Refined three-dimensional solution structure of insect defensin A

Bruno Cornet[1], Jean-Marc Bonmatin[1], Charles Hetru[2], Jules A Hoffmann[2], Marius Ptak[1,3] and Françoise Vovelle[1,3]*

[1]Centre de Biophysique Moléculaire (CNRS), rue Charles Sadron, 45071 Orléans Cedex 2, France, [2]Institut de Biologie Moléculaire et Cellulaire, UPR 9022 (CNRS), 15 rue Descartes, 67084 Strasbourg Cedex, France and [3]Université d'Orléans, rue Charles Sadron, 45071 Orléans Cedex 2, France

Background: Insect defensin A is a basic 4 kDa protein secreted by *Phormia terranovae* larvae in response to bacterial challenges or injuries. Previous biological tests suggest that the bacterial cytoplasmic membrane is the target of defensin A. The structural study of this protein is the first step towards establishing a structure–activity relationship and forms the basis for understanding its antibiotic activity at the molecular level.

Results: We describe a refined model of the three-dimensional structure of defensin A derived from an extensive analysis of 786 inter-proton nuclear Overhauser effects. The backbone fold involves an N-terminal loop and an α-helical fragment followed by an antiparallel β-structure. The helix and the β-structure are connected by two of the three disulphide bridges present in defensin A, forming a so-called 'cysteine-stabilized αβ' (CSαβ) motif. The N-terminal loop, which is locally well defined, can occupy different positions with respect to the other moieties of the molecule.

Conclusions: The CSαβ motif, which forms the core of the defensin A structure, appears to be a common organization for several families of small proteins with toxic properties. The distribution of amino acid side chains in the protein structure creates several hydrophobic or hydrophilic patches. This leads us to propose that the initial step in the action of positively charged defensin A molecules with cytoplasmic membranes may involve interactions with acidic phospholipids.

Structure 15 May 1995, **3**:435–448

Key words: antibacterial peptides, cysteine-stabilized αβ (CSαβ) motif, insect defensins, NMR, simulated annealing

Introduction

Insects produce a battery of antimicrobial proteins in response to bacterial challenges or injury. These proteins include, in addition to the ubiquitous lysozyme, four families of various sizes [1]. The first of these, the cecropin family, consists of 4 kDa α-helical proteins effective against both Gram-positive and Gram-negative bacteria [2–4]. The second, proline-rich, family comprises small (2–4 kDa) proteins and includes apidaecin [5], abaecin [6] and glycosylated species such as drosocins [7]. The third, glycine-rich, family consists of 8–27 kDa proteins such as attacins [8], diptericins [9] and coleoptericins [10]. The proline-rich and glycine-rich peptides are active against Gram-negative bacteria. The fourth family, which includes royalysin, insect defensins and sapecins, are 4–5 kDa proteins principally active against Gram-positive bacteria. Royalysin, produced in the honey, forms part of the bee's defence against bacterial invasion [11]. Inducible insect defensins are secreted into the insect hemolymph by the fat body cells [12], and a dozen active proteins have been characterized to date [1]. Sapecins, isolated from the culture medium of NIH-sape-4-cells [13], belong to the same family. Defensin A, isolated from the larvae of the flesh fly *Phormia terranovae*, and sapecin, isolated from the flesh fly *Sarcophaga peregrina*, are cationic peptides of 40 amino acid residues that differ only at position 34 [^{1}ATCDLLSGTGINHSACAAHCLLRGNRGGYCNG KG(A)VCVCRN^{40}]. Both peptides contain six conserved cysteine residues that form a stabilizing array of three disulphide bridges [14,15].

The activity of defensin A against Gram-positive bacteria was recently investigated using a strain of *Micrococcus luteus* as a test organism. Induced K^+ leakage, probably resulting from the formation of voltage-dependent channels and from partial depolarization of the cytoplasmic membrane, followed by ATP depletion and inhibition of respiration [16], indicates that the membrane is a potential target for defensins. In addition, it was shown that sapecins have an affinity for cardiolipin, a major acidic phospholipid constituent of some bacterial membranes [17].

Preliminary models of defensin A structure in water [18,19], and of sapecin structure in methanol [20], have been derived from two-dimensional (2D) ^{1}H NMR studies. Their overall folds involve an N-terminal loop and a short amphipathic α-helix followed by a C-terminal antiparallel β-structure. This architecture clearly differs from that of mammalian defensins in solution. Mammalian defensins lack helical structure and are organized as dimers of β-sheet structures [21,22]. Two of the three disulphide bridges present in insect defensins connect the helix and the β-sheet, forming a characteristic

*Corresponding author.

structural motif with similarities to the 'cysteine-stabilized α-helical' (CSH) motif identified previously [23]. Interestingly, this motif is also common to several peptides and proteins (scorpion toxins, plant γ-thionins) that exhibit various toxic activities.

In this report, we describe a refined model of the 3D structure of *P. terranovae* defensin A in water, based on detailed analyses of NMR constraints and on extensive sampling of conformational space using simulated annealing methods. Comparison with available structures of scorpion toxins and plant γ-thionins from the Protein Data Bank (PDB) [24,25] leads us to propose the αβ structure stabilized by two disulphide bridges as a general type of organization for several families of proteins.

Results

Structural statistics

Using an improved set of 786 constraints derived from inter-proton nuclear Overhauser effects (NOEs) (see the Materials and methods section), 50 structures were randomly generated and subjected to simulated annealing (SA) and energy minimization (EM). Seven structures that exhibited 'wrong' folds and large constraint violations were eliminated. The remaining 43 were considered consistent with NMR data (Table 1) because no violation >0.7 Å and only few violations >0.5 Å were found in annealed structures. The NMR reliability factor (R*f*), as proposed by James *et al.* [26], averaged over the NOE-constrained EM structures was 11%. An analysis of potential-energy contributions (Table 1) indicated favourable non-bonded and electrostatic interactions. The average deviations from ideal covalent geometry were very low (Table 1).

Description of the structure

In agreement with our preliminary model [18,19], the backbone fold of our refined model (Fig. 1) includes an N-terminal loop (residues 4–14) and an α-helix (residues 15–23) followed by an antiparallel β-sheet (residues 27–31 and 35–39) including an unusual three-residue turn connecting the two strands. The 43 structures are divided into two sets, A and B, on the basis of differences in the positions occupied by the N-terminal loop (see below).

According to backbone torsion-angle values (Table 2) and to energy and geometry criteria used to define hydrogen bonds (Table 3), residues 15–23 constitute a helix. Average (ϕ,ψ) values (from ϕ15 to ψ23) of

Table 1. Structural characteristics of the two sets of structures.†

Parameter	Structure set A	Structure set B	Structure set A+B
Number of structures	33	10	43/50
Number of constraints			
NOEs			786/904
H-bonds			13
Disulphide bonds			9
NOE violations			
Rmsd* (Å)	0.33±0.03 (0.27:0.38)	0.34±0.02 (0.29:0.37)	0.33±0.02 (0.27:0.38)
Number >0.5 Å*	8±2 (5:12)	9±2 (6:14)	8±2 (5:14)
R-factor (%)	11±1 (–380:–178)	11±1 (10:12)	11±1 (10:12)
X-PLOR energy (kJ mol^{-1})			
Total energy	–476±52 (–568:–367)	–443±56 (–535:–363)	–468±56 (–568:–363)
van der Waals energy	–681±16 (–727:–660)	–693±24 (–735:–656)	–685±19 (–785:–656)
Electrostatic energy	–443±25 (–497:–388)	–443±26 (–484:–405)	–443±25 (–497:–388)
Covalent geometry			
Bond length deviation (×10^{-4} Å)	89±5 (79:99)	89±4 (81:99)	89±5 (79:99)
Valence angle deviation (°)	2.41±0.05 (2.31:2.53)	2.44±0.03 (2.38:2.49)	2.41±0.05 (2.31:2.53)
Improper angle deviation (°)	0.20±0.02 (0.16:0.26)	0.20±0.01 (0.18:0.21)	0.20±0.02 (0.16:0.26)
Average rmsd values (Å)			
N, Cα, C′, for residues 6–11	0.6±0.2 (0.3:1.0)	1.1±0.2 (0.8:1.4)	0.7±0.2 (0.3:1.4)
N, Cα, C′, for residues 15–39	0.4±0.1 (0.2:0.8)	0.6±0.1 (0.4:0.8)	0.4±0.1 (0.2:0.8)
N, Cα, C′, for residues 1–40	0.9±0.2 (0.6:1.3)	2.7±0.7 (1.4:3.6)	1.3±0.8 (0.6:3.6)
Heavy atoms	1.3±0.2 (0.9:1.7)	3.0±0.6 (1.9:3.8)	1.7±0.8 (1.0:3.8)
All atoms	1.6±0.2 (1.2:2.0)	3.3±0.6 (2.2:4.1)	2.0±0.8 (1.2:4.1)
Favoured conformations (%)			
For residues 1–40	62±6 (52:77)	61±6 (48:71)	62±6 (48:77)
For residues 15–40	84±5 (75:90)	78±7 (65:90)	83±6 (65:90)

†The 43 structures are divided into two sets, A and B, on the basis of differences in the positions occupied by the N-terminal loop. All values are given for the energy-minimized structures except those marked by an asterisk, which are given for the simulated annealing structures. The numbers given are the average values±standard deviation, with the minimum and maximum values in parentheses. The rmsd values are given relative to the mean structure of the set of structures and calculated for the given part of the protein. The total energy does not include the constraint energy term.

