± 0.6
Dihedrals >1°	5.34 ± 1.76
Dihedrals >3°	0.205 ± 0.469
Energetic statistics[a,b]	
Bond	13.9 ± 0.8
Angle	107 ± 6
Dihedral	106 ± 5
Put-of-plane	3.08 ± 0.53
H bond	-32.8 ± 2.0
VDW	-207 ± 8
Electrostatic	-884 ± 16
Total	-893 ± 16
Experimental	29.0 ± 3.2
Structural statistics[a]	
BB[c,d]	0.813 ± 0.149
BB central core[e]	0.494 ± 0.071
Heavy[d]	1.29 ± 0.14
Heavy central core[e]	0.880 ± 0.065
PROCHECK[50]	
Residues in most favourable regions[f] (%)	75.1 ± 4.35
Residues in allowed regions[f] (%)	23.2 ± 4.24

[a] Average values for the rMD stage 39 structures ensemble.
[b] In kcal mol^{-1}
[c] Backbone atoms correspond to N, C^{α}, C.
[d] R.m.s.d. values (in Å) from residues 6 to 79.
[e] R.m.s.d. values (in Å) from residues: 6 to 10, 16 to 12, 29 to 43, 57 to 69 and 75 to 79.
[f] Average values from residues 6 to 79 for the rMD stage 39 structures.

In order to obtain the complete disulphide pattern, a more complex calculation involving the use of ambiguous distance constraints between cysteine S atoms was performed. The existence of the disulphide-bridge between residues 18 and 59 was confirmed by the results of calculation (SSI), starting from random conformations and using ambiguous distance restraints between each of the ten cysteine S atoms. Among the 77 calculated structures, 15 were characterised by low experimental energy ($E_{exp} < 50$ kcal/mol). As ambiguous restraints are comparable to overlapped NOEs, this type of restraint can be satisfied by more than two S^{γ} atoms (see equation (1) in Methods) a situation that is obviously unphysical and can be rejected (Nilges, 1995). Physically significant structures all contained the 18-59 bridge and two separate clusters of four cysteine residues in the two halves of the molecule. A completely unambiguous assignment of the disulphide-bridge partners in each cluster from a calculation starting from random Cartesian coordinates would require sampling a prohibitively large ensemble. We therefore used this structural ensemble as an initial coordinate set for the more efficient calculation (SSII), which has a reduced sampling radius and does not allow complete refolding of the molecule (Blackledge *et al.*, 1995). We also introduced the unambiguously assigned bridge 18-59 and ambiguous restraints only between the four S atoms of each cluster.

Based on nine reference structures, 90 new structures were calculated with such ambiguous restraints. Analysis of distances between S atoms over the ensemble allowed us to assign the disulphide topology (Figure 3). Only average distances between pairs 7-19, 13-31, 41-67 and 61-77 are below 2 Å. Moreover, only four calculated structures are characterised by different topologies: one was obviously unphysical, two were characterised by high experimental and physical energy. The remaining structure (characterised by topologies 41-77 and 61-67 in the C-terminal part) possessed close contacts (Cys67-H^{β}/Cys61-H^{β} 2.5 Å, Cys67-H^{β}/Pro43-H^{β} 2.5 Å and Cys41-H^{β}/Cys61-H^{α} 2.9 Å) with no corresponding observable NOE cross-peak. All these cases were therefore rejected. The calculations thus converged to a well-defined minimum with a statistically and energetically favourable disulphide-binding pattern.

Specific disulphide-bridge reduction and biochemical investigation of the major from

In order to verify the disulphide topology proposed above, in contradiction to that detected by classical enzyme proteolysis (*vide supra*), a new strategy was designed to produce and separate partially reduced and alkylated proteins using the

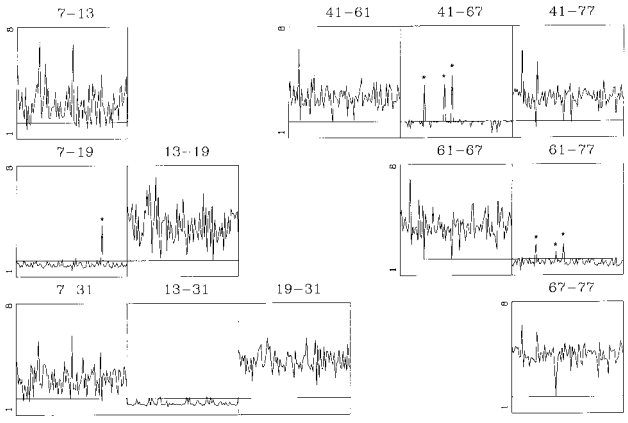

Figure 3. NMR determination of the disulphide-bridge topology. Distribution of the distances between the N-terminal and C-terminal cysteine S atoms for the SSII structure ensemble. The *x*-axis corresponds to each of the 90 conformers of the SSII ensemble. The *y*-axis corresponds to the distance (from 1 to 8 Å) between S atoms of the cysteine residues labelled in the graph title. Normal S-S distances for a disulphide-bridge (2.02 Å) are shown as horizontal lines. Stars correspond to the four exceptions (see Results).

water-soluble reduction agent TCEP (Gray, 1993). MALDI mass spectra measurements performed on partially reduced and alkylated MIT1 after reaction with an excess of 4-VP yielded a mass that is 420 Da higher than the mass of MIT1 (8510(±3) Da). This mass difference corresponds to the incorporation of four moles of the alkylating agent (105 Da) per mole of MIT1, implying two broken disulphide-bridges. To obtain more detailed information as to the origin of these mass difference, modified MIT1 was digested with trypsin and peptide mixtures were subsequently separated by rp-HPLC. The relevant peptides were measured by MALDIMS and submitted to Edman degradation. The location of the 4VP-Cys and of the diPTH-Cys has allowed us to assign two disulphide-bridges, 7-19 and 61-77. Two other peptide bonds were localised by measurement of mass of the peptide obtained by thermolysin hydrolysis at 65°C. The S-S bridges were located between cysteine residues 13 and 31, and 41 and 67. In both these experiments, mass spectra as well as UV chromatograms show that all the protein had been digested, confirming the presence of a major form (approximately 95%) and a minor form with two sets of S-S bridges and retrospectively confirmed that those obtained by NMR correspond to the major form.

Determination and statistical analysis of the tertiary structure

To refine our structures, we added this information as five covalent bonds during the final calculations (SAI, SAII and rMD). Based on low experimental values, we selected 46 structures for the rMD refinement stage (with complete AMBER4 force-field) from an SAII ensemble of 110 conformers. After this stage, seven structures with higher physical energy were discarded, leaving a representative ensemble of 39 structures. Figure 5(a) shows the superposition on the mean structure of the ensemble backbone atoms and cysteine side-chain atoms for residues 6 to 79. MIT1 is a flattened protein (35 Å × 25 Å × 15 Å), with most residues in the β-sheet region of the Ramachandran plot (Figure 6). The central core is composed of two successive, similar, three-strand β-sheets (with a $+2x, -1$ topology (Richardson, 1981); strands 6-7, 17-21, 29-33, 40-41, 65-71 and 74-79), which confer an obvious symmetry on the protein. Four fingers of varying length that protrude from the central core connect the secondary elements to each other.

