Int. J. Peptide Protein Res. **48**, 1996, 220–228
Printed in UK – all rights reserved

INTERNATIONAL JOURNAL OF
PEPTIDE & PROTEIN RESEARCH
ISSN 0367–8377

Three-dimensional structure of the RGD-containing snake toxin albolabrin in solution, based on ^{1}H NMR spectroscopy and simulated annealing calculations

K. JOHN SMITH[1], MAHESH JASEJA[1], XINJIE LU[2], JANICE A. WILLIAMS[3], EVA I. HYDE[1] and IAN P. TRAYER[1]

[1]*School of Biochemistry, University of Birmingham, Birmingham, UK,* [2]*Thrombosis Research Institute, Emmanuel Kaye Building, London, UK, and* [3]*Department of Applied Pharmacology, National Heart and Lung Institute, London, UK*

Received 25 October, revised 13 December 1995, accepted for publication 17 March 1996

Albolabrin is a snake toxin that contains a RGD-(Arg-Gly-Asp) sequence motif and competes with fibrinogen to bind to the integrin αIIbβ3 (GpIIb-IIIa) on platelets. It thus inhibits platelet aggregation and cell–cell adhesion. It shows a high sequence similarity to other disintegrins, yet the reported disulfide bonding pattern for this peptide differs from that of others in this family. Recently we reported the assignment of the ^{1}H-NMR spectrum of albolabrin and a preliminary description of its secondary structure [Jaseja, M., Smith, K.J., Lu, X., Williams, J.A., Trayer, H., Trayer, I.P. & Hyde, E.I. (1993) *Eur. J. Biochem.* **218**, 853–860]. Here we present a more detailed description of the secondary and the tertiary structure, based on the ^{1}H NMR results and simulated annealing methods.

The structure of albolabrin in solution was calculated using 318 distance and 18 dihedral angle restraints. The average atomic RMS deviation between 12 refined structures and the mean structure was 3.1 Å for the backbone. The protein appears to be highly mobile. Its structure is dominated by a series of turns and by three hairpins, each with a short region of distorted antiparallel β-pleated sheet, held together by six disulfide bridges. The most well defined area is the hydrophobic core, residues 21–47 and 57–67, which is clustered around F40 and has a backbone atomic RMS deviation of only 1.3 Å from the mean structure. The RGD adhesion sequence is found at the highly mobile tip of one of the β-hairpins, protruding from the body of the protein.

Many of these structural features are similar to those of other disintegrins, and differences in the disulfide bonding pattern of the disintegrins can be accomodated without significant energy penalty. Comparison of this structure with other proteins of similar function suggests that it is the RGD-loop, rather than the precise topology of the proteins, that is important to antagonist activity. © Munksgaard 1996.

Key words: disintegrin; snake venom; disulfide

Disintegrins form a family of homologous cysteine-rich polypeptides containing an Arg-Gly-Asp (RGD) motif, found in snake venoms (reviewed in ref. 1). The RGD sequence is the consensus sequence by which adhesive proteins such as fibrinogen, finronectin, vitronectin and von Willebrand factor recognise the integrin, αIIbβ3 (also known as Gp IIb–IIIa), which is the major cell–surface receptor for fibrinogen on activated platelets (2). The distintegrins compete for RGD-binding sites on the receptor and act as potent inhibitors of the binding of the adhesive proteins to activated platelets, and hence of platelet aggregation and cell–cell adhesion (3, 4).

The disintegrins have been divided into three classes, based on their length and number of cysteine residues. The long disintegrins, such as bitistatin, contain about 84 residues and 14 cysteines. Albolabrin, kistrin and flavoridin belong to the medium length class of disintegrins containing about 73 residues and 12 cysteines, whereas echistatin is a short disintegrin, containing 49 residues and only 8 cysteines. All the cysteine residues in these disintegrins are involved in disulfide bonds; however, the reported disulfide pattern of albolabrin (5) differs from that of kistrin (6) and flavoridin (7, 8), and echistatin has a third disulfide bonding pattern (7, 9). The structures of kistrin (6, 10, 11), flavoridin (12) and echistatin (9, 13–17) have been determined by NMR spectroscopy. We recently reported the assignment of the ^{1}H NMR spectrum of albolabrin (18), with a preliminary description of its secondary structure. In this paper we give a more detailed description of the

secondary and tertiary structure of albolabrin, calculated by simulated annealing methods.

EXPERIMENTAL PROCEDURES

Experimental constraints. Albolabrin was prepared and NMR data were acquired and assigned as described previously (18). Analysis of the NOESY spectrum collected at pH 5.0 and 310 K with a mixing time of 100 ms yielded a complete data set of NOEs of which 318 yielded constraints lower than the sum of the covalent bond lengths. Of these useful NOEs, 119 were non-sequential, 21 were medium range $[3<(i-j)<5]$ and 83 were long range $[6<(i-j)]$. Among the long-range NOEs, 58 were separated by more than 10 residues, and 21 by more than 20 residues. This gave an average of five useful NOE constraints per residue over the defined area of the protein. Changes in chemical shift at other conditions allowed us to confirm some of the ambiguous assignments but yielded no additional NOEs, nor were any additional NOE crosspeaks identified at 200 ms mixing time. The distribution of these NOE constraints is illustrated in Fig. 1.