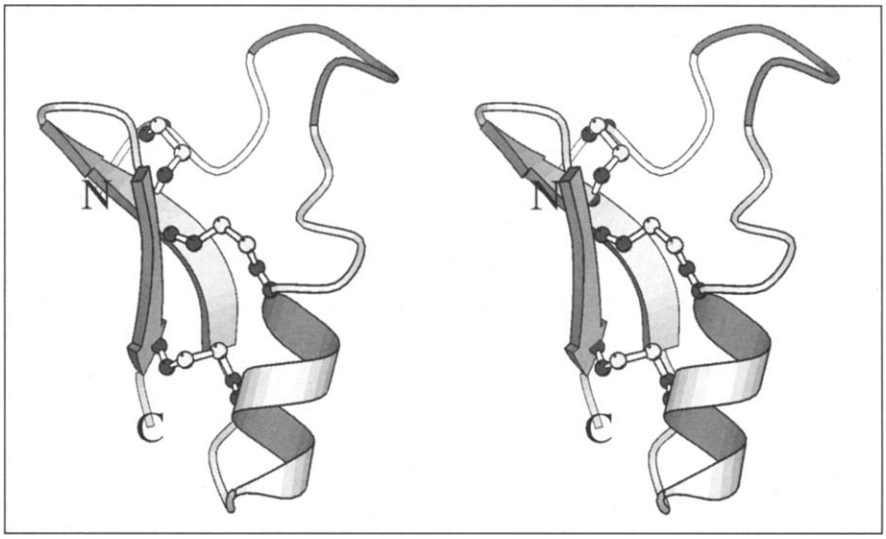

Fig. 1. Schematic stereoview of defensin A (drawn with MOLSCRIPT [55]).

Table 2. Backbone and side-chain torsion angles.

Residue	ϕ	ψ	χ^1	χ^2	χ^{ss}
Ala1	–	160 ± 89 (0.79)	–	–	–
Thr2	−119 ± 36 (0.12)	−178 ± 83 (0.64)	42 ± 63 (0.43)	–	–
Cys3	−119 ± 33 (0.11)	74 ± 64 (0.43)	−162 ± 42 (0.22)	−141 ± 34 (0.16)	−42 ± 61 (0.33)
Asp4	−126 ± 57 (0.29)	141 ± 61 (0.44)	−138 ± 80 (0.53)	–	–
Leu5	−9 ± 76 (0.66)	81 ± 52 (0.33)	−72 ± 23 (0.08)	163 ± 21 (0.06)	–
Leu6	−66 ± 62 (0.43)	−68 ± 15 (0.03)	−33 ± 5 (0.00)	−172 ± 6 (0.01)	–
Ser7	−131 ± 23 (0.06)	136 ± 64 (0.50)	−114 ± 78 (0.51)	–	–
Gly8	6 ± 74 (0.66)	−58 ± 58 (0.35)	–	–	–
Thr9	−110 ± 50 (0.26)	53 ± 39 (0.17)	63 ± 25 (0.07)	–	–
Gly10	−99 ± 47 (0.20)	−119 ± 52 (0.37)	–	–	–
Ile11	−71 ± 43 (0.27)	−26 ± 30 (0.13)	63 ± 57 (0.35)	75 ± 70 (0.37)	–
Asn12	48 ± 39 (0.22)	58 ± 66 (0.51)	169 ± 23 (0.07)	31 ± 70 (0.50)	–
His13	−70 ± 85 (0.66)	17 ± 77 (0.55)	−29 ± 57 (0.42)	−139 ± 55 (0.40)	–
Ser14	−29 ± 66 (0.48)	−18 ± 55 (0.39)	−145 ± 18 (0.05)	–	–
Ala15	−84 ± 61 (0.44)	−57 ± 5 (0.00)	–	–	–
Cys16	−57 ± 10 (0.01)	−52 ± 6 (0.01)	−40 ± 71 (0.41)	−178 ± 52 (0.34)	47 ± 52 (0.33)
Ala17	−51 ± 7 (0.01)	−57 ± 17 (0.04)	–	–	–
Ala18	−45 ± 17 (0.04)	−58 ± 2 (0.00)	–	–	–
His19	−49 ± 4 (0.00)	−56 ± 2 (0.00)	−177 ± 3 (0.00)	88 ± 3 (0.00)	–
Cys20	−54 ± 6 (0.01)	−52 ± 3 (0.00)	−53 ± 25 (0.05)	−77 ± 27 (0.05)	−61 ± 8 (0.01)
Leu21	−49 ± 4 (0.00)	−56 ± 2 (0.00)	−91 ± 6 (0.01)	−91 ± 38 (0.18)	–
Leu22	−46 ± 4 (0.00)	−54 ± 3 (0.00)	−88 ± 4 (0.00)	151 ± 9 (0.01)	–
Arg23	−54 ± 7 (0.01)	−48 ± 2 (0.00)	−91 ± (0.04)	−146 ± 43 (0.23)	–
Gly24	131 ± 3 (0.00)	53 ± 4 (0.00)	–	–	–
Asn25	−127 ± 6 (0.01)	178 ± 6 (0.01)	−56 ± 69 (0.37)	90 ± 80 (0.60)	–
Arg26	−134 ± 7 (0.01)	28 ± 27 (0.10)	49 ± 37 (0.10)	−115 ± 48 (0.22)	–
Gly27	−129 ± 9 (0.01)	−177 ± 35 (0.17)	–	–	–
Gly28	153 ± 45 (0.26)	72 ± 13 (0.02)	–	–	–
Tyr29	−136 ± 8 (0.01)	−175 ± 6 (0.01)	69 ± 8 (0.01)	7 ± 87 (0.93)	–
Cys30	−117 ± 6 (0.00)	160 ± 9 (0.01)	12 ± 31 (0.12)	95 ± 51 (0.24)	−42 ± 61 (0.33)
Asn31	−121 ± 10 (0.01)	162 ± 9 (0.01)	87 ± 11 (0.02)	−70 ± 30 (0.10)	–
Gly32	−4 ± 12 (0.01)	−72 ± 27 (0.07)	–	–	–
Lys33	−43 ± 36 (0.14)	−40 ± 25 (0.08)	−52 ± 72 (0.54)	101 ± 92 (0.60)	–
Gly34	110 ± 26 (0.09)	42 ± 8 (0.01)	–	–	–
Val35	−129 ± 5 (0.00)	160 ± 5 (0.00)	−177 ± 13 (0.03)	–	–
Cys36	−85 ± 10 (0.01)	142 ± 10 (0.02)	−99 ± 33 (0.14)	−135 ± 73 (0.47)	47 ± 52 (0.33)
Val37	−137 ± 5 (0.00)	157 ± 7 (0.01)	−161 ± 26 (0.10)	–	–
Cys38	−113 ± 5 (0.00)	143 ± 16 (0.04)	−56 ± 18 (0.04)	−56 ± 26 (0.05)	−61 ± 8 (0.01)
Arg39	−134 ± 9 (0.01)	173 ± 16 (0.04)	−131 ± 65 (0.49)	−97 ± 104 (0.82)	–
Asn40	−123 ± 10 (0.02)	–	−15 ± 42 (0.25)	45 ± 81 (0.71)	–

Average values, standard deviations and circular variances (in parentheses) are given for the 43 structures.

Table 3. Statistics on hydrogen bonds of defensin A.

Secondary structure	Donor		Acceptor		Occurrence
Extended N-terminal	Cys3	HN	Cys30	O	9
segment	Cys30	HN	Cys3	O	25
β-turn #8–9	Gly10	HN	Ser7	O	8
α-helix	Cys16	HN	His13	O	10
	Ala17	HN	His13	O	18
	Ala18	HN	Ser14	O	19
	His19*	HN	Ala15	O	32
	Cys20*	HN	Cys16	O	33
	Leu21*	HN	Ala17	O	32
	Leu22*	HN	Ala18	O	33
	Arg23*	HN	His19	O	33
	Gly24*	HN	Cys20	O	33
	Asn25*	HN	Leu21	O	NF
Turn 1	Asn25	HN	Cys20	O	21
	Asn25	HD21	Cys20	O	23
β-strand 1	Arg26	HN	Arg39	O	28
	Gly27*	HN	Arg39	O	3
	Tyr29*	HN	Val37	O	30
	Asn31*	HN	Val35	O	32
	Asn31	HD21	Val35	O	6
Turn 2	Gly34	HN	Asn31	O	24
β-strand 2	Val35*	HN	Asn31	O	27
	Val37*	HN	Tyr29	O	33
	Arg39*	HN	Gly27	O	NF

Values are given for the 33 structures of the A set. We give the number of structures with hydrogen bonding potential energy ≤−2.1 kJ mol^{-1}. (X-PLOR hydrogen bond potential energy takes into account the distance between donor and acceptor atoms and the directionality of the bond.) NF, not found. Asterisks indicate hydrogen bonds used as constraints during modelling.

−51±9°, −54±9° are typical of such structures. The existence of a regular i→i+4 hydrogen-bond network extending from the carbonyl oxygen of residue 13 to the amide hydrogen of residue 24 (O13 to NH24) in most structures as well as the (ϕ,ψ) angles suggest that this helix has an α_R conformation.

The C-terminal fragment (27–40) is a twisted two-stranded antiparallel β-sheet. Characteristic hydrogen bonds ($NH^{29}\rightarrow O^{37}$, $NH^{31}\rightarrow O^{35}$, $NH^{35}\rightarrow O^{31}$, $NH^{37}\rightarrow O^{29}$; see Table 3), as well as characteristic backbone torsion angles [27], are found in the central part of this β-sheet in the majority of the 33 structures of the A set: odd-numbered residues forming hydrogen bonds (Tyr29, Asn31, Val35, Val37 and Arg39) have low ϕ values (−131±9°) and high ψ values (170±16°); even-numbered residues that do not form hydrogen bonds (Cys30, Cys36 and Cys38) show an opposite trend (ϕ=−105±16°,ψ=149±15°). NH^{26} and NH^{27} form a branched hydrogen bond with O^{39}, whereas no $NH^{39}\rightarrow O^{27}$ hydrogen bond is formed, because NH^{39} is oriented towards the interior of the β-sheet and O^{27} points out towards the helix. So, the slow exchange rate of NH^{39} is more probably attributable to its low level of accessibility than to its involvement in hydrogen bonding. The presence of the Gly27–Gly28 pair introduces local distortions, because (ϕ,ψ) angles of Gly28 residues have non-canonical values of (−120°,90°) or (120°,90°).