The refined solution structure presents satisfactory resolution as judged by the rmsd of the backbone (C′, C^{α}, N) and heavy atoms with respect to the mean structure (Table 1 and Figure 5(a)). In general, the central core backbone atoms are very well defined with backbone rmsd of less than 0.6 Å. The extremities of the four protruding fingers were less well defined as judged by a higher rmsd value (Figure 4(b) and (c)). According to the

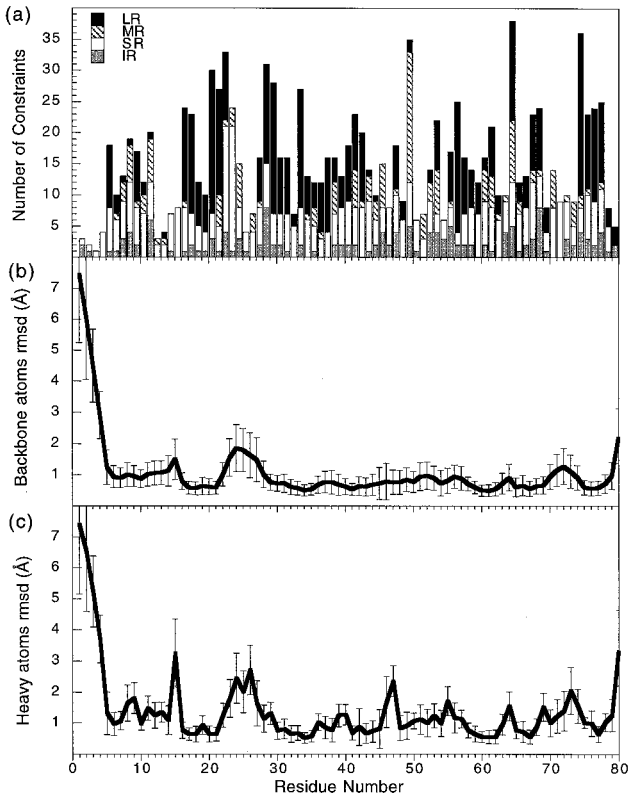

Figure 4. Correlation between the number of NOEs and the rmsd value. (a) Number of NOEs per residues used in the final round of the structure calculation. NOEs are grouped into intra-residue (IR), sequential $i-j=1$ (SR); medium-range $i-j<5$ (MR); and long-range $i-j \geqslant 5$ (LR). (b) and (c) Average rmsd values for the N, C^{α} & C′ (b) atoms heavy atoms (c) for the 39 final structures.

PROCHECK (Laskowski *et al.*, 1996) analysis of the family of 39 structures, 75% of the structure residues have their backbone conformation in the most favoured regions of the Ramachandran plot (Figure 6). In fact, only the Lys73 backbone conformation systematically sampled a disallowed region. Lys73 occupied the third position of the β-turn at the top of the fourth finger and is fully solvent-exposed. Spectral overlap precludes quantification of most sequential NOEs that could resolve this unfavourable conformation.

Analysis of the side-chain conformations indicated that disulphide-bridges 18-59 and 61-77 occurred predominantly in a left-handed conformation ($\chi_3 = -89.7(\pm 4.3)^{\circ}$, 74%; $\chi_3 = -87.9(\pm 4.2)^{\circ}$, 77%, respectively) 41-67 in predominantly right-handed ($\chi_3 = +92.5(\pm 5.0)^{\circ}$, 74%) while other disulphide-bridges were less well defined. The location of Phe50 side-chain just above Gly52 causes ring current effects that explain the unusual proton chemical shifts of Gly52 HN and Gly52 $H^{\alpha 1}$ (respectively 4.75 and 2.69 ppm), The Gln12 side-chain caps Arg9, which stabilises the helix turn by hydrogen bonding between Gln12 O^{ε} and the Arg9

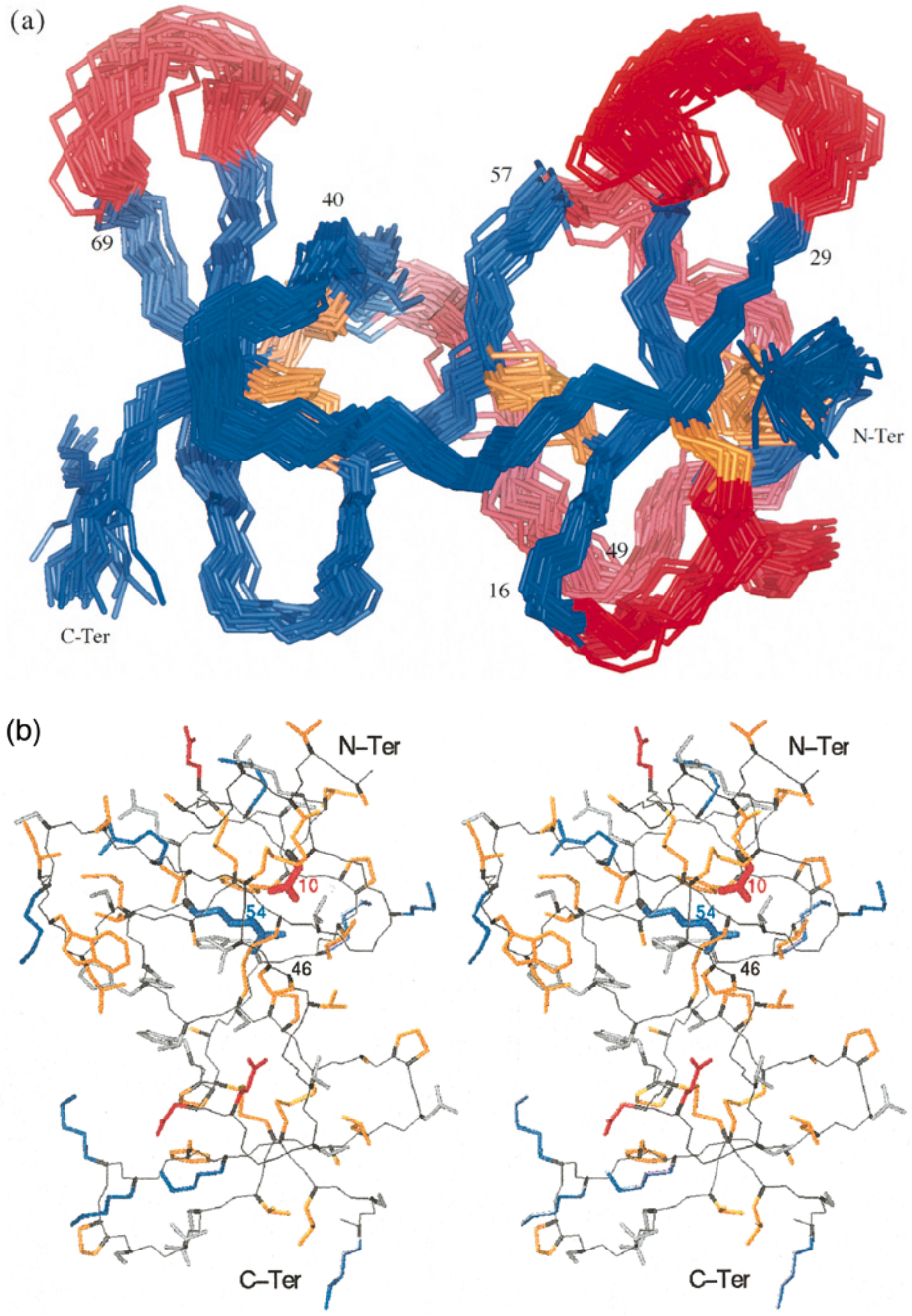

Figure 5. Solution structure of MIT1. (a) Backbone and disulphide-bridge heavy atoms from residues 5 to 80 of the 39 NMR conformers (calculations rMD, Table 1). N, C and C^{α} atoms from residues 6 to 79 of each structure were superimposed on the average structure atoms. Central core residue backbone atoms (5 to 10, 16 to 21, 29 to 43, 57 to 69 and 75 to 80) are displayed in blue, extremity residue backbone atoms of each finger (11 to 15, 22 to 28, 44 to 56 and 70 to 74) are displayed in red, and disulphide-bridges in yellow. (b) Stereo view of the conformer closest to the mean structure of the 39 conformers shown in (a). The following colours were used for the side-chains: blue, Arg and Lys; red, Glu and Asp; yellow, Ala, Cys, Ile, Leu, Met, Phe, Pro, Trp and Val; grey, Asn, Gln, Ser, Thr and His. Buried side-chains of the inner ionic bridge; Asp10 and Arg54 are displayed with a thicker stick and are labelled. The His46 O atom is labelled.

amide proton. The Arg54 side-chain is totally buried and interacts with the carbonyl group of His46 (*via* amino groups) and the carboxyl group of Asp10 (*via* H^{ε}, see Figure 5(b)). As the core of the protein is composed of few hydrophobic buried side-chains (20-21,65), the packing and the stability seem to be due to disulphide and ionic-bridges rather than true hydrophobic interactions.