The volumes of the NOESY crosspeaks were estimated by counting contours in the 2D plot, and classified into five groups corresponding to NOEs with intensities between strong and very weak. The upper and lower bounds of the corresponding constraints were 1.8–2.5 Å, 2.2–3.3 Å, 3.0–4.0 Å, 3.5–5.0 Å and 4.0–6.0 Å. 6.0 Å was used as a conservative value for the upper bound of the very weakest contacts. No stereospecific assignments were made. Pseudoatom corrections were added to the upper bounds (19) and an additional 0.5 Å was added to upper distance constraints involving methyl groups (20). Eighteen dihedral angle restraints were also identified (Fig. 1). Thirteen $^3J_{NH-\alpha H}$ coupling constants larger that 8 Hz were observed, and the corresponding ϕ dihedral angles restrained to $-125\pm50°$ (≥ 8 Hz) or $-125\pm40°$ (≥ 9 Hz). In addition five ψ angle restraints (of $140\pm60°$) were added in regions of the polypeptide where the NOE, hydrogen bonding and ϕ dihedral pattern indicated an extended β-sheet secondary structure. In some calculations, one additional dihedral angle constraint was applied to χ^3 across the disulfide bridge between residues Cys34–Cys38 ($90\pm20°$). Without this restraint the disulfide bond was distorted, with the dihedral most frequently lying around 50–60°, which lies at the limits of the predicted maximum allowed static range (21).

Thirteen slowly exchanging amide protons were observed in a hydrogen-exchange experiment, in which the polypeptide was freeze-dried from H_2O (pH 6.5), dissolved in D_2O and a COSY spectrum collected within 30 min. The temperature coefficients of the chemical shifts of these slowly exchanging backbone amide protons (measured in TOCSY spectra collected between 295–310 K) were below 3 ppb/K. From these, 11 hydrogen bonds were identified by examination of the secondary structure observed in preliminary calculations in the absence of hydrogen bond constraints. The corresponding carbonyl acceptors were identified by comparison of the observed NOEs with those found in regular secondary structures (i.e. β-turn or β-sheet). These structures were modelled and gave the lowest violations of the NOEs in subsequent calculations. The hydrogen bond distances were constrained to O–N=2.0–3.3 Å and O–H=1.8–2.3 Å. Other backbone amide resonances which exchange slowly with D_2O (Fig. 1) can be accounted for by steric protection from the solvent.

Calvete *et al.* (5) have reported a disulfide pattern for albolabrin. Direct chemical evidence was presented for the disulfide bridges between cysteines 29–59 and 47–66, and indirect evidence indicated that the other linkages were between cysteines 6–15, 8–16, 21–35 and 34–38. This disulfide bridge pattern was introduced into our structural calculations as either pseudo-NOE constraints (distances constrained were S_γ–$S_\gamma=2.02\pm0.02$ Å and S_γ–$C_\beta=3.0\pm0.5$ Å) or as covalent constraints. All the observed NOEs are consistent with the disulfide bonding pattern determined by Calvete *et al.* (5), but other disulfide patterns cannot be excluded from the NMR data, and calculations were also performed with other disulfide bonding patterns (see below).

Calculations

Structure determination was carried in X-PLOR 3.1 (22) using a hybrid protocol in which a subset of atoms from each residue (C_α, H_α, N, HN, C, $C_{\beta*}$, $C_{\gamma*}$) was embedded using the distance geometry routine [similar to the dg_sub_embed protocol; (22)]; the remainder of the atoms were placed by template fitting and then the polypeptide coordinates were allowed to evolve under the applied NOE distance constraints during a series of simulated annealing steps [similar to the dgsa and refine protocols (22)]. A total of 100 structures was calculated for each set of constraints (described below), and, for each, a subset of structures containing the lowest violations of the experimental NOE and dihedral angle constraints was chosen. These were refined by minimising the averaged coordinates produced over the last 4 ps of a 5 ps low-temperature simulated annealing trajectory. The parallhdg.pro parameter set in XPLOR was used for the minimisation, with the force constants set to the following values: NOE 50 kcal mol^{-1} $Å^{-2}$ (square-well potential), van der Waals repulsion term 4 kcal mol^{-1} $Å^{-4}$ (repel function, with the van der Waals scaling factor set to 0.75), dihedral constraints 200 kcal mol^{-1} rad^{-2}. Electrostatics were excluded from all calculations, and disulfide bridges were intro-

duced as pseudo-NOEs in the 'dgsa' protocol, and thereafter as covalent constraints.

In addition to the explicit upper and lower bounds estimated from NOESY spectra, as described above, two sets of mild 'repulsive' constraints were applied in the structure calculations. The first set of repulsive constraints consisted of lower bounds for certain sequential inter-residue distances where the corresponding *sequential* NOEs were definitely *not* observed in any of the spectra collected (23). Hence, lower bounds for unobserved $d_{NN}(i, i+1)$ contacts were set to 3.3 Å, and in the reverse turn between residues Asp17 and Lys22, the lower bounds for $d_{\alpha N}(i, i+1)$ contacts for residues Ala19–Thr20 and Cys21–Lys22 were set to 3.3 Å, and the lower bound for the $d_{\beta N}(i, i+1)$ contact for residues Thr20–Cys21 was set to 2.8 Å. The introduction of these additional constraints excluded many very short contacts which were seen in the structures calculated with a constraints set corresponding only to the observed NOEs, and hence improved the agreement of the calculated structure with the NOESY data.

The second set of repulsive constraints were introduced after a detailed examination of structures calculated with the dataset described above revealed many short contacts, all of which appeared inconsistent with the measured NMR data. In total, over 1200 *non-sequential* inter-residue distances involving NH, H_α and H_β protons were identified as occasionally lying within 5 Å in calculated structures. Of these, 894 were found not to give NOEs; the remainder could not be unambiguously assigned because of overlap, and so were not included in the calculations. An explicit lower bound of 3.5 Å was assigned to these non-sequential distances (23). This is a conservative estimate for the cutoff distance in our NOESY spectra, to allow for the possibility that crosspeaks were absent due to mobility in the protein.