The three-residue turn connecting the helix to the β-sheet is very well defined and corresponds to the 'α-β-arch' described by Efimov [28] as a standard structure in proteins. The helix is interrupted by the Gly24 residue in α_L conformation followed by two residues, Asn25 and Arg26, residues in β and γ_L conformations, respectively. This turn is stabilized by an $NH^{25}\rightarrow O^{20}$ hydrogen bond, accounting for the low temperature coefficient of NH^{25} which was initially attributed to an $NH^{25}\rightarrow O^{21}$ hydrogen bond (Table 3). The Asn25 side chain is inserted between the helix and the first strand of the sheet. This reduces its accessibility to the solvent considerably (Fig. 2a). Its position is well defined experimentally by 28 NOEs involving β,β′ protons and 22 NOEs involving δ,δ′ protons. Both sets of protons are stereospecifically assigned. This insertion of Asn25 is compatible with different values for the χ1 and χ2 angles (Table 2).

The three-residue turn connecting the two strands of the β-sheet exhibits a very unusual $\alpha_R\alpha_R\alpha_L$ conformation stabilized by two hydrogen bonds, $HN^{35}\rightarrow O^{31}$ and $HN^{34}\rightarrow O^{31}$. The peptide backbone in this turn adopts a position almost perpendicular to the average plane of the β-sheet.

The (4–14) loop is poorly defined experimentally, presumably due to its high flexibility. However, the (6–11) fragment exhibits very similar conformations in most of the 43 structures (average root mean square deviation [rmsd] of ~0.7 Å; see Fig. 3a). In 12 structures, a type I β-turn is stabilized by an $NH^{10}\rightarrow O^{7}$ hydrogen bond. As discussed below, the (4–14) loop seems to undergo a rotation about two hinges. In most structures, residues 1 and 2 preceding this loop adopt roughly the same orientation, parallel to the first strand of the β-sheet. This local conformation is stabilized by the Cys3–Cys30 disulphide bridge and also by two hydrogen bonds involving residues 3 and 30 (see Table 3).

In order to analyze the side-chain conformations, χ1 and χ2 circular variances (CVs) were calculated (Table 2). CV values <0.2 correspond to a tight clustering of χ values and are indicative of a staggered conformation, whereas larger values generally correspond to multiple conformations. PROCHECK [29] analysis of our structures (data not shown) indicates that even for larger CV values, the conformations fall in favourable regions of the χ1, χ2 maps. The solvent accessibility [30] for each residue is presented in Figure 2a. In a small protein like defensin A, most side chains are external and there is no direct correlation between accessibilities and dispersion of side-chain conformations as seen from the

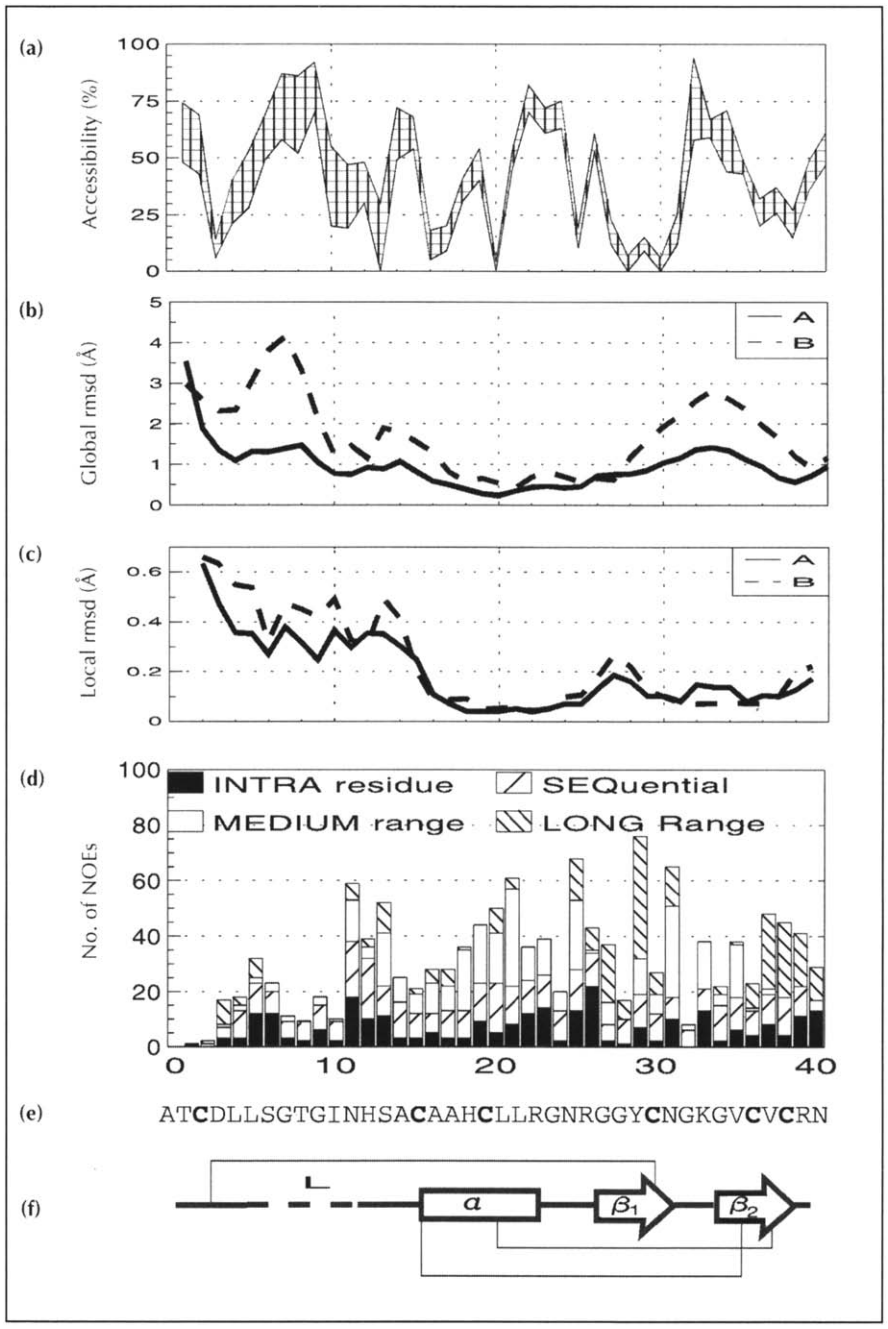

Fig. 2. Structural data for the two sets of defensin A structures plotted as a function of sequence. **(a)** Accessibility, (i.e. the ratio of average±rms accessible surface for the refined structure set) *versus* the maximum accessible surface of the initial set of (ϕ,ψ) randomized structures. **(b)** Average rmsds from the mean structure for backbone heavy atoms (N, Cα and C′). Data for the two sets of structures are plotted separately. **(c)** Three-residue average rmsds from the mean structure for backbone heavy atoms (N, Cα and C′). Data for the two sets of structures are plotted separately. **(d)** NOE distribution. Only the 786 NOEs used in the final step of modelling are represented. **(e)** Amino acid sequence. The six cysteine residues are show in bold type. **(f)** Schematic representation of the secondary structure of defensin A. β-strands are represented by arrows, the α-helix by a rectangle and the N-terminal loop by a broken line. The 3–30, 16–36 and 20–38 disulphide bridges are shown by thin lines.

CV [31]. The two histidine residues, His13 and His19, exhibit very different conformational behaviour. His13, located in the second hinge (see below), shows large CVs (CVχ1=0.42; CVχ2=0.40). His19, located in the well-defined helical part of the structure, is characterized by very low CVs (CVχ1=CVχ2=0.00). Side chains with high mobilities are found in the turn connecting the two β-strands (Lys33) and in the N-terminal part (Arg39, Asn40). The high CVχ2 value of Tyr29 (0.93) is accounted for by the fact that the aromatic plane of the residue occupies the same position in the 43 structures but the χ2 values differ by 180°.

As previously described [18,19], small hydrophobic patches (Leu5/Leu6, Ala15/Ala17/Ala18, Leu21/Leu22, Val35/Val37) are distributed within the primary structure. The Tyr29 side chain is on the same side of the β-sheet as the bulky γ-methyl groups of Val35 and Val37. The helix possesses an amphipathic character; the charged residues, His19 and Arg23, point towards the β-sheet, whereas the hydrophobic residues, Ala15, Ala17, Ala18, Leu21 and Leu22, are located on the other face of the helix. The distribution of amino acids on the protein surface is illustrated in Figure 4.

Disulphide bridges

The overall stability of the 3D structure is ensured by the three disulphide bridges (3–30, 16–36 and 20–38). Among the six cysteine residues, four (16, 20, 36, 38) are sufficiently well defined to allow stereospecific assignment