Structural homology with colipase

Comparison of our structure with available databases using the program Dali (Holm & Sander, 1993) revealed that MIT1 is a very close structural homologue of pig colipase (van Tilbeurgh *et al.*, 1992, 1993; Breg *et al.*, 1995; Egloff *et al.*, 1995a,b; Hermoso *et al.*, 1996; and

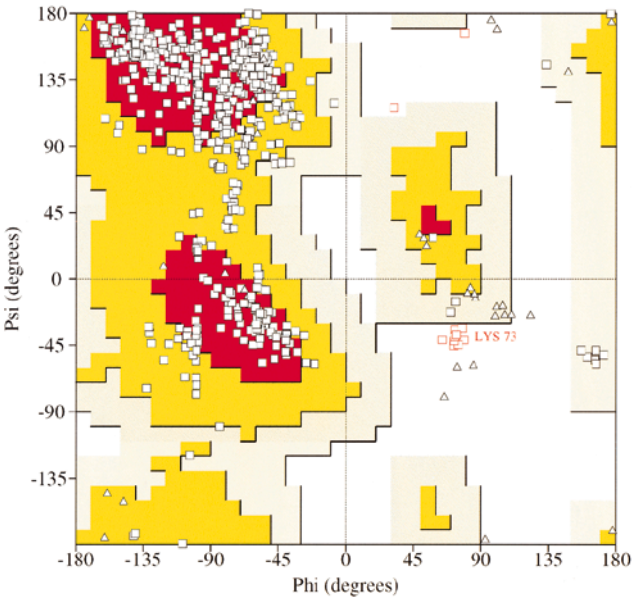

Figure 6. Ramachandran plot for the 39 final structures of MIT1. Each open square represents the backbone conformation of structures residues (6 to 79) for the 39 MIT1 final structures. Only Lys73 sampled the disallowed region (PROCHECK-NMR; Laskowski *et al.*, 1996).

see Figure 7). Colipase is a 90 amino acid residue protein of the pancreatic secretion involved in the digestion of fatty acids. It acts as a cofactor by specifically restoring the activity of pancreatic lipase inhibited by bile salt. This cofactor anchors pancreatic lipase to the lipid/water interface. With specific contact, colipase stabilises the open conformation of lipase, thereby facilitating hydrolysis of triglyceride (Lowe, 1997). Colipase is abundant among mammals, and has been detected in some bird and fish species (Bosc-Bierne *et al.*, 1984; Sternby *et al.*, 1984), but until now has not been reported in the reptilian class.

Both structures contain a well-defined central β-sheet region held together by disulphide-bridges, from which protrude four fingers. These two proteins possess identical disulphide-bridge patterns and similar secondary structure motifs (see Figure 8(a)), consisting mainly of two three-stranded inverse β-sheets. In both cases, the first strand of β-sheet I is composed of only two residues and is partially disordered in both NMR ensembles (our results; Breg *et al.*, 1995). The five other strands (βI-2 to βII-3) are equivalent in the two structures. We note that most of the β-turns are conserved, particularly the type II β-turn (36-39, 39-42); and a 3_{10} helical turn at residues 10 to 12 and 15 to 17. Both the N-term-

Figure 7. Superimposition of colipase MIT1. The two average structures were superimposed on 61% of their backbone atoms (5 to 10, 13-14, 18 to 22, 27 to 44, 57 to 68 and 75 to 80: MIT1 residue numbering is used for the superimposition) with rmsd of 1.3 Å. The colipase backbone is shown in green (from residues 15 to 90) and in MIT1 in purple (from residues 5 to 80).

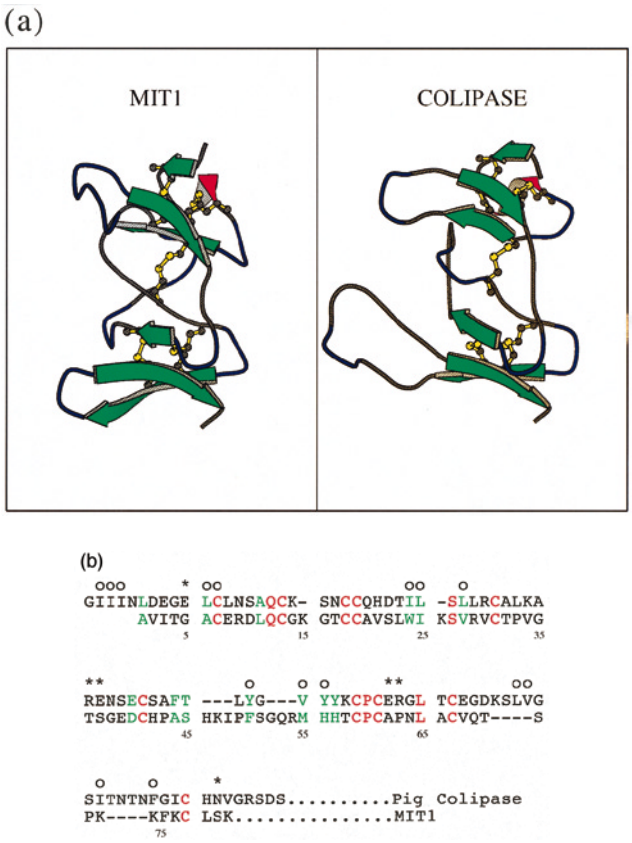

Figure 8. Comparison of the secondary and primary structures. (a) Ribbon diagram of MIT1 and colipase secondary structural elements (MOLSCRIPT; Kraulis, 1991). Helices in red, strands in green and β-turns in blue. (b) sequence alignment of pig colipase and MIT1. Red, sequence identity; green, sequence similarly. o and *, labelled colipase residue involved in lipids and lipase interactions, respectively.

inal part and the extremity of the four fingers of the two structural homologues appear to be disordered in the NMR ensemble (our results and colipase solution structures) and show an increased *B*-factor for the colipase crystallographic structures. Because of the number of deletions or insertions in the fingers, their superimposition is more difficult to interpret. In fact, the four fingers protrude in the same direction but differ in length; fingers I and II are longer by one residue in MIT1, finger III possesses six residues more than the third colipase finger, and the last one is shortened by eight residues in MIT1 (see Figure 8(a) and (b)).

On the basis of structural homology, we have aligned the sequence of the two proteins and found only 14 residues strictly conserved (see Figure 8(b)). Despite this weak sequence identity, the ten cysteine residues are well aligned and the disulphide patterns are the same. It is interesting to note that a well-conserved motif among colipase sequences, (50)YxxYYxCPC(58), is replaced by a homologous sequence in MIT1, (50)FxxxxxHHxCPC(61). Moreover, in spite of an insertion of three residues between Phe50 and His56 in MIT1 *versus* colipase, different side-chain orientations allow the aromatic ring of MIT1 (Phe50) and of its colipase homologue (Tyr50) to be structurally equivalent. Finally, it appears that most of the hydrophobic residues in the fingers (I to III) involved in the interaction between colipase and lipids (Cozzone *et al.*, 1981; van Tilbeurgh *et al.*, 1993; Egloff *et al.*, 1995a; Hermoso *et al.*, 1997) are conserved in MIT1 or replaced by residues with close physico-chemical properties (see Figure 8(b)). In both cases, most of the hydrophobic residues are partially or completely solvent-exposed. In contrast, none of the charged groups involved in the interaction between colipase and lipase (van Tilbeurgh *et al.*, 1992, 1993; Egloff *et al.*, 1995a,b; Hermoso *et al.*, 1996; Lowe, 1997) are conserved in MIT1 (see Figure 8(b)). Consequently, despite the fact that the two proteins are close structural homologues, their electrostatic surfaces are dissimilar (see Figure 9). A particular feature of MIT1 is a positively charged face and, on the opposite face, a negatively charged pocket (Val21 O, Thr32 O^{γ}, Pro33 O, Glu39 O^{ε}).

Colipase digestion

In view of this obvious structural homology, we have tested the biochemical properties of colipase. Analogous tryptic and peptic digestion was performed on porcine colipase, again under conditions that should preserve existing disulphide-bridges, in order to determine whether this protein also presents protease resistance. All the results obtained in this study are comparable to those obtained with MIT1, confirming analogous resistance.