The structural calculations were repeated several times, each time with an additional subset of constraints. At first, lower bounds were set to a uniform value of 1.8 Å and only upper bounds were used, then lower bounds were introduced according to the observed NOEs, then repulsive constraints were introduced for sequential non-contacts, and finally repulsive constraints were introduced for longer-range non-contacts. There were no significant differences in the topology of the structures calculated with each constraint set, and, as observed by Holak *et al.* (23), the introduction of repulsive constraints did not bias the calculations, but instead improved the agreement of the calculated structures with the observed NOESY data.

The NOEs observed for albolabrin defined the secondary structural elements of the protein quite well (Fig. 3), but the relatively small number of long-range NOEs meant that the three-dimensional topology of the protein was less well defined. In particular, some difficulty was encountered in determining the handedness of the structure, hence both left- and right-handed structures were refined. The problem of determining global topology in a protein of this type, which is relatively flat and where most information is restricted to short-range contacts, is well known (see for example ref. 24). Three different folds were identified arising from the refinement of the left- and right-handed structures produced by X-PLOR/DG. The right-handed enantiomer refined to give the fold presented in this paper (Fig. 3), while the left-handed enantiomer gave a mixed set of final structures, in which either the *C*-terminus lay to the left of the body of the protein (effectively a mirror image of the right-handed structure), or in which the *C*-terminus between residues 60 and 64 crossed under the axis of the protein at 42–43, from the right to lie on the left-hand side. At each stage of the calculation the left-handed set of structures showed a higher-energy penalty on average than the right-handed set, indicating that the right-handed enantiomer is folded correctly.

In order to confirm that the fold observed was not biased by the method of calculation (e.g. by the way in which X-PLOR samples conformational space), the calculations were repeated with an independent method using DIANA 2.1.1 [Tripos Associates, Inc., 1993 (25, 26)]. Constraints were preprocessed in HABAS for analysis of local conformation (27), and then a standard REDAC protocol was applied in DIANA (DIANA manual, Tripos Associates, Inc.). Using the REDAC procedure with repulsive constraints, additional constraints were generated only in the regions 1–9, 51–53, in which there are few constraints. Using the REDAC procedure without repulsive constraints yielded similar constraints to that with repulsive constraints. In both cases, of the structures which passed an arbitrary energy penalty cutoff, most were of the right-handed fold previously identified in X-PLOR [e.g. using the full repulsive set of constraints, of the 39 structures from 100 started, which passed a cutoff, 36 (90%) were of the correct fold]. The structures calculated from DIANA showed less variation than those from X-PLOR (i.e. lower r.m.s.d.) but overall the lowest energy structures generated were very similar.

RESULTS

Secondary structure

The secondary structure of albolabrin consists largely of a series of turns, with little regular secondary structure (see the diagonal NOE map in Fig. 1), as in other disintegrins. Only three short sections of antiparallel β-sheet are present, giving three hairpin loops at Cys15–Pro25, Cys34–Met41 and Thr45–Cys59. Turns are found at Ser10–Asn13, Leu24–Ala27, Gly26–Cys29, Asn60–Ala64 and Asn69–His72. Of the five prolines in the protein, only

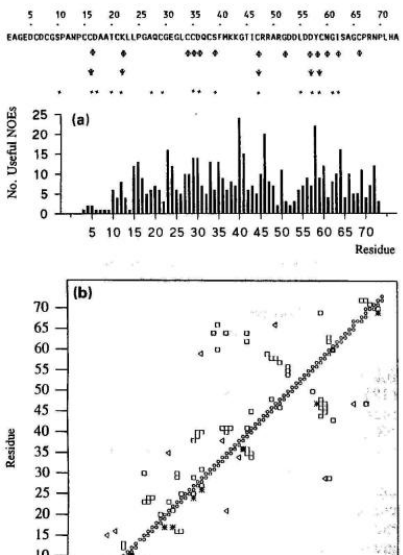

FIGURE 1

Distribution of NOE constraints observed for albolabrin. Top: The sequence of albolabrin with the dihedral angle constraints. The ϕ-symbol corresponds to a $^3J_{NH-\alpha H}$ coupling constant value ≥ 8 Hz, and the ψ-symbol to a ψ-dihedral restricted to the β-region of the Ramachandran plot. The amide protons protected from exchange with D_2O are marked with an '$\times$'. (a) Distribution of the 318 'useful' NOEs which provided an inter-proton distance restraint more restrictive than the maximum covalent distance (each NOE is represented twice in the histogram). (b) A diagonal map of NOEs observed for albolabrin. Sequential NOEs are represented by circles and non-sequential NOEs by squares (at least one NOE for each symbol). NOEs involving the sidechain resonances are plotted above the diagonal, and those involving only mainchain resonances are plotted below the diagonal. The triangles represent disulfide bridges and the stars represent hydrogen bonds.

Pro14 and Pro67 are not at the second position of a β-turn. The assignment of the type of turn observed is based on the NOEs and the Ramachandran plots for the two central residues of the turn. These structures are discussed briefly below.