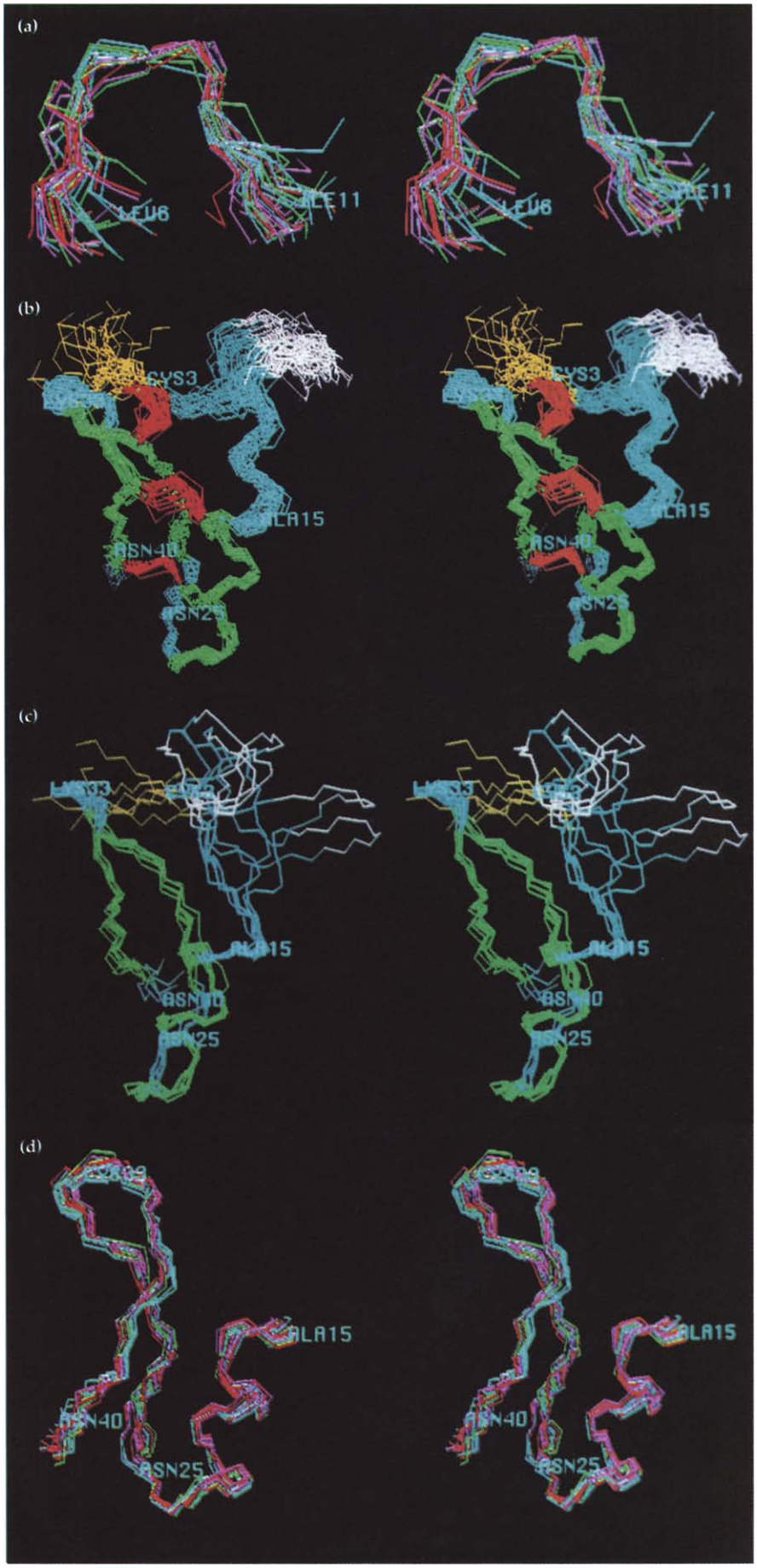

Fig. 3. Stereoviews of fitted structures. Only the backbone heavy atoms (N, Cα and C′) are shown. **(a)** N-terminal loop (6–11) for the 43 structures. **(b)** Superposition of the 33 structures of set A. Helix and strands are coloured green, the three disulphide bridges red and the N-terminal flexible loop white. **(c)** Superposition of the 10 structures of the B set. The colour scheme is the same as for (b). In (b) and (c), the structures are fitted to the αβ motif (residues 15–39). **(d)** Superposition of the αβ motif (15–40) for the 43 structures. (Drawn with SYBYL [56].)

of the β-protons. In spite of the fact that no torsional constraints were introduced in our calculations, the Cys20–Cys38 bridge is defined with good precision in a *gauche*$^+$ conformation (Table 2). The two other bridges, Cys3–Cys30 and Cys16–Cys38, are more poorly defined in *gauche*$^-$ and *gauche*$^+$ conformations respectively, with average χss angle values outside the classical ±90° minima [32].

Structure variability

The high value of the average rmsd calculated from the atomic coordinates of the N, Cα and C′ atoms (1.3±0.8 Å) may be interpreted as a large dispersion of conformations. In fact, the 43 structures found to be in agreement with NMR data can be divided into two sets using rmsd criteria. The first set (A) includes 33 structures exhibiting very similar folds (Fig. 3b), with an average rmsd around 0.9 Å. The average rmsd for the 10 remaining structures (set B) is much larger (2.7 Å). However, a best fit superposition of these structures reveals that they differ only by the orientation of the N-terminal loop (Fig. 3c). This is underlined by the global rmsd plot (see the Materials and methods section) of backbone heavy atoms (Fig. 2b) in which high values are located in the N-terminal part and, to a lesser extent, in the turn of the β-sheet. The local three-residue rmsd plot [33] (Fig. 2c) shows that the N-terminal loop probably has moderate internal flexibility as confirmed by the superposition of the 43 structures of the fragment comprising residues 6–11 (average rmsd of 0.7 Å; see Fig. 3a). This loop can adopt different orientations with respect to the rest of the molecule, the transition involving residues 4/5 and 12–15. The variability of the structures is also illustrated by the (ϕ,ψ) CVs (Table 2), which are especially high in the N-terminal region. Moreover, the ratio of residues lying in the most favourable regions of the Ramachandran plot (Table 1; Fig 5) indicates that the lack of precision of the structure is mainly confined to the N terminus and particularly to the hinges which are most flexible. On the contrary, the residue 15–40 segment shows a low average rmsd (0.4 Å) and a high percentage (83%) of favourable conformations (Table 1).

Discussion

On the basis of an expanded set of experimental data, the solution structure of defensin A has been refined to high precision, as measured by the rmsd of atomic positions (0.4 Å on the secondary-structure elements). The peptide backbone fold is similar to that found in the preliminary model, but these new calculations provide better insight into the details of the 3D structure. The helical fragment is a well defined regular α_R-helix, whereas the β-sheet exhibits a pronounced right-handed twist and involves an unusual turn formed by three residues having $\alpha_R\alpha_R\alpha_L$ conformations. These two secondary-structure

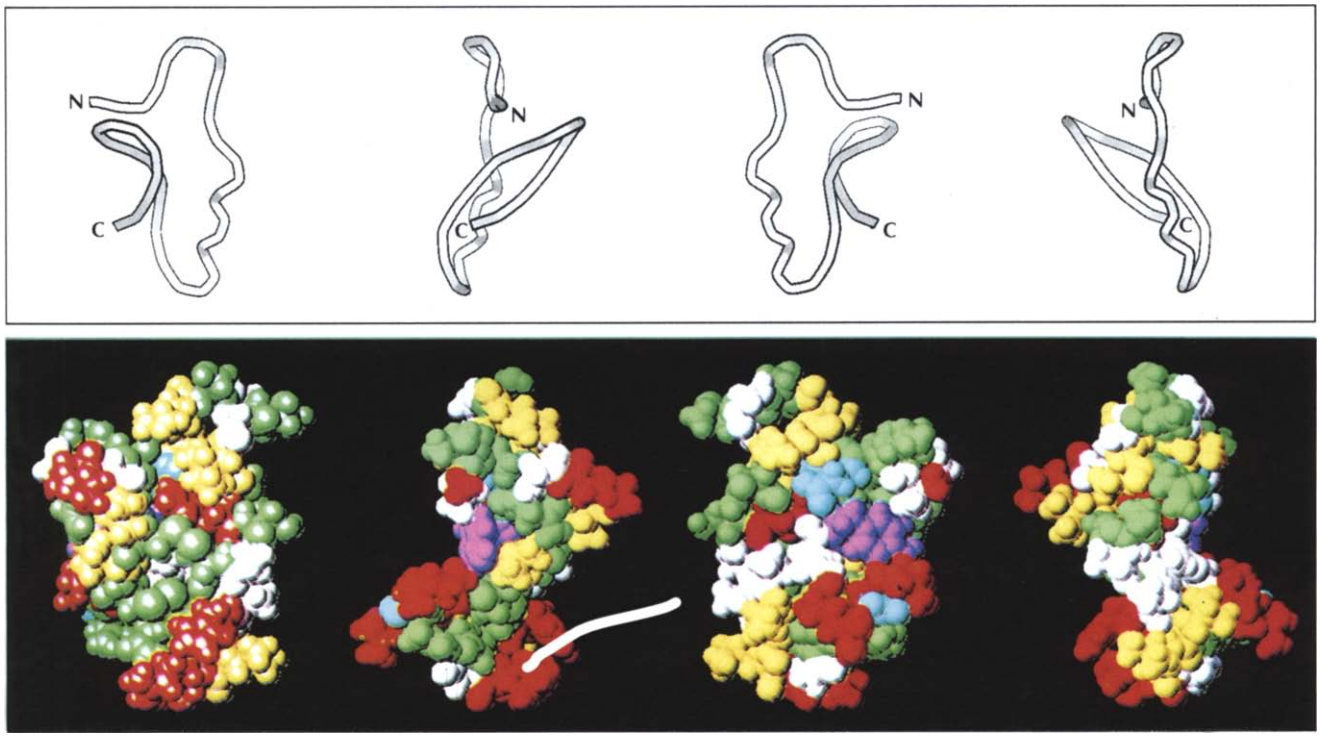

Fig. 4. Distribution of amino acids by type. (Top) Tube representation of the Cα trace of the different orientations shown below. (Bottom) Space-filling view with four different orientations of defensin A. Hydrophobic residues are shown in yellow, polar residues in green, aromatic residues in purple, positively charged residues in red and negatively charged residues in cyan. Glycines are shown in white. Hydrophobic residues are not packed in a core as usually found in larger proteins. Instead, they are distributed over the surface. Positively charged side chains are preferentially located in two regions of the molecule, with Lys33 in one location and Arg23, Arg25 and Arg39 in another.

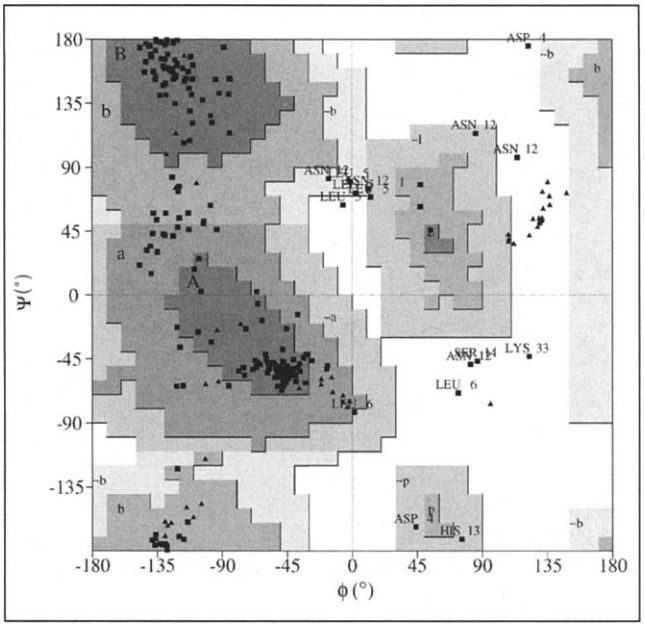

Fig. 5. Ramachandran plot of the six structures having the largest number of residues in the favourable regions of the Ramachandran map. (Drawn with PROCHECK [29].) Glycines are represented by triangles and all other residues by squares. Residues falling outside the favourable regions are labelled. They mostly belong to the N-terminal loop.

elements, locked together by two disulphide bridges, constitute the stable domain of defensin A which, according to their sequence similarities, should be present in most insect defensins. On the other hand, the N-terminal loop exhibits special behaviour by adopting several orientations with respect to the αβ-motif. Other defensins [1] and sapecins [34,35] have deletions or extensions within this loop region which seem to be related to the kinetics of action, the level of activity and antibiotic specificity. For example, *Aeschna* defensin (a member of the Paleoptera order) has a large deletion in the region of the N-terminal loop relative to *Phormia* defensin A [36]. In growth-inhibition assays performed using various bacterial strains, the *Aeschna* defensin appears systematically more active than that of *Phormia*.