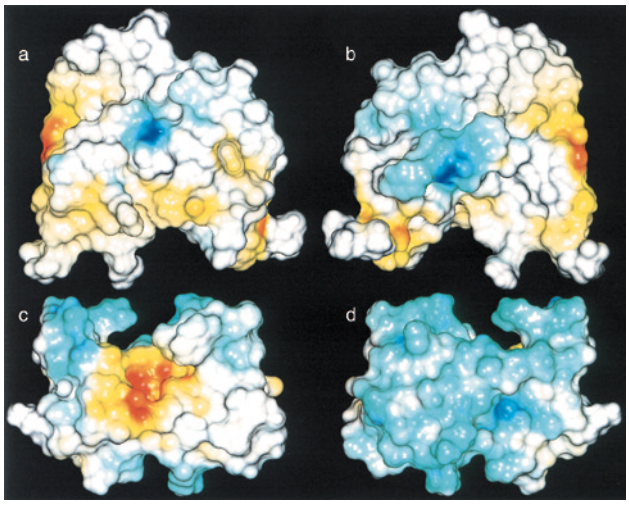

Figure 9. Electrostatic potentials map of colipase and MIT1. (a) and (b) The exposed surface of colipase, (c) and (d) were coloured with the electrostatic potential (Gilson *et al.*, 1987) by linear interpolation between red ($\phi(r) < -3$ kT, negative), white ($\phi(r = 0$ kT, neutral) and blue ($\phi(r) > 3$ kT, positive). (a) and (c) Surfaces were displayed with the same orientation as Figure 5. (b) and (d) Representations were turned by 180° with respect to the vertical axis.

Discussion

NMR determination of disulphide bridge topology

The correct assignment of disulphide-bridges is of fundamental importance for successful structure determination of cysteine-rich proteins by NMR. In the absence of direct chemical shift or scalar coupling evidence, and in view of the fact that no individual interproton distance provides unambiguous evidence for the identification of disulphide-bond partners (Klaus *et al.*, 1993), it is often necessary to resort to biochemical techniques to determine the assignment. The case of MIT1 is particularly complex, given the comparatively large number of unassigned disulphide-bridges present in the protein (ten cysteine residues out of a total of 80 residues), and the high resistance of the protein to classical specific endoproteases. We therefore decided to employ a more robust method based on the use of a floating assignment of disulphide-bond partners, whose assignment is determined only by the NOE-derived distance dataset. This method, proposed and demonstrated on model systems (Nilges, 1995), allowed us to unambiguously assign all five bonds and confidently propose the most probable topology from the 945 possible combinations. To our knowledge, this is the most complex system for which the method has been applied, and the first time that the method has been used to resolve a real experimental problem.

Protease resistance

We have shown that MIT1 exhibits high resistance to digestion by commonly used specific endoproteases. This resistance is unusual, as many other toxins are degraded under similar conditions (Kolbe *et al*., 1993; Gasparini *et al*., 1994; Chung *et al*., 1995; Olamendi-Portugal *et al*., 1996; Calvette *et al*., 1997). It is even resistant to high proteolytic enzyme concentrations under conditions that can unfold other types of proteins (2 M urea, or after preheating at 60°C). As for MIT1, biochemical investigation of colipase cysteine topology with endoproteases (Erlanson *et al*., 1974) led to a disulphide-bridge pattern in contradiction with the high-resolution structure. We have shown that colipase displayed a similar resistance to classical endoproteases. This is probably because colipase must coexist in a protease environment containing trypsin, chymotrypsin and elastase. The structural homology of MIT1 and colipase indicates that the resistance of MIT1 to proteases has similar topological origins. As these proteins are two close structural homologues and display the same kind of unusual behaviour in the presence of endoproteases, this suggests that the contradiction between structural and biochemical data can be attributed to a specific tertiary fold rather than an experimental artefact. This conclusion is reinforced by the fact that all biochemical results are reproducible, and the absence of systematic violations during the structural calculation. Possible explanations for the erroneous assignment of disulphide-bridges may be the presence of a small proportion of protein with a different disulphide-bridge topology that is not enzyme-resistant or a rearrangement of the disulphide-bridges in the presence of the protease. This latter possibility has been excluded by peptic and tryptic digestions performed under conditions that should have preserved existing disulphide-bridges, and which provided results comparable to those obtained at higher pH values.

Function of MIT1

Despite a very limited sequence identity (14/80), it is still possible to superimpose more than 60% of the backbone atoms of the average solution structures of the two proteins to an rmsd of 1.3 Å (1.4 to 1.5 Å with the crystallographic structure of colipase complexed with lipase). The absence of conservation of the charge distribution on the electrostatic surface makes it unlikely that MIT1 and colipase share a common physiological partner. As expected, no clear activity of MIT1 on lipase was detected, and colipase failed to contract or to relax ileum smooth muscle (data not shown). As most of the hydrophobic surface amino acid residues involved in colipase/lipid interaction seem to be conserved, MIT1 may be able to interact with lipids, which may in turn be responsible for an MIT1 activity on the membrane of intestinal smooth muscle cells.

Like colipase, MIT1 is resistant to classical digestive enzymes such as trypsin, pepsin and carboxypeptidase Y. As MIT1 is relatively abundant in snake venom, it may be that after the mamba has killed its prey (due to the action of other mamba toxins) and has ingested it, a non-negligible amount of ''intact'' MIT1 arrives with the rest of the prey in the mamba intestine and plays a role in intestinal contraction favouring the overall digestive process.

Molecular evolution

Homologies between proteins of the pancreatic exocrine secretion and other proteins contained in snake venom have been observed in the past. This has led to the hypothesis that the venom glands of serpents are relatively new structures that probably appeared about 150 to 200 million years ago. They may have been shaped after another exocrine gland that is evolutionarily much older, i.e. the pancreas (Kochva, 1987). According to this hypothesis, we propose that colipase and MIT1, which possess structural and biochemical homology, have a pancreatic common ancestor. Structural similarities between colipase and components of a snake toxin family have been proposed (van Tilbeurgh *et al*., 1992). Proteins of this family, which includes curaremimetic toxins (Hatanaka *et al*., 1994), neuronal toxins (Oswald *et al*., 1991), muscarinic toxins (Ducancel *et al*., 1991), cytotoxins (Rees *et al*., 1987), fasciculins (Le Du *et al*., 1992) and calciseptins (De Weille *et al*., 1991), express diverse functions, but conserve the same overall structure: the ''three-finger fold''. This intriguing observation led to the hypothesis that colipase has evolutionary relationships with this snake toxin family; however, a detailed topological comparison reveals fundamental differences between the two types of molecule and, until now, there lacked biochemical support for such hypothesis (Ménez, 1995; Housset & Fontecilla-Camps, 1996). The recognition of a close structural homologue of colipase in snake venom could be the first biochemical evidence for such a hypothesis (see Figure 10(b)). Moreover, it is interesting to note that protein CM2 (Joubert *et al*., 1982), which is related to another snake toxin family, disintegrin, displays 29% sequence identity with colipase (see Figure 10(a)). Based on this sequence homology, one, may propose that disintegrins could be related to colipase through evolution (see Figure 10(b)).