The first element of regular structure is at residues Ser10–Asn13, which form a type I β-turn. A small hairpin is seen over the region Cys15–Leu24, with residues Cys16–Asp17 and Leu23–Lys22 forming a short segment of antiparallel β-strand, stabilised by hydrogen bonds between the backbone of Lys22 and Asp17 and also by the charge–charge interaction between their sidechains. An irregular loop stretching between residues Asp17–Cys21 is seen at the end of the hairpin, forming a type III β-turn at positions Asp17–Thr20. Two overlapping type II β-turns are seen at positions Leu24–Ala27 and Gly26–Cys29. This turn motif is continued until residue Glu31. Residues Cys34–Met41 form a highly constrained hairpin loop with a type I′ β-turn between residues Asp36–Ser39 at its tip. This loop contains a disulfide bridge within the hairpin (Cys34–Cys38), an external disulfide bridge (Cys21–Cys35), and in addition appears to be hydrogen bonded across the neck of the hairpin (Asp36–Ser39).

The pairing of sequences Ile46–Cys47 with Asp57–Tyr58 in a short region of antiparallel β-sheet creates a turn in the backbone in the region of the RGD sequence (residues 51–53). The sheet is distorted by a β-bulge, resulting from a bifurcated hydrogen bond between Asp57NH and both Cys47CO and Arg48CO. The structure of the hairpin may be stabilised by interactions between the charged sidechains, which give rise to a positively charged strand (Arg48, Arg49, Arg51) and a negatively charged strand (Asp53, Asp54, Asp56, Asp57). The upper limit of the β-sheet is formed by residues Thr45 and Cys59 and the lower limit is defined by the NOE between Ala50NH and Asp56αH. Hence, in albolabrin, the RGD sequence is located in a seven residue loop (Ala50–Asp56). Because of the importance of this RGD loop to disintegrin function, we detail the NOEs more fully. The methyl of Ala50 within this loop acts as a convenient reporter group, with weak or medium NOEs observed to each of Asp56 NH and αH, Leu55 NH and αH, and Asp54 αH and βH. The distance constraints arising from these NOEs are accommodated by movement of residues Leu55 and Asp56 in towards the centre of the loop. Weak or medium d_{NN} NOEs are observed between residues Ala50–Arg51, Gly52–Asp53 and Asp53–Asp54 (medium), and an 8.0 Hz $^3J_{NH-\alpha}$ coupling and slowly exchanging amide proton for residue Leu55. However, a search of calculated structures failed to identify any potential hydrogen-bond acceptor for Leu55, and hence the protection is probably due to steric hindrance of solvent access. All three residues RGD are in a β-conformation but no $d_{\alpha N}(i, i+2)$ NOEs are observed, so a tight β-turn is not present. As in the other disintegrins, the conformation about the RGD sequence can only be described as an open and mobile bent conformation, probably adopting transient turn-like structures, in which the residues RGD are almost fully exposed to the solvent.

A loosely defined, open or 'wide' turn exists for

residues Asn60–Ser64, which is defined largely by the tertiary fold of the protein, rather than by the local secondary structure. Finally, an approximately type I β-turn was found for residues Asn69–His72.

Three-dimensional structure calculations

We observe only a small number of long-range NOE contacts in albolabrin, and many of these are weak (Fig. 1b). We can exclude the possibility that this is due to broad linewidths of resonances, since strong intraresidue and sequential NOEs were seen over the whole protein. It is therefore likely that internal mobility in albolabrin is responsible for the low intensity, or absence, of these long-range NOEs. As found for other disintegrins, when the structure of albolabrin was calculated in the absence of disulfide constraints it failed to converge. In albolabrin, NOEs to support the disulfide bridge arrangement were almost entirely absent (Fig. 1b), and a search of inter-cysteine distances for disulfide bonds in these structures of albolabrin was inconclusive. The tertiary structure was therefore calculated using different disulfide bonding patterns.

Disulfide bonding pattern

Initially the disulfide bonding pattern determined chemically by Calvete *et al.* (5) was used, i.e. 6–15, 8–16, 21–35, 34–38, 29–59 and 47–66. Incorporation of the different pairs of disulfide bridges possible from Calvete's paper in structural calculations indicated that either pair of disulfides 6–15 and 8–16 or 6–16 and 8–15 could readily be accommodated. However, the replacement of Cys34–Cys38 by Cys35–Cys38 introduced significant violations of both long- and short-range constraints involving residues in the hairpin Cys34–Met41, so this alternative arrangement is unlikely to be correct.

Calculations were also performed with the disulfide bridge pattern corresponding to that found in flavoridin and kistrin (6–8, 10) i.e. disulfides 6–21, 8–16, 15–38, 29–35, 34–59, 47–66. As in other disintegrins, (6, 8, 9, 17), several possible patterns of disulfide bridges have approximately equivalent calculated energies, with the correct disulfide pattern having only slightly lower energy than other patterns. However, the disulfide pattern found in flavoridin and kistrin produced significantly larger violations of the observed NOEs than for structures calculated with the albolabrin pattern of disulfide bridges, with the most often violated constraints being key contacts in the protein core. Hence, we are confident that the structure discussed below is the one most compatible with our data.