Self-association

It was recently demonstrated (R Maget-Dana, personal communication) that defensin A, which is highly soluble in water, is also capable of forming a stable monolayer at the air–water interface. Preliminary circular dichroism (CD) and fluorescence studies (R Maget-Dana and JM Bonmatin, personal communication) of defensin A in water reveal concentration effects (variations of the percentage of helicity and of the polarization ratio, respectively) at values lower than those required for NMR experiments. This suggests that defensin A can self-associate in spite of its high positive charge. The existence of a single species was detected by NMR at acidic pH and millimolar concentrations, as shown by the assignment of each proton to a single peak in 2D maps. The observed line-widths are consistent with a monomer or a small oligomer not larger than a dimer or tetramer (for oligomers possessing symmetry elements). After back-calculations of the nuclear Overhauser effect spectroscopy (NOESY) maps, 30 NOEs that were not compatible with our models were rejected and not used in the molecular modelling of defensin A. These incompatibilities may result from different sources such as local dynamics, spin-diffusion effects or intermolecular contacts, with the assumption that several NOEs come out from intermolecular interactions. Attempts to build a model of self-association using Nilges' method [37] led to several plausible dimeric structures. However, these models could not be validated due to the lack of experimental data. Accordingly, NMR experiments on ^{15}N-labelled defensin A should help to eliminate ambiguities between intramolecular and intermolecular NOEs. The structure presented here must be considered as the basic structure which may self-associate at some specific pH and under certain concentration conditions. Which form of the protein is active at the very low concentrations (usually ~10^{-6} M) present in the biological assays [16] remains to be determined.

Comparison with other proteins

As already mentioned, one characteristic feature of defensin A is the stabilization of its structure by disulphide bridges. Tamaoki *et al.* [23] previously listed CSH motifs in a variety of small peptides and proteins. In this motif, the helical conformation of the sequence Cys-Xaa-Xaa-Xaa-Cys is stabilized by a disulphide bridge to the sequence Cys-Xaa-Cys, belonging to a segment in extended form. Apamin [38] is a simple example of a small peptide formed by only one helical fragment connected to an extended strand. In defensin A, the helical fragment is connected via the two disulphide bridges (16–36, 20–38) to an antiparallel two-stranded β-structure. The overall fold is stabilized by a third (3–30) bridge. The sequence involving the two secondary-structure elements can be summarized as Cys^{I}-$(Xaa)_i$-Cys^{II}-Xaa-Xaa-Xaa-Cys^{III}-$(Xaa)_j$-$Cys^{I'}$-$(Xaa)_k$-$Cys^{II'}$-Xaa-$Cys^{III'}$, where the $(Xaa)_i$ segment corresponds to the N-terminal loop of defensin A, the $(Xaa)_j$ segment to the end of the helix, the first β-strand and the turn between them, and $(Xaa)_k$ to the turn between the two β-strands. Similar motifs in which the β-sheet can involve two or three strands are present in several proteins having very different origins and activities. Therefore, we propose to call them 'cysteine-stabilized αβ' (CSαβ) motifs. Structures of various scorpion toxins and plant γ-thionins sharing the CSαβ motif are compared in Table 4 and their structures are superimposed in Figure 6. According to the general definition of Orengo and Thornton [39], which does not necessarily refer to the presence of disulphide bridges, γ-thionins are classified as βαββ proteins. On the same basis, defensin A is an αββ protein. In spite of their structural similarities, these proteins exhibit different activities. Scorpion toxins inhibit Ca^{2+}-activated K^+-channels through five different binding sites. Defensins and sapecins permeabilize membranes, probably by inducing the formation of voltage-dependent channels [16]. Sapecin B can also inhibit K^+-channels [34,35]. Finally, the activity

Table 4. Comparison of CSαβ motifs from homologous proteins.

Protein name	PDB entry code	Ref.	N	Disulphide bridges		Secondary structures				Segment length		
				Cysteine	Rmsd (Å)	α	β1	β2	Rmsd (Å)	i	j	k
Defensin A	–	–	33	16,20,36,38	–	*14–23	27–31	35–39	–	12	9	5
Charybdotoxin	2crd	[57]	12	13,17,33,35	0.69±0.19	11–20	25–29	32–36	1.38±0.15	5	10	4
PO_5-NH_2 toxin analogue	1pnh	[58]	25	8,12,26,28	0.67±0.15	6–15	18–22	25–29	1.29±0.13	4	8	4
Insectotoxin	1sis	[59]	10	16,20,31,33	0.63±0.14	–	–	–	–	10	5	4
Variant-3 neurotoxin	2sn3	[60]	1	25,29,46,48	0.71±0.15	23–32	38–42	45–49	1.17±0.12	8	11	4
Variant-1 neurotoxin	1vna	[61]	26	25,29,46,48	0.75±0.23	23–32	38–42	45–49	1.50±0.19	8	11	4
γ1-H thionin	1gpt	[62]	8	20,24,41,43	0.59±0.15	18–27	31–35	40–44	0.99±0.17	5	9	6
γ1-P thionin	1gps	[62]	8	20,24,41,43	0.69±0.16	18–27	31–35	40–44	1.18±0.16	5	9	6

Columns from left to right are as follows: protein names; Protein Data Bank (PDB) entry codes; bibliographic references (Ref.); number of structures available (N); cysteine residue numbers; rmsds after best fit superposition on these cysteines; CSαβ secondary structure sequences (residue ranges for each element), rmsds after best fit on these motifs and the number of residues involved in the segments between cysteines (see text for definition of i, j and k indices). *Ser14 is not exactly part of the α-helix but has an approximately right-handed α-helical conformation and is superimposable with other proteins containing the CSαβ motif.

of cytolytic γ-thionins is neither fully understood nor fully characterized, but these proteins are capable of changing membrane permeability and are toxic for bacterial, animal and plant cells [40].

The only conserved residue in the sequences of most proteins containing a CSαβ motif is a glycine, Gly28 in defensin A. This residue, located at the beginning of the first β-strand, seems necessary to allow flexibility between the helix and the β-sheet and its special behaviour was noted previously. Structures of seven proteins containing a CSαβ motif were extracted from the PDB and compared with the structure of defensin A (Table 4; Fig. 6). The α-β-arches [28] (i.e., the turns connecting helices and β-sheet) present conformations very similar to that found in defensin A. In γ-thionins, this turn (Gly24-Asn25-Arg26 for defensin A) is also three residues long and starts with a glycine in an α_L conformation followed by two residues with β and α_R/γ_L conformations. It is stabilized by a special hydrogen-bonding pattern, as in defensin A. In insectotoxin, the helix is one turn shorter and also terminates with a glycine in an α_L conformation. In PO_5-NH_2 toxin analogue, the helix is one residue shorter and the resulting constraint is balanced by an unusual and energetically unfavourable conformation, denoted ϵ by Efimov [41]. In the two neurotoxin variants and charybdotoxin, where the turns are larger, the α_L glycine is replaced by one or several residues in extended conformation. In charybdotoxin and γ-thionins, the side chains of residues Thr23 and Trp29, which correspond to Asn25 in defensin A, are also buried and stabilized by hydrogen bonding.

Three-residue turns connecting the two strands of a β-sheet have been systematically analyzed [42,43]. They are classified into two groups depending on the hydrogen-bonding pattern. When the two strands are connected by canonical (i→i+4) hydrogen bonds, the turn is classified as 3:3. When these bonds are replaced by (i→i+2) hydrogen bonds, the turn is classified as 3:5. These very tight 3:5 turns are unusual and exhibit a variety of conformations, whereas the 3:3 turns are more common and have a well defined $\alpha_R\alpha_R\alpha_L$ conformation. In defensin A, a 3:3 $\alpha_R\alpha_R\alpha_L$ conformation is predominant but the hydrogen-bonding pattern is a mixture of the characteristic 3:3 and 3:5 forms (Fig. 7). Unlike defensin A, scorpion and plant toxins include canonical two-residue β-turns or longer segments connecting the two β-strands.

Biological implications

Defensin A is excreted in the hemolymph of the flesh fly *Phormia terranovae* in response to bacterial challenge or tissue injury [12] and is principally active against Gram-positive bacteria. The antibiotic properties of this 4 kDa protein stem from its ability to induce K^+ leakage in the cytoplasmic membrane [16]. The present work provides a refined model of defensin A, the first member of a large family of proteins involved in the insect immune system to be solved with this precision.

The structure of defensin A includes an N-terminal loop and a small αβ motif stabilized by two disulphide bonds. The N-terminal loop can adopt different orientations in space through a hinge motion. In contrast, the 'cysteine-stabilized αβ' (CSαβ) motif adopts only one conformation. Comparison with other insect defensin sequences, in which deletions or insertions occur in the region of the N-terminal loop, suggests that this region should be important for the activity of these proteins.