Recent studies revealed that to acquire venomous function, a few cysteine-rich structural motifs with unrelated functions had been selected by nature (Ohno *et al*., 1998). These highly stable architectures have been exploited to express different functions on the surface of the molecule. The evolutionary mechanisms that govern this general principle of functional diversity associated with structural economy remains unknown (Ohno *et al*., 1998). The particular relationship between snake venoms glands and the pancreas (Kochva, 1987)

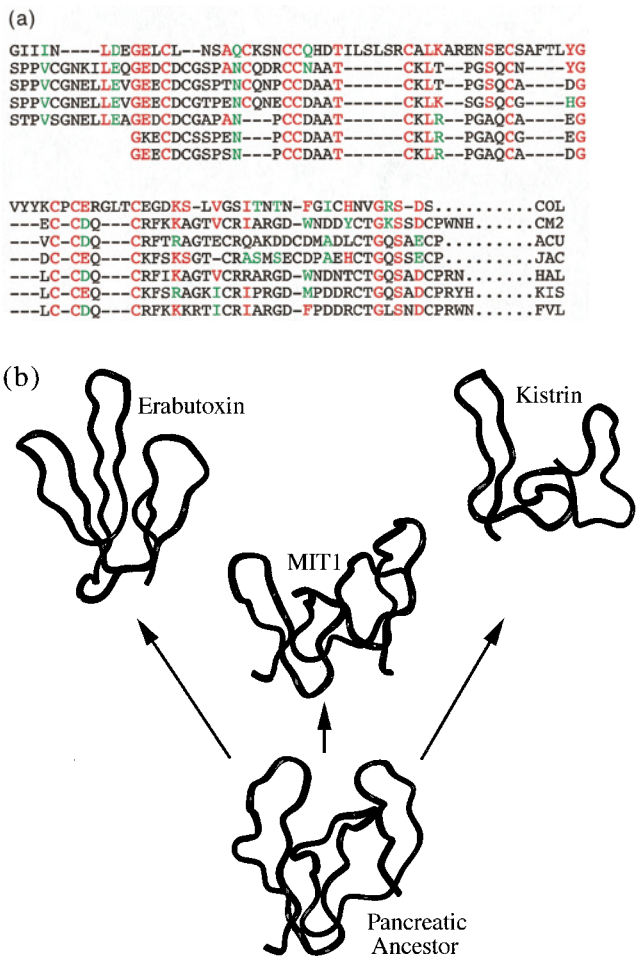

Figure 10. Molecular evolution of snake toxins. (a) Sequence alignment of pig colipase with some disintegrins: red, sequence identity; green, sequence similarity. Col, pig colipase; CM2, disintegrin from *Bitis arietans* (SWISS-PROT accession number P17497); ACU, metalloproteinase disintegrin-like protein from *Agkistrodon contortrix laticinctus* (EMBL accession number U86634); JAC, jarahagin-C protein from *Bothrops jararaca* (PIR accession number JC2245); HAL, prepro-halystatin 3 from *Gloydius halys* (EMBL accession number D28871); KIS, disintegrin kistrin from *Agkistrodon rhodostoma* (PDB accession number 1KST); FVL, flavoridin from *Trimeresurus flavoviridis* (PDB accession number 1RVL). (b) Proposed evolution of three families of snake venom proteins, The colipase scaffold is very well conserved in the MIT1 structure and is related to disintegrin and three-finger toxin motifs. Only residues 25 to 90 are shown for colipase as the pancreatic ancestor. Only residues 15 to 80 are shown for MIT1. As the CM2 structure has not been determined, kistrin (residues 12 to 59; Alder *et al.*, 1991) was chosen to represent the disintegrin scaffold (based on sequence homology; Joubert *et al.*, 1982.). Erabutoxin b (residues 1 to 62; Rees *et al.*, 1987) represents the three-finger toxin architecture (based on structure homology as shown by van Tilbeurgh *et al.* (1992) and discussed by Housset & Fontecilla-Camps (1996) and Ménez (1995)).

could explain that a number of these stable structural motifs correspond to secreted pancreatic proteins. A particularly well-known case is that of secretory phospholipase A_2 (sPLA$_2$, Lambeau *et al.*, 1997). Structural homologues of pancreatic sPLA$_2$ are present in snake venoms; some of them have acquired toxic function with different types of toxicity (Lambeau *et al.*, 1989, 1997). Pancreatic and venom sPLA$_2$ can even share the same membrane receptors (Lambeau *et al.*, 1994). Another example is the high degree of structural homology between the trypsin inhibitory protein and BPTI and the snake dendrotoxins (Lancelin *et al.*, 1994) that block voltage-sensitive K^+ channels (Halliwell *et al.*, 1986) to calcicludine (Schweitz *et al.*, 1994), which blocks voltage-sensitive Ca^{2+} channels. In MIT1, the colipase scaffold has been well conserved; in particular, the disulphide-bond pattern, hydrophobic surface residues, and resistance to protease have been retained (see Figures 7 and 8). Like BPTI/dendrotoxins homologues, specific modifications of these surfaces charges (see Figure 9) would then have allowed MIT1 to acquire a new function in snake venom. So the existence of this new architecture suggests that, as for other selected motifs, there might be other structural analogues of colipase in snake venoms with other types of toxicity.

Conclusion

The solution conformation of MIT1 has been determined using NMR spectroscopy and structure calculations. MIT1 is a cysteine-rich protein found in black mamba venom, for which no sequence homology has been recognised, and which belongs to no recognised family proteins. The protein is highly resistant to classical specific endoproteases, a characteristic that forced us to use more complex methods to resolve the important problem of disulphide-bridge assignment. The use of floating assignment for the S-S bonds during the simulated annealing calculation has allowed us to unambiguously assign the correct disulphide-bridge topology from the 945 possible combinations. This result permitted us to recognise the homology of this snake venom protein with mammalian pancreatic colipase and has allowed us to propose hypothesis for the possible origin and function of this protein.

Methods

Sample

Because of the lack of information on the topology of the disulphide-bridges, we preferred working on the extracted protein of natural venom rather than a synthetic one, to avoid any non-native disulphide-bridge arrangement errors during synthesis. MIT1 was purified from *Dendroaspis polylepis polylepis* as described (Schweitz *et al.*, 1990; MIT1 corresponds to fraction G4 in that work). MIT1 displays 96% sequence identity with another protein purified in 1980 from black mamba venom (Joubert & Strydom, 1980); we assumed that these two proteins corresponded to the same component of the venom.

Biochemical investigations

In the first set of experiments, native MIT1 was digested with trypsin (10 mM NH_4HCO_3 (pH 7.2) at 37°C, pH 7.2 for 18 hours with substrate to enzyme ratio (E/S) 1/100 (w/w) or at pH 6.5 for two hours with $E/S = 1/10$, w/w) and with *Staphylococcus aureus* V8 proteinase (50 mM NH_4HCO_3 (pH 7.2) for six hours at 37°C, $E/S = 1/100$, w/w). For the peptic hydrolysis reactions, 400 µl of immobilised pepsin (Pierce Chemical Company, Rockford, USA) was deposited into the filter cup of an Ultrafree-MC filter (30,000 Da cut-off) and washed three times with 200 µl of HCl (pH 1.9) to eliminate possible contamination. Dissolved MIT1 (4 nmol) was digested with immobilised pepsin for two to 30 minutes at 37°C at pH 1.9 with $E/S = 50/1$, w/w. The reactions were stopped by high-speed centrifugation at 12,000 *g* and the released peptides were purified by rp-HPLC (Brownlee, C_{18}, 5 µm, 2.1 mm × 100 mm). Absorbance was monitored at 214 nm. Elution was performed isocratically for two minutes at 100% solvent A (water containing 0.1% (v/v) TFA), then a linear gradient was run from 0 to 50% solvent B (acetonitrile/water, 90/10 (v/v) containing 0.08% TFA) over 40 minutes with a flow-rate of 0.2 ml/minute, then from 50% to 100% solvent B over five minutes. The peptide sequences of MIT1 peptides were obtained using an Applied Biosystems sequenator (477A) coupled to an on-line PTH-amino acid analyser (120A). All sequencing reagents were obtained from Applied Biosystems.

In a second set of experiments, MIT1 was dissolved in 90 µl of water containing 0.1% TFA, followed by the addition of 10 µl of a TCEP solution (100 mg/ml). After reduction for six minutes at 55°C, the reactions were quenched by injection of the mixture onto the rp-HPLC column and the TCEP was eluted using water containing 0.1% TFA. In order to modify the free thiol groups, the column was rapidly flushed with 10 mM acetate buffer (pH 7.5). Then 100 µl of 50 mM 4-vinylpyridine was injected onto the column in excess relative to total SH and the reaction was allowed to proceed for 15 minutes on the column (Gray, 1993). The alkylated proteins were measured and submitted to a tryptic digest. To circumvent the protease-resistance, proteolysis of MIT1 was performed at high temperature with thermolysin (10 mM NH_4HCO_3 (pH 7.0) for one hour at 65°C, $E/S = 1/10$, w/w). The released peptides were separated as described above.