Tertiary structure

Table 1 shows the deviations from experimental constraints and idealised geometry for the twenty lowest energy structures calculated with X-PLOR using the full data set of constraints. In our previous paper (18) we demonstrated that albolabrin exists as a monomer, and in agreement with this, all observed NOEs are compatible with a single calculated structure. The 12 best structures as judged by the lowest value of the NOE and dihedral angle violations were analysed. In the final structures no more than one constraint is violated by more than 0.4 Å (none by more than 0.5 Å), and the ϕ angles show no violations greater than 10°. Analysis of the Ramachandran plot showed that the ϕ,ψ angle values of the residues in the calculated structures lie in the allowed regions (data not shown). Five non-glycine residues are found in the right-hand side of the plot, namely Gln28, Cys29, Gln37, Cys38, Met41, the first four of which are involved in turns. In agreement with the intense $C_\alpha H_i$–$C_\beta H_{i+1}$ NOE crosspeaks, all five prolines in the calculated structures adopt the *trans*-conformation. The conformations of the side chains of the six disulfides are not defined absolutely, but four show a distinct conformationals preference; Cys29–Cys59, Cys47–Cys66 and Cys34–Cys38 for a right-handed spiral and Cys8–Cys16 for a left-handed spiral (28).

Figure 2 shows a side-by-side stereo representation of one calculated structure, and an alignment of all twelve structures is illustrated in Fig. 3. Locally, each of the loops of albolabrin is well defined (Fig. 4a), but, due to the paucity of long range NOEs, their relative orientations are more loosely defined. When the whole structure is aligned, the average atomic RMS difference from the geometric mean for all backbone atoms is 3.1 Å and for all non-hydrogen atoms is 3.8 Å (Fig 4b). Residues 1–9 are unstructured in our calculations as they show no NOEs. Residues 1–14 fail to align with the body of the protein as a result of the small number of distance constraints at Pro14. In the hairpin Cys15 to Leu24, the stem of the hairpin is well defined, but the orientation of the loop Asp17–Cys21 is unconstrained

TABLE 1

Deviations from experimental constraints and idealised geometry for the 20 lowest-energy structures[a]

139.5±32.1 (73.3–191)	F(NOE constraints) (kcal mol^{-1})
15.4±8.0 (5.5–33.0)	F(repulsive constraints) (kcal mol^{-1})
58.2±12.4 (28.2–78.9)	F(Repel) (kcal mol^{-1})
2.9±0.6	Dihedral constraints r.m.s.d. (°)
5.2±0.4 ($\times 10^{-3}$)	Bond r.m.s.d. (Å)
0.75±0.06	Angle r.m.s.d. (°)
0.58±0.08	Improper r.m.s.d. (°)

[a] Structures were calculated as in the methods section. The energies of the structures were calculated at the end of the 'refine' protocol in X-PLOR following 500 steps of Powell minimisation. Mean values ± standard deviations of the calculated energies are reported, with the maximum and minimum values for the 20 structures given in parentheses. Deviations in bond lengths, angles and impropers are reported as r.m.s. deviations from ideal values.

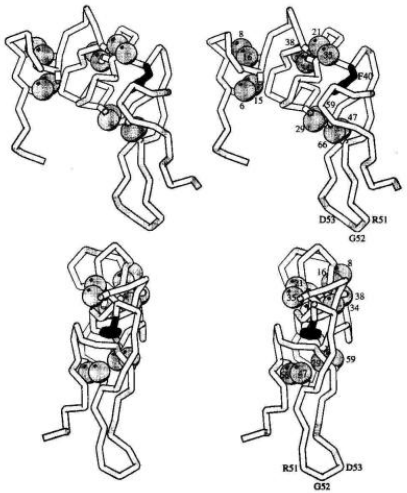

FIGURE 2

Calculated structure of albolabrin. The figure shows two approximately orthogonal views of a single structure of albolabrin, chosen as the calculated structure closest to the geometrical mean of the 12 aligned structures seen in Fig. 4. Side-by-side stereo representations are shown. The α-carbon trace for the whole protein is shown, with the sidechain carbon atoms for the residue Phe40 and van der Waal spheres at the sulfur atoms of the cysteine residues. The cysteines and the RGD sequence that binds to αIIbβ3 are numbered. This figure was produced using the program MOLSCRIPT (44).

relative to the stem. Likewise, the orientation of the hairpin Cys15 to Leu24 relative to the body of the protein is poorly defined since the only constraint in this region is the disulfide bridge Cys21–Cys35. The RGD-containing hairpin at residues 47–57 aligns poorly with the remainder of the protein, suggesting a degree of independent mobility. When only the hairpin itself is aligned, it is clear that the loop at the tip of the hairpin containing the RGD sequence is particularly poorly defined, probably indicating a high degree of mobility (Fig. 3). A prominent, fully solvent exposed shoulder consisting of residues Lys42 and Lys43 lies vertically above the RGD loop but its position relative to the core of the protein is poorly defined, which may indicate mobility in this area.

The best defined and most densely packed region of albolabrin extends from residues 22–47 and 57–67, and forms the core of the protein. On restricting the region compared to residues 22–47 and 57–67 (Fig. 3), the average atomic RMS difference from the geometric mean for backbone atoms is reduced to 1.3 Å and for all non-hydrogen atoms is 2.1 Å. Phe40 is the central residue of the core of the protein (Fig. 2) and has numerous NOEs (Fig. 1). Phe40 is substantially buried in the structure (Fig. 1), together with residues Leu23, Ala27, Gln28, Leu33, Cys34, Ser39, Met45, Cys47, Asn60, Gly61, Ile62 and Ala64, which thus form the protein core. The βCH_2 resonances of residue Gln28 experience a particularly large upfield shift as a result of their interaction with Phe40. Four of the six disulfide bridges (21–35 and 34–38; 29–59 and 47–66) make a substantial contribution to the core, but lie at the periphery.