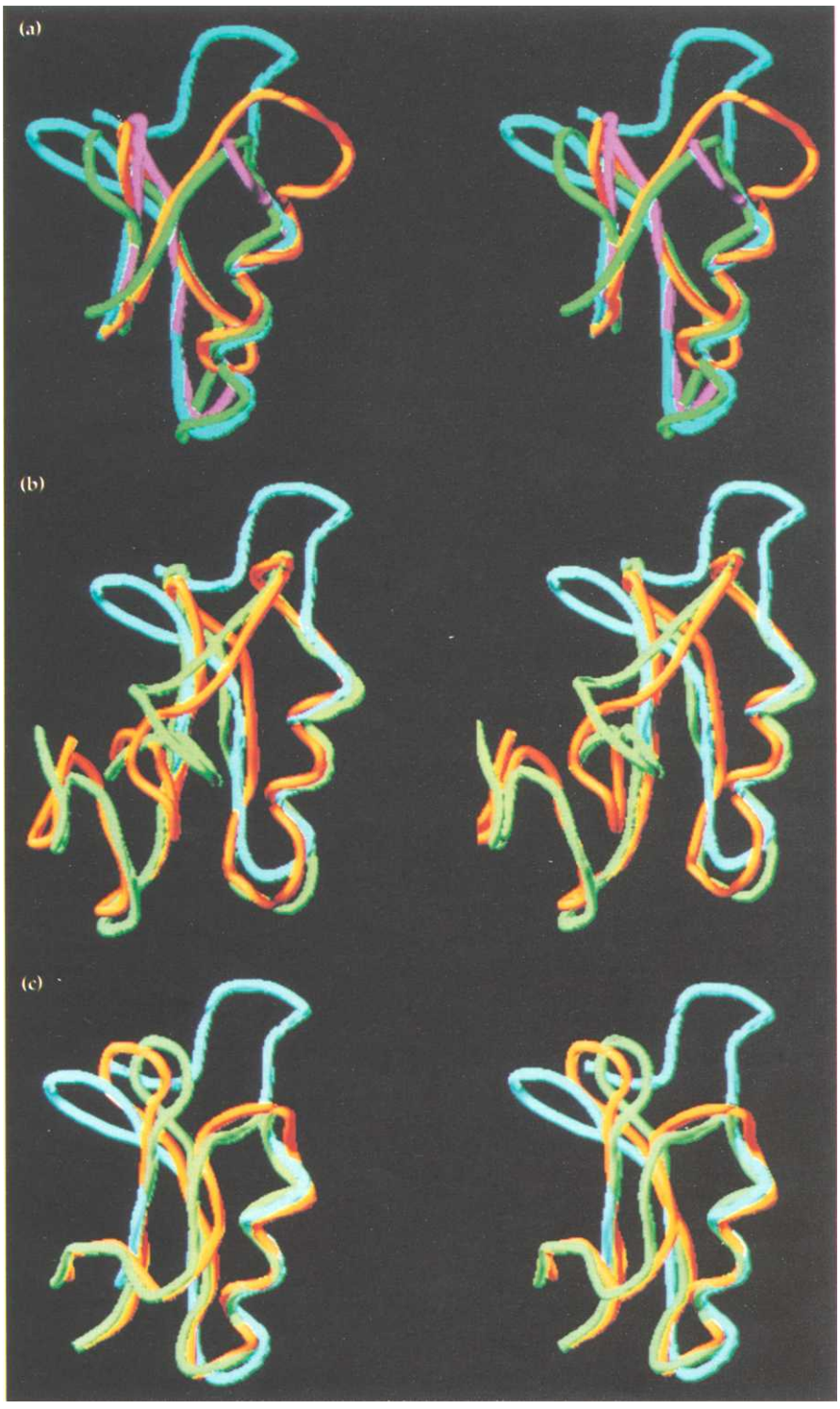

Fig. 6. Stereo superpositions of insect defensin A with several structurally homologous proteins exhibiting the CSαβ motif. Coordinates are fitted to the Cα atoms of the four cysteines involved in the CSαβ motif. These plots were obtained using SYBYL [56]. In all plots defensin A is in cyan. **(a)** Superposition with short scorpion toxins. Charybdotoxin is shown in green, PO_5-NH_2 toxin analogue in magenta and insectotoxin in orange. **(b)** Superposition with long scorpion toxins. Variant-1 is shown in green and variant-3 in orange. **(c)** Superposition with γ-thionins. γ1-H is shown in orange and γ1-P in green. (See Table 4 for the reference details.)

Experiments performed on *Micrococcus luteus* [16] and on giant liposomes indicate that the cytoplasmic membrane is a potential target for defensin A. Considering the distribution of amino acids forming polar and hydrophobic patches on the protein surface, it would be premature to propose a precise model of interaction of this protein with membranes. Nevertheless, as defensin A is positively charged, it could be argued that it can strongly perturb membrane organization by interacting with negatively charged phospholipids.

The CSαβ motif is shared by a number of small cysteine-rich proteins and appears as a general type of organization for several families of proteins. Striking similarities between structures of insect defensins, scorpion toxins and plant γ-thionins illustrate this point. On the other

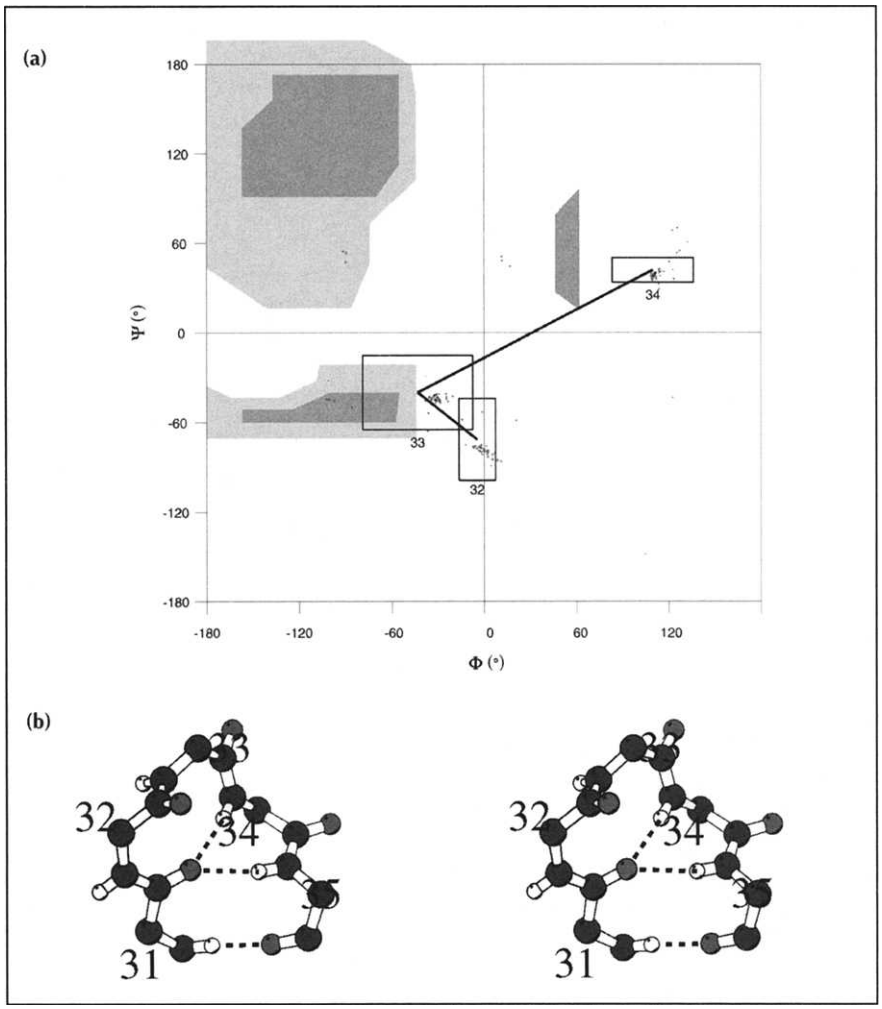

Fig. 7. Conformation of the unusual three-residue turn between the two strands of the β-sheet. **(a)** Ramachandran plot of the 43 energy-minimized structures. Squares represent the average±rms values of torsion angles. Lines connect the sequential residues. **(b)** Stereoview of the turn. Hydrogen bonds are represented by dashed lines. (Drawn with MOLSCRIPT [55].) The average conformation and $NH^{34}\rightarrow O^{31}$ hydrogen bonds correspond to a 3:5 turn. $NH^{35}\rightarrow O^{31}$ hydrogen bond corresponds to a 3:3 turn.

hand, no analogies are found with mammalian defensins [21,22].

Other defensins, from different insect species [1], can now be modelled by homology methods using defensin A as a template. This study constitutes a basis for future investigations of structure–function relationships in this family of defence proteins, complementing the NMR studies. It should also shed light on the evolution of the insect immune system and, more generally, on the phylogeny of small cysteine-rich proteins exhibiting antibiotic or toxic properties.

Materials and methods

NMR data

Two-dimensional ^1H NMR experiments were carried out on a recombinant *P. terranovae* defensin A produced by a strain of *Saccharomyces cerevisiae* kindly provided by Transgène (Strasbourg, France). Methods of spectra recording, data processing and 2D map assignments are described elsewhere [18]. NOE intensities constituted the primary experimental data used to derive the 3D structure. A complete analysis of NOESY maps recorded at a mixing time of 150 ms in either ^1H_2O or ^2H_2O provided an initial set of 705 dipolar connectivities from which 507 were sufficiently well defined (i.e. without overlap) for integration of cross-peak volumes using Bruker's UXNMR program. For backbone and side-chain β-protons, NOE intensities were converted into inter-proton distance ranges. The latter were calculated on the basis of a $1/r^6$ dependence of NOEs using the average intensity of 16 well-defined intra-residue β-methylene proton cross-peaks as a reference. Lower limits and upper limits were chosen by considering intervals of ±20% or ±50% for the integrated volumes, depending on whether cross-peak integration was made on both sides or on one side of the diagonal. Other side-chain connectivities classified as so-called strong, medium or weak were converted into distances ranging from a common lower limit of 1.8 Å (sum of the van der Waals radii) to upper limits of 2.5 Å, 3.5 Å or 5.5 Å, respectively. The upper limit was calculated by direct $1/r^6$ conversions of the smallest volumes measured within each class. Using a very long mixing time (300 ms) allowed us to detect very small NOEs, hence 199 additional constraints were added with an upper limit of 6.5 Å. This upper-limit value was chosen to circumvent errors from spin diffusion [44] and T_1 relaxation. The validity of all NOE-derived constraints was carefully tested in a first approach including back-calculations (see below). Distance constraints corresponding to the three disulphide bridges 3–30, 16–36 and 20–38 determined by mass spectrometry [15] were also introduced. These distance constraints range from 2.0–2.1 Å for S_i^γ–S_j^γ bonds and from 2.5–3.5 Å for C_i^β–S_j^γ bonds and S_i^γ–C_j^β bonds. Hydrogen-bond constraints involving backbone amide protons which

correspond to N→O distances of 2.6–3.3 Å were established according to NMR experiments: NH[27,29,31,35,37] had very low exchange rates and were involved in the β-sheet stabilization and NH[19,21,22,23,24,25] had small temperature coefficients and were engaged in the helix stabilization. Note that the secondary structures mentioned above were deduced unambiguously from a characteristic NOE connectivity pattern [18,19]. During the modelling procedures, NH[20] and NH[39] appear hydrogen bonded in most of the calculated structures. From this, re-examination of $^1H/^2H$ exchange and temperature coefficients data allowed us to find these additional hydrogen bond constraints (see above).