Mass spectrometry

Electrospray ionisation mass spectrometry (ESIMS) was performed using a SCIEX API III+ triple quadruple mass spectrometer (Perkin-Elmer Sciex Instruments, Thornhill, Canada) equipped with a nebuliser-assisted electrospray (ionspray) source. Calibration was performed in positive mode using poly(propylene glycol) ions. Mass spectra were analysed using a Quadra 950 data system (Apple Computer Inc., Cuppertino, USA). MacBioSpec software (Perkin-Elmer Sciex) was used to calculate the mass of the protein and peptides from their sequence.

Matrix assisted laser desorption ionization mass spectrometry (MALDIMS) spectra were recorded on a Voyager Elite XL instrument from Perseptive Biosystems (Framingham, USA). 2,5-Dihydrobenzoic acid dissolved in 50% acetonitrile solution was used as the matrix: 1 µl of the protein or peptide mixture (1 to 5 pmol/µl) was mixed with 1 µl of the matrix solution; 1 µl of this mixture was deposited on the target and dried.

NMR analysis

NMR experiments were recorded at 600 MHz ^{1}H-frequency on a Bruker AMX 600 NMR spectrometer. Assignments were made at 25, 32 and 40°C at pH 4.5 in 90% ^1H_2O/10% ^2H_2O or 99% ^2H_2O, using a 2 mM sample. Spectra used for resonance assignments included the 2D experiments DQF-COSY (Rance *et al*., 1983), clean TOCSY (Griesinger *et al*., 1988), ^{13}C-HSQC (Bodenhausen & Ruben, 1980), ^{13}C-HSQC-TOCSY (Medvedeva *et al*., 1993) and NOESY (Kumar *et al*., 1980) with 100, 150 and 200 ms mixing time. (Assignments at 32°C are available as Supplementary Material.) Slowly exchanging amide protons were identified in a series of experiments after dissolution in ^2H_2O.

$^3J_{HNH^\alpha}$ scalar coupling constants were measured from DQF-COSY after deconvolution to avoid overestimation due to linewidth (Schwöbel, 1997). Backbone ϕ angles were constrained to $-80° < \phi < -40°$, $-160° < \phi < -80°$ and $-140° < \phi < -100°$ for small (<5.5 Hz), large (>8 Hz) and very large (>9 Hz) $^3J_{HNH^\alpha}$ couplings, respectively. $^3J_{H^\alpha H^\beta}$ couplings qualitatively measured from DQF-COSY spectra and NOE data yielded stereospecific assignments for β-protons (Redfield, 1993). χ_1 angles were constrained in structure calculation with a precision of ±30°.

Peak volumes for NOESY experiments with 50 and 100 ms mixing time were integrated using Felix (10/95, MSI), and calibrated using several known distances between backbone protons in well-identified β-sheet stands (Wüthrich, 1986). Distances were constrained in the structure calculation with a precision of ±20% of the corresponding measured NOE peak volume. A set of unambiguously assigned NOE cross-peaks was used to determine preliminary structures. The determination of the global fold served to assign ambiguous distance restraints. Classical corrections were applied for pseudoatoms (Wüthrich *et al*., 1983; Fletcher *et al*., 1996). A correction of 0.5 Å was added to the upper limit in case of weakly overlapping NOE cross-peaks. Lower limits were conserved only for HN, H^α and H^β where neither residue displayed evidence of obvious dynamic processes (disordered residues, large linewidth). Based on the presence of 17 hydrogen bonds N-O (2.7 to 3.2 Å), NH to O (1.7 to 2.2 Å) distance constraints (Sunnerhagen *et al*., 1997) were added for structure calculation. The criterion for a hydrogen bond was significant amide exchange protection and evidence of an acceptor in previously calculated structures identified with the program STRIDE (Frishan & Argos, 1995).

Determination of the protein fold

Structure calculations were performed using DISCOVER interfaced to the InsightII program for the visualisation and result analysis (version 10/95, MSI). The AMBER4 covalent force-field was used for all calculations for the simulated annealing (SA) protocols, in which a simple quartic non-bond term was employed (Blackledge *et al*., 1995). All ω dihedrals were forced to *trans* except that preceding Pro49, as these was strong evidence from the NOE data that this peptide bond was in the *cis* conformation.

The global fold of the protein was determined using a 60 ps SA protocol (call SAI) at a nominal temperature of

1000 K starting from randomised initial coordinates and containing 10 ps of slow cooling to 100 K. Experimental distances restraints were scaled to a maximum of 50 kcal mol^{-1} $Å^{-2}$ (200 kcal mol^{-1} $Å^{-2}$ for dihedral restraints) during the first 30 ps. The molecule was then minimised in the complete AMBER force-field. Determination of the global fold structural ensemble was used as starting coordinate for a second SA protocol (calculation called SAII) starting with a 5 ps scaling period at 2000 K followed by 2 ps of sampling before smoothly cooling to 100 K over 11 ps. The structures were minimised as in SAI.

NMR determination of the disulphide-bridge pattern

The disulphide covalent bond introduces a fixed distance of 2.02 Å between two sulphur atoms that may be far apart in the primary sequence. Because of steric repulsion, cysteine side-chain dihedral angles are restrained around 90° for the S^{γ}-S^{γ} bonds and 103° for the C^{β}-S^{γ} bonds (Heck *et al*., 1994). Knowledge of the disulphide-bridge pattern therefore considerably restricts the conformational space accessible to the polypeptide chain.

In order to include the existence of five disulphide-bridges as structural information, we used ambiguous restraints (from 0 to 2.02 Å, 100 kcal mol^{-1} $Å^{-2}$ as force constant) between one specific Cys S^{γ} and other possible Cys S^{γ}, as proposed by Nilges (1995). The constraints were applied using the algorithm available in the Discover package for the interpretation of overlapped NOE cross-peaks and, according to equation (1), each ambiguous restraint forces a given Cys S^{γ} atom to be close to at least one of the N other Cys S^{γ} atoms:

$$r_{\text{effective}} = \left(\frac{1}{N} \left(\sum_{i \neq j} \frac{1}{(r_{S_i - S_j})^6} \right) \right)^{\frac{-1}{6}}$$

$$F_{\text{NOE overlapped}} = 0, \quad \text{if} \quad r_{\text{effective}} < r_{\text{u}} \qquad (1)$$

or:

$$F_{\text{NOE overlapped}} = k_{\text{u}}(r_{\text{effective}} - r_{\text{u}})^2, \quad \text{if} \quad r_{\text{effective}} \geqslant r_{\text{u}}$$

As Discover allows the use of only ambiguous restraints between H atoms, we have modified the AMBER4 force-field library to include a modified cysteine residue, with an otherwise transparent H^{γ} proton superposed on the S^{γ} atom. The lower ambiguous distance restraints are set to 0 but, because of steric repulsion, two Cys S^{γ} atoms could not be closer than 1.66 Å. We added this ambiguous information to previously identified NMR restraints to generate structures with a modified version of previous SA protocols, followed by minimisation with a simple quartic non-bond term used in the sampling period to avoid electrostatic repulsion between cysteine S atoms.

In the first calculation (called SSI), with the SA protocol starting from randomised initial coordinates, ambiguous restraints were applied between all cysteine S atoms. Among the 77 structures calculated, nine were selected and were used as starting coordinates for the second SA protocol, which produced 90 new structures. In this calculation (called SSII), we introduced the unambiguously assigned bridge 18-59 and ambiguous restraints only between the four S atoms each cluster (7,13,19,31 and 41,61,67,77). SSI and SSII are the counter parts of SAI and SAII with suitable modification for introducing ambiguous S-S restraints.

Structure refinement

Determination of the disulphide-bridge pattern allowed us to introduce corresponding covalent restraints, 2.0 to 2.1 Å for $d(S^{\gamma}\text{-}S^{\gamma})$ and 3.03 to 3.1 Å for $d(C^{\beta},S^{\gamma})$. Final structures were calculated using the two-stage simulated annealing protocol described above (SAI to SAII). In all, 46 structures were selected on the basis of low experimental energies, which were then refined using a restrained molecular dynamic protocol (Blackledge *et al*., 1995; calculated rMD). In this stage, the use of the complete force-field led to a physically more viable structure. The effects of solvent were simulated implicitly using distance-dependent (Brooks *et al*., 1983; Weiner *et al*., 1984) and reduced charges on the polar side-chains except for Asp10 and Arg54 as these side-chains were completely buried. The weight of the experimental constraints was reduced by a factor of 2. After equilibration, the temperature was regulated using weak coupling to a thermal bath (Berendsen *et al*., 1984). The molecule was allowed to evolve at 750 K for 10 ps, then it was slowly cooled to 300 K over a period of 5 ps, where a further 12 ps of restrained dynamics was performed. Finally, the structure was minimised using a conjugate gradient algorithm.