Overall, albolabrin is an asymmetric protein with a maximum length of ca. 40 Å, and a maximum width and depth of ca. 10–12 and 32 Å respectively, the latter distance being measured from Lys42 to the tip of the RGD loop at Gly52. The hydrophobic protein core is surrounded by charged and polar residues that form an interface with the solvent. Outside the protein core, the tertiary fold is stabilised principally by the covalent disulfide bridges which tie together the loops comprising the secondary structure. This stabilisation is probably an entropic effect arising from a decrease in conformational freedom in the unfolded state. The disulfide bridges are clustered into pairs (6–15 and 8–16; 21–35 and 34–38; 29–59 and 47–66). An irregular β-hairpin carrying the RGD adhesion sequence protrudes from the body of the protein. The antiparallel β-strands of the RGD-containing hairpin are somewhat twisted, so that the strand Asp56–Cys59 faces into the body of the protein, placing the broad loop Asn60–Ala64 in such a position that these residues can contribute to the hydrophobic core of the protein. The disulfide bridge Cys47–Cys66 folds the *C*-terminal segment back towards the RGD hairpin. This places the His72 close to the highly charged RGD-containing hairpin, and in many structures an electrostatic interaction is indicated with residue Asp57.

The charge distribution of albolabrin is highly asymmetric. The whole of the *N*-terminal portion of the protein forms a negative pole extending as far as the innermost strand of the RGD hairpin (Asp53–Cys59; only residue Lys22 and the free *N*-terminus break the pattern). The line of demarcation between positive and negative poles runs down the centre of the RGD-containing hairpin, so that the outermost strand of the hairpin (Thr45–Arg51) and the *C*-terminal portion of the protein are positively charged (only the free *C*-terminus breaks the pattern).

DISCUSSION

Comparison with kistrin and flavoridin

Comparison of the secondary and tertiary structural elements of albolabrin with those reported for other disintegrins (6, 10, 12, 14, 17) shows that most

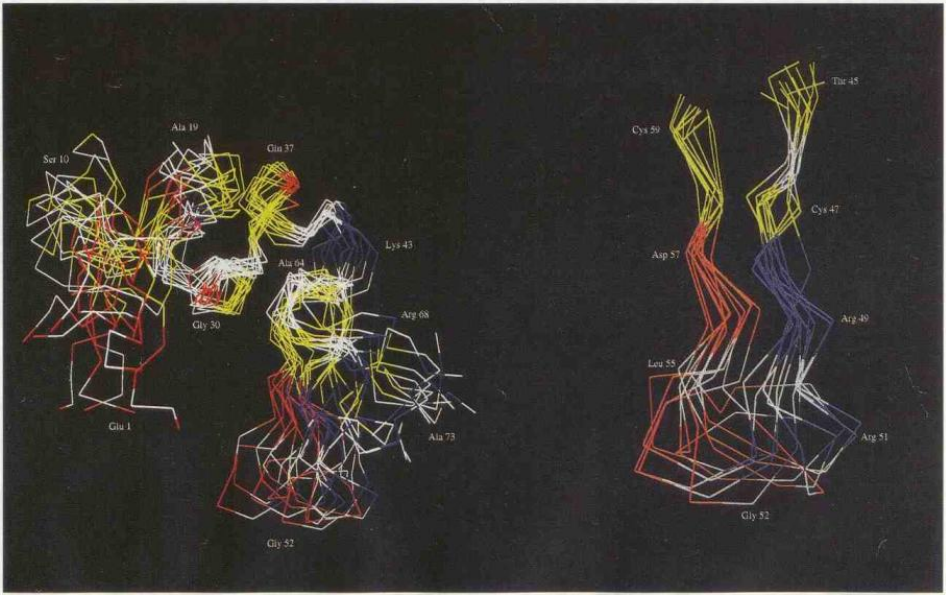

FIGURE 3

Illustration of the alignment of the 12 chosen calculated structures. Left: The 12 calculated structures of albolabrin, aligned on the backbone atoms of residues 22–47 and 57–67. Only the α-carbons are shown. Each residue is coloured by its polarity (blue = basic, red = acidic, yellow = polar, white = hydrophobic). Right: An alignment of the residues of the RGD-containing hairpin (Thr45–Cys59). Selected residues have been numbered.

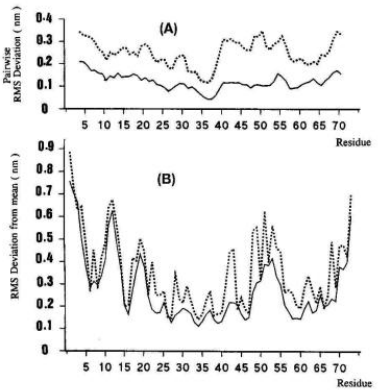

FIGURE 4

Alignment of the 12 chosen calculated structures. Panel A (above): The mean pairwise RMS deviation (in Å) for pentapeptide segments of the calculated structures. Panel B (below): The RMS atomic deviation (in Å) of each residue from the geometric mean structures after alignment over the whole structure. In each panel the solid line represents the backbone atoms (αC, N, C), and the dotted line represents all non-hydrogen atoms.