Computational procedure

Our modelling strategy included validation of experimental constraints, a wide exploration of conformational space, back-calculations of NMR data and energy minimization. Data validation and conformational space exploration were performed using the standard simulated annealing (SA) protocol of X-PLOR 2.9 [45–50]. Starting structures were generated using random (ϕ,ψ) torsion angles and then energy minimized (500 steps) to avoid steric contacts. These structures were subjected to 20 ps of high temperature (1000 K) molecular dynamics (MD) using the 'repel' potential energy function of X-PLOR (initial force constant C_{rep}=0.002, scaling of the van der Waals radii k_{rep}=1) and a soft square-well potential for the constraint energy term (slope of the asymptote c=0.1). An additional 10 ps of high-temperature MD allowed us to gradually increase the van der Waals repulsion term until C_{rep}=0.1 and the soft square-well potential until c=1. During a cooling phase, the temperature was decreased to 300 K by 25 K steps every 0.5 ps using a full square-well potential and smaller atoms (k_{rep}=0.8). Finally, 500 steps of energy minimization were performed. During all SA phases, the force constant of the constraint energy term was kept at 207 kJ mol^{-1} $Å^{-2}$.

We checked the consistency of NOE cross-peak assignments in order to validate each NOE-derived distance constraint and to proceed to stereospecific assignments of side-chain β- or γ-protons. The first round of calculations was performed as follows. Ten randomly generated structures were subjected to SA by considering only the nine distance constraints representing disulphide bridges and the eleven experimentally established hydrogen bonds. In the first round, all NOE constraints were inactivated. The resulting structures were used in successive rounds of SA during which stereospecific assignments were made using the GLOMSA program [51], while NOE constraints were gradually introduced. At the end of each round, an NOE-derived constraint was validated only when its violation was <1 Å for at least nine of the ten structures. When no diastereospecific assignment was established for a prochiral pair of protons, we used the largest upper bound for the two related constraints with the $1/r^6$ averaging option of X-PLOR. In this case, both protons were selected and the distance was averaged according to $(<r^{-6}>)^{-1/6}$. This is equivalent to applying the constraint to the nearest proton of the methylene pair. When stereospecific assignments were established, the distances were specified explicitly. For degenerate protons such as methyl or aromatic ring protons, we used the $1/r^6$ summation of X-PLOR. In this option, all combinations of proton pairs were summed: $(\Sigma r^{-6})^{-1/6}$.

An improvement of the constraint set was finally achieved through a first round of back-calculations. A simulation of a NOESY experiment with a mixing time of 150 ms was performed by using the standard procedure of X-PLOR 3.1. An isotropic correlation time of 2 ns, as deduced from Stokes's formula for a 4 kDa protein, was used to calculate NOE intensities. These intensities were calibrated with respect to the average intensity of ββ′-methylene cross-peaks. A careful comparison between experimental and calculated 2D maps led to the back-assignment of ~20 NOE cross-peaks. Additional rounds of SA calculations led to the elimination of 143 distance constraints corresponding to violations >0.5 Å. Finally, we were left with a total of 786 NOE-derived distance constraints, including 371 backbone–backbone as well as 415 side-chain–backbone and side-chain–side-chain constraints. Here, examination of the structures obtained during the last round of NOE-constraint validation and re-examination of NMR data enabled us to introduce two additional hydrogen bond distance constraints. From the 143 rejected NOE-derived distance constraints, 30 are incompatible with our models (the others correspond to the smallest upper bound of the two related β-methylene constraints when no diastereospecific assignment was possible).

Using this improved data set, 50 structures were randomly generated and subjected to SA. Minimization was performed using the CHARMM force-field [52] including standard electrostatic, Lennard–Jones and hydrogen-bond potential energy terms, in addition to covalent energy terms. The NMR constraints were applied as NOE intensities using the relaxation pseudo-energy function included in X-PLOR 3.0 [53]:

$$E_{NMR}=\sum_{i=1}^{N}\left[\sqrt[6]{I_i^c}-\sqrt[6]{k_s I_i^o}\right]^2$$

where I^c is the calculated NOE intensity, I^o is the intensity of the experimental NOE and k_s is a constant to allow the adjustment between experimental and calculated ββ′-methylene intensities. A cut-off of 10 Å was applied to non-bonded interactions and a distance-dependent dielectric function (ϵ=4r) was used for the calculation of the electrostatic terms.

Reliability between experimental and back-calculated NOE intensities was quantified through the calculation of a reliability factor (Rf) [26]:

$$Rf\,(\%)=\frac{\sum_{i=1}^{N}\left(\sqrt[6]{I_i^c}-\sqrt[6]{k_s I_i^o}\right)}{\sum_{i=1}^{N}\sqrt[6]{k_s I_i^o}}$$

Data analysis

The procedure described above produced 50 structures that we compared by calculating the rmsd of Cα coordinates for each pair of structures. The pairwise rmsd threshold defining a set of structures was chosen to be 1 Å. The similarities of each set were appreciated by calculating rmsds for N, Cα and C′ backbone atom coordinates, with respect to an average structure and by calculating both the average rmsd and the standard deviation of these rmsds. In order to define global superpositions of molecular regions, the average rmsd per residue was calculated from the N, Cα and C′ coordinates within the whole family of structures (global rmsd plot). Local analogies were analyzed by considering the local rmsd plot, also called the three-residue rmsd plot, defined by Berndt *et al.* [33]. This average rmsd for N, Cα and C′ atoms of a tripeptide fragment was calculated with respect to the average structure of the fragment. By sliding this tripeptide window along the sequence, variations in the structures were detected and analyzed.

Another method currently used to display the variability of the local conformation is based on the calculation of the circular

variance (CV) of the ϕ, ψ, $\chi 1$, $\chi 2$ or χss torsion angles [54,31] as:

$$CV_\theta = 1 - \frac{\sqrt{(\sum_{i=1}^{N}\cos\theta_i)^2 + (\sum_{i=1}^{N}\sin\theta_i)^2}}{N}$$

where θ represents a given torsion angle and N the number of θ values.

The distribution of ϕ,ψ angle pairs in the Ramachandran map was analyzed using the PROCHECK program [29]. The ratio of residues that fall within the most energetically favoured regions can be considered as a good criterion for assessing the quality of the structure. The conformation nomenclature (α_R, α_L, β, γ_R, γ_L, δ and ϵ) used here is as defined by Efimov [41] and quoted by Sibanda *et al.* [42].

Coordinates of insect defensin A have been deposited with the Brookhaven Protein Data Bank (entry code 1 ICA).

Acknowledgements: We thank Prof M Rico and Dr M Bruix (Madrid, Spain) for kindly providing us with the coordinates of γ-thionins. This work was supported by the Centre National de la Recherche Scientifique (CNRS), the Région Centre, and the Université d'Orléans. It forms part of the thesis of Bruno Cornet, supported by a grant from Ministère de la Recherche et de l'Enseignement Supérieur. We thank the Transgène company (Strasbourg, France) for providing us with samples of recombinant defensin A.