In all, 39 final structures were selected on the basis of the experimental and physical energy function. The quality of the structure was investigated using PROCHECK-NMR (Laskowski *et al*., 1996). The final MIT1 structures have been deposited in the Brookhaven protein data bank (PDB, accession code 1imt). Colipase structures were taken from the PDB, accession codes 1lpa, 1lpb, 1eth, 1pcn and 1pco.

Calculation of electrostatic surfaces

As the mobility of solvent-exposed side-chains can significantly alter the electrostatic surface, the charge distribution has been determined for a series of structures. After calculating an electrostatic grid for the individual structures, the mean of each grid value is taken to represent the conformational average, whose surface gives a more realistic representation of the effective encounter surface of the molecule than any individual conformer. The electrostatic grids were calculated with the program DelPhi (Gilson *et al*., 1987), implemented in the MSI package, as described (Blackledge *et al*., 1996). Dielectric constants of 80 and 4 were used to represent the solvent and the proteins interior, respectively. An ionic strength of 0.145 M was assumed for the solvent and the Stern layer was taken to be 2.0 Å. Charges were taken from the AMBER4 force-field. Using focusing calculations of the $65 \times 49 \times 37$ points matrix, the final resolution achieved was 0. 7 Å per grid point. The charged exterior of the protein was visualised by projecting the grid values onto a Connolly solvent-accessibility surface.

Acknowledgements

We thank Drs J.-C. Fontecilla-Camps, P. Gans, J.-M. Lancelin and J.-P. Simorre for invaluable advice, and J.-P. Andrieu for Edman sequencing. Dr. C. Chapus is gratefully acknowledged for the biochemical tests of

MIT1 activity on pig lipase and for providing the colipase sample. This work was supported by the Commissariat à l'Energie Atomique CEA, the Centre National de la Recherche Scientifique CNRS, and Molecular Simulations, Inc. This is publication 529 of the Institut de Biologie Structurale Jean-Pierre Ebel.

References

Adler, A., Lazarus, R. A., Dennis, M. S. & Wagner, G. (1991). Solution structure of kistrin, a potent platelet aggregation inhibitor and GP IIb-IIIa antagonist. *Science,* **253**, 445–448.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.

Blackledge, M. J., Medvedeva, S., Poncin, M., Guerlesquin, F., Bruschi, M. & Marion, D. (1995). Structure and dynamics of ferrocytochrome *c*553 from *Desulfovibrio vulgaris* studied by NMR spectroscopy and restrained molecular dynamics. *J. Mol. Biol.* **245**, 661–681.

Blackledge, M. J., Guerlesquin, F. & Marion, D. (1996). Comparison of low oxidoreduction potential cytochrome *c*553 from *Desulfovibrio vulgaris* with the class I cytochrome *c* family. *Proteins: Struct. Funct. Genet.* **24**, 178–194.

Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters,* **69**, 185–189.

Bosc-Bierne, I., Rathelot, J., Bechis, G., Delori, P. & Sarda, L. (1984). Evidence for the existence of pro-Colipase in chicken pancreas and pancreatic juice. *Biochimie,* **66**, 413–416.

Breg, J. N., Sarda, L., Cozzone, P. J., Rugani, N., Boelens, R. & Kaptein, R. (1995). Solution structure of porcine pancreatic proColipase as determined from ^{1}H homonuclear two-dimensional and three-dimensional NMR. *Eur. J. Biochem.* **227**, 663–672.

Brooks, B. R., Broccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. J. (1983). CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.

Calvette, J. J., Schrader, M., Raida, M., McLane, M. A., Romero, A. & Niewiarowski, S. (1997). The disulphide bond pattern of bitistatin, a disintegrin isolated from the venom of viper *Bitis arietans*. *FEBS Letters,* **416**, 197–202.

Chung, D., Gaur, S., Bell, J. R., Ramachandran, J. & Nadasdi, L. (1995). Determination of disulphide bridge pattern in omega-conopeptides. *Int. J. Pept. Protein Res.* **46**, 320–325.

Cozzone, P. J., Canioni, P., Sarda, L. & Kaptein, R. (1981). 360-MHz nuclear magnetic resonance and laser photochemically induced dynamic nuclear polarization studies of bile salt interaction with porcine colipase A. *Eur. J. Biochem.* **114**, 119–126.

De Weille, J. R., Schweitz, H., Maes, P., Tartar, A. & Lazdunski, M. (1991). Calciseptine, a peptide isolated from black mamba venom, is a specific blocker of the L-type calcium channel. *Proc. Natl Acad. Sci. USA,* **88**, 2437–2440.

Ducancel, F., Rowan, E. G., Cassar, E., Harvey, A. L., Ménez, A. & Boulain, J.-C. (1991). Amino acid sequence of a muscarinic toxin deduced from cDNA nucleotide sequence. *Toxicon,* **29**, 516–520.

Egloff, M.-P., Sarda, L., Verger, R., Cambillau, C. & van Tilbeurgh, H. (1995a). Crystallographic study of the structure of colipase and of the interaction with pancreatic lipase. *Protein Sci.* **4**, 44–57.

Egloff, M.-P., Marguet, F., Buono, G., Verger, R., Cambillau, C. & van Tilbeurgh, H. (1995b). The 2.46 Å resolution structure of the pancreatic lipase-colipase complex inhibited by a C11 allyl phosphonate. *Biochemistry,* **34**, 2751–2762.

Erlanson, C., Charles, M., Astier, M. & Desnuelle, P. (1974). The primary structure of porcine colipase II. II. The disulphide bridges. *Biochim. Biophys. Acta.* **359**, 198–203.

Fletcher, C. M., Jones, D. N. M., Diamond, R. & Neuhaus, D. (1996). Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures. *J. Biomol. NMR,* **8**, 292–310.

Frishman, D. & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins: Struct. Funct. Genet.* **23**, 566–579.

Gasparini, S., Kiyatkin, B., Drevet, P., Boulain, J. C., Tacnet, F., Ripoche, P., Forest, E., Grishin, E. & Ménez, A. (1994). The low molecular weight protein which co-purifies with alpha-latrotoxin is structurally related to crustacean hyperglycemic hormones. *J. Biol. Chem.* **269**, 19803–19809.

Gilson, M. K., Sharp, K. A. & Honing, B. H. (1987). Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comput, Chem.* **9**, 327–335.

Gray, W. R. (1993). Echistatin, Disulphide bridges: selective reduction and linkage assignment. *Proteins Sci.* **2**, 1749–1755.

Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R. R. (1988). Clean TOCSY for ^{1}H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110**, 7870–7872.

Halliwell, J. V., Othman, I. B., Pelchen-Matthews, A. & Dolly, J. O. (1986). Central action of dendrotoxin: selective reduction of a transient K conductance in hippocampus and binding to localized acceptors. *Proc. Natl Acad. Sci. USA,* **83**, 493–497.

Harvey, A. L. (1991). *Snake Toxins*, Pergamon, New York.

Hatanaka, H., Oka, M., Kohda, D., Tate, S., Suda, A., Tamiya, N. & Inagaki, F. (1994). Tertiary structure of erabutoxin b in aqueous solution as elucidated by two-dimensional nuclear magnetic resonance. *J. Mol. Biol.* **240**, 155–166.

Heck, S. D., Kelbaugh, P. R., Kell, M. E., Thadeio, P. F., Saccomano, N. A., Stroh, J. G. & Volkman, R. A. (1994). Disulfide bind assignment of ω-agatoxins IVB and IVC: discovery of a D-serine residue in ω-agatoxin IVB. *J. Am. Chem. Soc.* **116**, 10426–10436.