features are well conserved despite the differences in disulfide bonding pattern. In all the medium-length disintegrins, albolabrin, kistrin and flavoridin, two of the short regions of β-sheet are conserved: that at the neck of the hairpin containing the RGD sequence and also that across the neck of the hairpin between residues Cys34–Met41 (residues Gly33–Lys37 in kistrin; residues Cys33–Phe38 in flavoridin). Most of the turns are conserved, including the open turn between residues Asn60–Ala64 (residues Thr58–Cys64 in kistrin; residues Leu60–Asp63 in flavoridin) which is important in the protein core. The same residues appear to be involved in forming the central hydrophobic core in albolabrin, kistrin and flavoridin (e.g. the close association of Gln26 with Phe38 in kistrin). The central role of Phe40 in the protein core is confirmed by the conservation of this residue in all disintegrin sequences. The essential role of the residue Phe40 and residues in the loop Asn60–Ala64 appears to be confirmed by alanine scanning mutagenesis of non-cysteine residues in kistrin, where the only mutants which could not be expressed were for residues Phe38 and Gln60, presumably because the mutation to alanine resulted in the loss of important long-range contacts (29). However, the topology of the hydrophobic area differs between albolabrin and kistrin, in that it involves more extensively residues in the short β-sheet at the top of the hairpin containing the RGD sequence in kistrin, and also in that there are cysteine residues placed more centrally within the core (the disulfide bridges 32–57 and 33–27 in kistrin, 32–57 and 45–62 in flavoridin).

Examination of the NOEs in all the disintegrins shows that there are very few long-range contacts between residues outside the core of the protein. All the proteins appear to be highly mobile, in particular at the RGD loop. The presence of a small protein core and hence a highly solvent-exposed structure may contribute to this internal mobility. Although the disulfide bridges stabilise the tertiary fold, locally they may destabilise the protein structure by the introduction of dihedral stresses and by the disruption of potential tertiary contacts between elements of secondary structure in the region of the disulfide bridge, as shown by protein engineering in a variety of proteins (30–33). This would lead to increased mobility. The lack of long-range contacts in the disintegrins would allow rearrangements of disulfide bridges to be made between homologous proteins, without disrupting important non-covalent contacts between the loops and hairpins. This is confirmed by the calculations performed with different disulfide binding patterns having similar energies. The precise arrangement of the loops in space and the location of the most mobile regions in a particular protein depend on the arrangement of disulfide bridges. For example, in flavoridin there is a mobile *N*-terminal domain (residues 1–25) and a better defined *C*-terminal domain (12). Applying the disulfide bridge pattern of flavoridin to a structural calculation for albolabrin produced a narrow 'bridge' between the *N*- and *C*-terminal domains formed from the polypeptide chain of residues 25–26 and a disulfide bridge Cys15–Cys38. This would appear to correspond to the mobile hinge region observed in flavoridin. Although there is a dramatic decrease in potency of the disintegrins when the disulfides are chemically reduced, so that the disulfide bonding is important for activity (e.g. ref. 5), the precise arrangement of the loops may not be essential to the function of the disintegrins. Nonhomologous proteins such as decorsin (34) and mambin (35) have a similar potency to the disintegrins, but have very different structures (36, 37), and a functional RGD loop can be introduced into other proteins such as lysozyme and immunoglobulins (38–40). In all these proteins the tripeptide RGD sequence is at the tip of a hairpin, projecting from the body of the protein. The mobility of this loop and the internal flexibility of the disintegrins may be important to their function. This may contribute to their ability to bind to unactivated αIIbβ3, where the binding site is thought to be buried, as well as to the activated integrin. In addition a disintegrin protein might be able to adopt different conformations when bound to different receptors. A similar function has been attrib-

uted to segmental mobility in antigen binding sites of antibodies (41–43).

The asymmetrical charge distribution seen in albolabrin is preserved in the other disintegrins. This may play a role in disintegrin function, as in alanine scanning mutagenesis of kistrin changes of Arg46 and Arg66 were among the few mutations that gave decreased potency of platelet aggregation (29). The charge distribution may be important for interaction with the $\alpha IIb\beta 3$ receptor, especially as there is no obvious exposed large hydrophobic patch on these proteins.

ACKNOWLEDGEMENTS

This work was supported by the Science and Engineering Research Council UK (K.J.S.) and the Wellcome Trust UK (M.J.). X.L. thanks Prof. V. Kakkar for continued support. We thank Mr. A.J. Pemberton for maintenance of the NMR and computer facilities. Coordinates for albolabrin will be submitted to the Brookhaven Protein DataBank.