References

1. Hoffmann, J.A. & Hetru, C. (1992). Insect defensins: inducible antibacterial peptides. *Immunol. Today* **13**, 411–415.
2. Boman, H.G., Faye, I., Gudmundsson, G.H., Lee, J.Y. & Lidholm, D.A. (1991). Cell-free immunity in *Cecropia*. A model system for antibacterial proteins. *Eur. J. Biochem.* **201**, 23–31.
3. Holak, T.A., *et al.*, & Clore, G.M. (1988). The solution structure of the antibacterial peptide cecropin A: a nuclear magnetic resonance and dynamical simulated annealing study. *Biochemistry* **27**, 7620–7629.
4. Durell, S.R., Raghunathan, G. & Guy, H.R. (1992). Modelling the ion channel structure of cecropin. *Biophys. J.* **63**, 1623–1631.
5. Casteels, P., Ampe, C., Jacobs, F., Vaeck, M. & Tempst, P. (1989). Apidaecins: antibacterial peptides from honeybees. *EMBO J.* **8**, 2387–2391.
6. Casteels, P., *et al.*, & Tempst, P. (1990). Isolation and characterisation of abeacin, a major antibacterial response peptide in the honeybee (*Apis mellifera*). *Eur. J. Biochem.* **187**, 381–386.
7. Bulet, P., *et al.*, & Hoffmann, J.A. (1993). A novel inducible anti-bacterial peptide of *Drosophila* carries an O-glycosylated substitution. *J. Biol. Chem.* **268**, 14893–14897.
8. Kockum, K., Faye, I., von Hofsten, P., Lee, J.Y., Xanthopoulos, K.G. & Boman, H.G. (1984). Insect immunity. Isolation and sequence of two cDNA clones corresponding to acidic and basic attacins. *EMBO J.* **3**, 2071–2075.
9. Dimarcq, J.L., *et al.*, & Hoffmann, J.A. (1988). Insect immunity. Purification and characterisation of a family of novel antibacterial proteins from immunized larvae of the dipteran *Phormia terranovae* and complete amino-acid sequence of the predominant member, diptericin A. *Eur. J. Biochem.* **171**, 17–22.
10. Bulet, P., *et al.*, & Hoffmann, J.A. (1991). Insect immunity. Isolation from a coleopteran insect of a novel inducible antibacterial peptide and of new members of the insect defensin family. *J. Biol. Chem.* **266**, 24520–24525.
11. Fujiwara, S., Imai, J., Fujiwara, M., Yaeshima, T., Kawashima, T. & Kobayashi, K. (1990). A potent antibacterial protein in royal jelly. *J. Biol. Chem.* **265**, 11333–11337.
12. Lambert, J., *et al.*, & Hoffmann, D. (1989). Insect immunity: isolation from immune blood of the dipteran *Phormia terranovae* of two insect antibacterial peptides with sequence homology to rabbit lung macrophage bactericidal peptides. *Proc. Natl. Acad. Sci. USA* **86**, 262–266.
13. Matsuyama, K. & Natori, S. (1988). Purification of three antibacterial proteins from culture medium of NIH-Sape-4, an embryonic cell line of *Sarcophaga peregrina*. *J. Biol. Chem.* **263**, 17112–17116.
14. Kuzuhara, T., Nakajima, Y., Matsuya, K. & Natori, S. (1990). Determination of the disulfide array in sapecin, an antibacterial peptide of *Sarcophaga peregrina* (flesh fly). *J. Biochem. (Tokyo)* **107**, 514–518.
15. Lepage, P., *et al.*, & Van Dorsselaer, A. (1991). Determination of disulfide bridges in natural and recombinant insect defensin A. *Eur. J. Biochem.* **196**, 735–742.
16. Cociancich, S., Ghazi, A., Hetru, C., Hoffmann, J.A. & Letellier, L. (1993). Insect defensin, an antibacterial peptide, forms voltage-dependant channels in *Micrococcus luteus*. *J. Biol. Chem.* **268**, 19239–19245.
17. Matsuyama, K. & Natori, S. (1990). Mode of action of sapecin, a novel antibacterial protein of *Sarcophaga peregrina*. *J. Biochem. (Tokyo)* **108**, 128–132.
18. Bonmatin, J.M., *et al.*, & Achestetter, T. (1992). Two-dimensional ^1H NMR study of recombinant insect defensin A in water: resonance assignments, secondary structure and global folding. *J. Biomol. NMR* **2**, 235–256.
19. Bonmatin, J.M., *et al.*, & Ptak, M. (1992). Progress in multidimensional NMR investigations of peptide and protein 3-D structures in solution. From structure to functional aspects. *Biochimie* **74**, 825–836.
20. Hanzawa, H., *et al.*, & Arata, Y. (1990). ^1H nuclear magnetic resonance study of the solution conformation of an antibacterial protein, sapecin. *FEBS Lett.* **269**, 413–420.
21. Pardi, A., Zhang, X.L., Selsted, M.E., Skalicky, J.J. & Yip, P.F. (1992). NMR studies of defensin antimicrobial peptides. 2. Three-dimensional structures of rabbit NP-2 and human HNP-1. *Biochemistry* **31**, 11357–11364.
22. Ganz, T. & Lehrer, R.I. (1994). Defensins. *Curr. Opin. Immunol.* **6**, 584–589.
23. Tamaoki., H., *et al.*, & Sakakibara, S. (1991). Solution conformation of endothelin determined by means of ^1H-NMR spectroscopy and distance geometry calculations. *Protein Eng.* **4**, 509–518.
24. Bernstein, F.C., *et al.*, & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
25. Abola, E.E., *et al.*, Weng, J.C. (1987). Protein Data Bank. In *Crystallographic Databases — Information Contents, Software Systems, Scientific Applications.* (Allen, F.H., Bergenhoff, G., & Sievens, R., eds), pp. 107–132, Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester.
26. James, T.L., Gochin, M., Kerwood, D.J., Pearlman, D.A., Schmitz, U. & Thomas, P.D. (1991). Refinement of three-dimensional protein and DNA structures in solution from NMR data. In *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy.* (Hoch, J.C., ed), pp. 331–348, Plenum Press, New York.
27. Chothia, C. (1983). Coiling of β-pleated sheets. *J. Mol. Biol.* **163**, 107–117.
28. Efimov, A.V. (1993). Standard structures in proteins. *Prog. Biophys. Mol. Biol.* **60**, 201–239.
29. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
30. Lee, B. & Richard, F.M. (1971). The interpretation of protein structures. Estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
31. MacArthur, M.W. & Thornton, J.M. (1993). Conformational analysis of protein structures derived from NMR data. *Proteins* **17**, 232–251.
32. Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 167–330.
33. Berndt, K.D., Güntert, P. & Wüthrich, K. (1993). Nuclear magnetic resonance solution structure of dendrotoxin K from the venom of *Dendroaspis polylepis polylepis*. *J. Mol. Biol.* **234**, 735–750.
34. Shimoda, M., Tagaki, H., Kurata, S., Yoshioka, T. & Natori, S. (1994). Inhibition of the $Ca^{(2+)}$-activated $K^{(+)}$-channel by sapecin B, an insect antibacterial protein. *FEBS Lett.* **339**, 59–62.
35. Kim, J.I., *et al.*, & Sato, K. (1994). Synthesis and characterization of sapecin and sapecin B. *FEBS Lett.* **342**, 189–192.
36. Bulet, P., *et al.*, & Hoffmann, J.A. (1992). A novel insect defensin mediates the inducible antibacterial activity in larvae of the dragon fly *Aeschna cyanea* (Paleoptera, Odonata). *Eur. J. Biochem.* **209**, 977–984.
37. Nilges, M. (1993). A calculation strategy for the structure determination of symmetric dimers by ^1H NMR. *Proteins* **17**, 297–309.
38. Pease, J.H.B. & Wemmer, D.E. (1988). Solution structure of apamin determined by nuclear magnetic resonance and distance geometry. *Biochemistry* **27**, 8491–8498.
39. Orengo, C.A. & Thornton, J.M. (1993). Alpha plus beta folds revisited: some favoured motifs. *Structure* **1**, 105–120.

40. Carrasco, L., Vazquez, D., Hernandez-Lucas, C., Carbonero, P. & Garcia-Olmedo, F. (1981). Thionins: plant peptides that modify membrane permeability in cultured mammalian cells. *Eur. J. Biochem.* **116**, 185–189.
41. Efimov, A.V. (1986). Standard conformations of a polypeptide chain in irregular protein regions. *Mol. Biol. (Moscow)* **20**, 250–260.
42. Sibanda, B.L., Blundell, T.L. & Thornton, J.M. (1989). Conformation of β-hairpins in protein structures. A systematic classification with application to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol.* **206**, 759–777
43. Sibanda, B.L. & Thornton, J.M. (1993). Accommodating sequence changes in β-hairpins in proteins. *J. Mol. Biol.* **229**, 428–447.
44. Clore, G.M. & Gronenborn, A.M. (1985). Assessment of errors involved in the determination of interproton distance ratio and distances by means of one- and two-dimensional NOE measurements. *J. Magn. Reson.* **61**, 158–164.
45. Clore, G.M., Gronenborn, A.M., Brünger, A.T. & Karplus, M. (1985). Solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of *Escherichia coli*. Combined use of ^{1}H nuclear magnetic resonance and restrained molecular dynamics. *J. Mol. Biol.* **186**, 435–455.
46. Brünger, A.T., Clore, G.M., Gronenborn, A.M. & Karplus, M. (1986). Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: application to crambin. *Proc. Natl. Acad. Sci. USA* **83**, 3801–3805.
47. Clore, M.G., Brünger, A.T, Karplus, M. & Gronenborn, A.M. (1986). Application of molecular dynamics with interproton distance restraints to three-dimensional protein structure determination. A model study of crambin. *J. Mol. Biol.* **191**, 523–551.
48. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. *Science* **235**, 458–460.
49. Nilges M., Clore G.M. & Gronenborn, A.M. (1988). Determination of the 3-D structure of proteins from interproton distance data by hybrid distance geometry–dynamical simulated annealing calculations. *FEBS Lett.* **229**, 317–324.
50. Nilges M., Clore G.M. & Gronenborn, A.M. (1988). Determination of the three-dimensional structure of proteins from interproton distance data by simulated annealing from a random array of atoms. *FEBS Lett.* **239**, 129–136.
51. Güntert, P., Braun, W. & Wüthrich, K. (1991). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217**, 517–530.
52. Brooks, B., Brucoli, R., Olafson, B.O., States, D., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
53. Nilges, M., Habazettl, J., Brünger, A.T. & Holak, T.A. (1991). Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. *J. Mol. Biol.* **219**, 499–510.
54. Mardia, K.V. (1972). *Statistics of Directional Data*. Academic Press, London.
55. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
56. Clark, M., Cramer, R.D. & van Opdenbosch, N. (1989). Validation of the general Tripos 5.2 force field. *J. Comput. Chem.* **10**, 982–1012.
57. Bontems, F., Roumestand, C., Gilquin, B., Menez, A. & Toma, F. (1991). Refined structure of charybdotoxin: common motifs in scorpion toxins and insect defensins. *Science* **25**, 1521–1523.
58. Meunier, S., *et al.*, & Darbon, H. (1993). Solution structure of PO_5-NH_2, a scorpion toxin analog with high affinity for the apamin-sensitive potassium channel. *Biochemistry* **32**, 11969–11976.
59. Arseniev, A.S., Kondarov, V.I., Maiorov, V.N. & Bystrov, V.F. (1984). NMR solution spatial structure of 'short' insectotoxin I5A. *FEBS Lett.* **165**, 57–62.
60. Zhao, B., Carson, M., Ealick, S.E. & Bugg, C.E. (1992). Structure of scorpion toxin variant-3 at 1.2 Å resolution. *J. Mol. Biol.* **227**, 239–252.
61. Lee, W., Jablonsky, M.J., Watt, D.D. & Krishna, N.R. (1994). Proton nuclear magnetic resonance and distance geometry/simulated annealing studies on the variant-1 neurotoxin from the New World scorpion *Centruroides sculpturatus Ewing*. *Biochemistry* **33**, 2468–2475.
62. Bruix, M., *et al.*, & Rico, M. (1993). Solution structure or γ1-H and γ1-P thionins from barley and wheat endosperm determined by ^{1}H-NMR: a structural motif common to toxic arthropod proteins. *Biochemistry* **32**, 715–724.

Received: **4 Dec 1994**; revisions requested: **11 Jan 1995**; revisions received: **14 Mar 1995**. Accepted: **16 Mar 1995**.