Hermoso, J., Pignol, D., Kerfelec, B., Crenon, I., Chapus, C. & Fontecilla-Camps, J. C. (1996). Lipase activation by nonionic detergents. The crystal structure of the porcine lipase-colipase-tetraethylene glycol monooctyl ether complex. *J. Biol. Chem.* **271**, 18007–18016.

Hermoso, J., Pignol, D., Penel, S., Roth, M., Chapus, C. & Fontecilla-Camps, J. C. (1997). Neutron crystallographic evidence of lipase-colipase complex activation by a micelle. *EMBO J.* **16**, 5531–5536.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.

Housset, D. & Fontecilla-Camps, J. C. (1996). The structures and evolution of snake toxins of the three-finger folding type. In *Protein Toxin Structure* (Parker, M., ed.), pp. 271–289, Springer, New York.

Joubert, F. J. & Strydom, D. J. (1980). Snake venom. The amino acid sequence of protein A from *Dendroaspis polylepis polylepis* (black mamba) venom. *Hoppe Seylers Z. Physiol. Chem.* **361**, 1787–1794.

Joubert, F. J., Haylett, T., Strydom, D. J. & Taljaard, N. (1982). Snake venom: protein CM2 from *Bitis arietans* (puff adder) venom. *Hoppe Seylers Z. Physiol. Chem.* **363**, 1087–1096.

Klaus, W., Broger, C., Gerber, P. & Senn, H. (1993). Determination of the disulphide bonding pattern in proteins by local and global analysis of nuclear magnetic resonance data. Application to flavoridin. *J. Mol. Biol.* **232**, 897–906.

Kochva, E. (1987). The origin of snakes and evolution of the venom apparatus. *Toxicon,* **25**, 65–106.

Kolbe, H. V. J., Huber, A., Cordier, P., Rasmussen, U. B., Bouchon, B., Jaquinod, M., Vlasak, R., Délot, E. & Kreil, G. (1993). Xenoxins, a family of peptides from dorsal gland secretion of *Xenopus laevis* related to snake venom cytotoxins and neurotoxins. *J. Biol. Chem.* **268**, 16458–16464.

Kraulis, P. K. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Kumar, A., Ernst, R. R. & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1–6.

Lambeau, G., Barhanin, J., Schweitz, H., Qar, J. & Lazdunski, M. (1989). Identification and properties of very high affinity brain membrane-binding sites for a neurotoxic phospholipase from the Taipan venom. *J. Biol. Chem.* **264**, 11503–11520.

Lambeau, G., Ancian, P., Barhanin, J. & Lazdunski, M. (1994). Cloning and expression of a membrane receptor for secretory phospholipases A_2. *J. Biol. Chem.* **269**, 1575–1578.

Lambeau, G., Cupillard, L. & Lazdunski, M. (1997). Membrane receptors for venom phospholipases A_2. In *Venom Phospholipase A_2 enzymes: Structure, Functions and Mechanism* (Kini, R. M., ed.), pp. 389–412, J. Wiley and Sons, Chichester.

Lancelin, J.-M., Foray, M.-F., Poncin, M., Hollecker, M. & Marion, D. (1994). Proteinase inhibitor homologues as potassium channel blockers. *Nature Struct. Biol.* **1**, 246–250.

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

Le Du, M. H., Marchot, P., Bougis, P. E. & Fontecilla-Camps, J. C. (1992). 1.9 Å resolution structure of fasciculin-1, and anti-acetylcholinesterase toxin from green mamba snake venom. *J. Biol. Chem.* **267**, 22122–22130.

Lowe, M.-E. (1997). Colipase stabilizes the lid domain of pancreatic triglyceride lipase. *J. Biol. Chem.* **272**, 9–12.

Medvedeva, S., Simorre, J.-P., Brutscher, B., Guerlesquin, F. & Marion, D. (1993). Extensive ^{1}H NMR resonance assignment of proteins using natural abundance gradient-enhanced ^{13}C-^{1}H correlation spectroscopy. *FEBS Letters,* **333**, 251–256.

Ménez, A. (1995). Les venins et toxines de serpents. In *La Fonction Venimeuse* (Goyffon, M. & Heurtault, J., eds), pp. 200–220, Masson, Paris.

Ménez, A., Bontems, F., Roumestan, C., Gilquin, B. & Toma, F. (1992). Structural basis for functional diversity of animal toxins. *Proc. Roy. Soc. Edinburgh, ser. B,* **99**, 83–103.

Nilges, M. (1995). Calculation of proteins structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660.

Ohno, M., Ménez, R., Ogawa, T., Danse, J. M., Shimohigashi, Y., Fromen, C., Ducancel, F., Zinn-Justin, S., Le Du, M. H., Boulain, J. C., Tamiya, T. & Ménez, A. (1998). Molecular evolution of snake toxins: is the functional diversity of snake toxins associated with a mechanism of accelerated evolution? *Prog. Nucl. Acid. Res. Mol. Biol.* **59**, 307–364.

Olamendi-Portugal, T., Gomez-Lagunas, F., Gurrola, G. B. & Possani, L. D. (1996). A novel structural class of K^+-channel blocking toxin from the scorpion *Pandinus imperator*. *Biochem. J.* **315**, 977–981.

Oswald, R. E., Sutcliffe, M. J., Bamberger, M., Loring, R. H., Braswell, E. & Dobson, C. M. (1991). Solution structure of neuronal bungarotoxin determined by two-dimensional NMR spectroscopy: sequence-specific assignments, secondary structure, and dimer formation. *Biochemistry,* **30**, 4901–4909.

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

Rees, B., Samama, J. P., Thierry, J. C., Gilbert, M., Fischer, J., Schweitz, H., Lazdunski, M. & Moras, D. (1987). Crystal structure of a snake venom cardiotoxin. *Proc. Natl Acad. Sci. USA,* **84**, 3132–3136.

Redfield, C. (1993). Resonance assignment strategies for small proteins. In *NMR of Macromolecules* (Roberts, G. C. K., ed.), pp. 71–100, IRL, Oxford.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.

Schweitz, H., Bidard, J.-N. & Lazdunski, M. (1990). Purification and pharmacological characterization of peptide toxins from the black mamba (*Dendroaspis polylepis*) venom. *Toxicon,* **28**, 847–856.

Schweitz, H., Heurteaux, C., Bois, P., Moinier, D., Romey, G. & Lazdunski, M. (1994). Calcicludine, a venom peptide of the Kunitz-type protease inhibitor family, is a potent blocker of high-threshold Ca^{2+} channels with a high affinity for L-type channels in cerebellar granule neurons. *Proc. Natl Acad. Sci. USA,* **91**, 878–882.

Schwöbel, T. (1997). Simultaneous handling of several NMR spectra of proteins. Diplomarbeit, University of Karlsruhe.

Sternby, B., Engstrom, A. & Hellmann, U. (1984). Purification and characterization of pancreatic colipase from dogfish (*Squalus acanthius*). *Biochim. Biophys. Acta,* **789**, 159–163.

Sunnerhagen, M., Nilges, M., Otting, G. & Carey, J. (1997). Solution structure of the DNA-binding domain and model for the complex of multifunctional hexameric arginine repressor with DNA. *Nature Struct. Biol.* **4**, 819–826.

van Tilbeurgh, H., Sarda, L., Verger, R. & Cambillau, C. (1992). Structure of the pancreatic lipase-procolipase complex. *Nature,* **359**, 159–162.

van Tilbeurgh, H., Egloff, M.-P., Martinez, C., Rugani, N., Verger, R. & Cambillau, C. (1993). Interfacial activation of the lipase-procolipase complex by mixed micelles revealed by X-ray crystallography. *Nature,* **362**, 814–820.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984). A new force field for molecular mechanical representation of nucleics acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784.

Wishart, D. S. & Sykes, B. D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of proteins secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR,* **4**, 171–180.

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

Wüthrich, K., Billeter, M. & Braun, W. (1983). Pseudo-structures for the 20 common amino acids for use in studies of proteins conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949–961.

Edited by P. E. Wright

(Received 27 April 1998; accepted 2 July 1998)

http://www.hbuk.co.uk/jmb

Supplementary material comprising one Table is available from JMB Online