REFERENCES

1. Williams, J.A. (1992) *Pathol. Biol.* **40**, 813–821
2. Ruoslahti, E. & Pierschbacher, M.D. (1987) *Science* **227**, 491–479
3. Gould, R.J., Polokoff, M.A., Friedman, P.A., Huang, T-F. Cook, J.J. & Niewiarowski, S. (1990) *Proc. Soc. Exp. Biol. Med.* **195**, 168–171
4. Dennis, M.S., Henzel, W.J., Pitti, R.M., Lipari, M.T., Napier, M.A., Diesher, T.A., Bunting, S. & Lazarus, R.A. (1990) *Proc. Natl. Acad. Sci. USA.* **87**, 2471–2475
5. Calvete, J.J., Schafer, W., Soszka, T., Lu, W., Cook, J.J., Jameson, B.A. & Niewiarowski, S. (1991) *Biochemistry* **30**, 5225–5229
6. Adler, M., Carter, P., Lazarus, R.A. & Wagner, G. (1993) *Biochemistry* **32**, 282–289
7. Calvete, J.J., Wang, Y., Mann, W., Schafer, W., Niewiarowski, S. & Stewart, G.J. (1992) *FEBS Lett.* **309**, 316–320
8. Klaus, W., Broger, C., Gerber, P. & Senn, H. (1993) *J. Mol. Biol.* **232**, 897–906
9. Cooke, R.M., Carter, B.G., Murray-Rust, P., Hartshorn, M.J., Herzyk, P. & Hubbard, R.E. (1992) *Protein Eng.* **5**, 473–477
10. Adler, M., Lazarus, R.A., Dennis, M.S. & Wagner, G. (1991) *Science* **253**, 445–448
11. Adler, M., & Wagner, G. (1992) *Biochemistry* **31**, 1031–1039
12. Senn, H. & Klaus, W. (1993) *J. Mol. Biol.* **232**, 907–925
13. Cooke, R.M., Carter, B.G., Martin, D.M.A., Murray-Rust, P. & Weir, M.P. (1991) *Eur. J. Biochem.* **202**, 323–328
14. Dalvit, C., Widmer, H., Bovermann, G., Breckenridge, R. & Metternich, R. (1991) *Eur. J. Biochem.* **202**, 315–321
15. Chen, Y., Pitzenberg, S.M., Garsky, V.M., Lumma, P.K., Sanyal, G. & Baum, J. (1991) *Biochemistry* **30**, 11625–11636
16. Saudek, V., Atkinson, R.A. & Pelton, J.T. (1991) *Biochemistry* **30**, 7369–7372
17. Atkinson, R.A., Saudek, V., & Pelton, J.T. (1994) *Int. J. Peptide Protein Res.* **43**, 563–572
18. Jaseja, M., Smith, K.J., Lu, X., Williams, J.A., Trayer, H., Trayer, I.P. & Hyde, E.I. (1993) *Eur. J. Biochem.* **218**, 853–860
19. Wüthrich, K., Billeter, M. & Braun, W. (1983) *J. Mol. Biol.* **169**, 949–961
20. Clore, G.M., Gronenborn, A.M., Nilges, M. & Ryan, C.A. (1987) *Biochemistry* **26**, 8012–8023
21. Pullman, B. & Pullman, A. (1974) *Adv. Protein. Chem.* **28**, 347–526
22. Brünger, A.T. (1992) *X-PLOR 3.1 Manual: A System for X-Ray Crystallography and NMR*, Yale University Press
23. Holak, T., Nigles, M. & Oschkinat, H. (1989) *FEBS Lett.* **242**, 218–224
24. Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R.C., Laue, D.E. & Thomas, J.O. (1993) *EMBO J.* **12**, 1311–1319
25. Güntert, P., Braun, W. & Wüthrich, K. (1991) *J. Mol. Biol.* **217**, 517–530
26. Güntert, P., Qian, Y.P., Otting, G., Muller, M., Gehring, W. & Wüthrich, K. (1991) *J. Mol. Biol.* **217**, 531–540
27. Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. (1989) *J. Am. Chem. Soc.* **111**, 3997–4004
28. Richardson, J.S. (1981) *Adv. Protein Chem.* **34**, 167–337
29. Dennis, M.S., Carter, P. & Lazarus, R.A. (1993) *Proteins: Struct. Func. Genet.* **15**, 312–321
30. Pace, C.N., Grimsley, G.R., Thomson, J.A. & Barnett, B.J. (1988) *J. Biol. Chem.* **263**, 11820–11825
31. Katz, B.A. & Kossiakoff, A. (1986) *J. Biol. Chem.* **261**, 15480–15488
32. Villafranca, J.E., Howell, E.E., Oatley, S.J., Xuong, N. & Kraut, J. (1987) *Biochemistry* **26**, 2182–2189
33. Wetzel, R. (1987) *Trends Biochem. Sci.* **14**, 478–482
34. Seymour, J.L., Henzel, W.J., Nevins, B., Stults, J.T. & Lazarus, R.A. (1990) *J. Biol. Chem.* **265**, 10143–10147
35. McDowell, R.S., Dennis, M.S., Louie, A., Shuster, M., Mulkerrin, M.G. & Lazarus, R.A. (1992) *Biochemistry*, **31**, 4766–4772
36. Krezel, A.M., Wagner, G.M Seymour-Ulmer, J., & Lazarus, R.A. (1994) *Science* **264**, 1944–1947
37. Sutcliffe, M.J., Jaseja, M., Hyde, E.I., Lu, X. & Williams, J.A. (1994) *Nature Struct. Biol.* **1**, 802–807
38. Lee, G., Chan, W., Hurle, M.R., DesJarlais, R.L., Watson, F. Sathe, G.M. & Wetzel, R. (1993) *Protein Eng.* **6**, 745–754
39. Yamada, T., Matsushima, M., Inaka, K., Ohkubo, T., Uyeda, A., Maeda, T., Titani, K., Sekiguchi, K. & Kikuchi, M. (1993) *J. Biol. Chem.* **268**, 10588–10592
40. Hashino, K., Shimojo, T., Kimizuka, F., Kato, I., Maeda, T., Sekiguchi, K. & Titani, K. (1992) *J. Biochem.* **112**, 547–551
41. Rini, J.M., Schulzegahmen, U., & Wilson, I.A. (1992) *Science* **255**, 959–965
42. Westhof, E., Altschuh, D., Moras, D., Bloomer, A.C., Mondragon, A., Klug, A. & van Regenmortel, M.H.V. (1984) *Nature (London)* **311**, 123–126
43. Tainer, J.A., Getzoff, E.D., Alexander, H., Houghten, R.A., Olson, A.J., Lerner, R.A. & Hendrickson, W.A. (1984) *Nature (London)* **312**, 127–134
44. Kraulis, P. (1991) *J. Appl. Crystallogr.* **24**, 946

Address:

Prof. *I.P. Trayer*
School of Biochemistry
University of Birmingham
Birmingham
B15 2TT
UK
Tel: 021–414–5401
Fax: 021–414–3982