Three-Dimensional Structure of Echistatin, the Smallest Active RGD Protein

Vladimir Saudek, R. Andrew Atkinson, and John T. Pelton*

Marion Merrell Dow Research Institute, 16 rue d'Ankara, B.P.447/R9, 67009 Strasbourg Cedex, France

Received April 9, 1991; Revised Manuscript Received May 23, 1991

ABSTRACT: Echistatin is a 49 amino acid protein isolated from the venom of a viper (*Echis carinatus*) and is one of the smallest natural adhesive ligands that interacts with integrin-type receptors through an Arg-Gly-Asp (RGD) sequence. The structure of echistatin in aqueous solution has been determined by nuclear magnetic resonance spectroscopy. Nuclear Overhauser spectra yielded 490 interatomic distance constraints, which were used in distance geometry calculations. The chain is shown to fold in a series of irregular loops to form a rigid core stabilized by four cystine cross-links. From this core an irregular hairpin and the C-terminus protrude. The core and the hairpin are further stabilized by a network of hydrogen bonds. The RGD sequence is located in a mobile loop at the tip of the hairpin. The mobility and its significance for activity are discussed.

Cell–matrix and cell–cell adhesion are mediated by surface receptors called integrins (Hynes, 1987). Their natural ligands are large proteins with complex functions enabling communcation between the interior and exterior of the cell. The main recognition site in many of these protein ligands is a simple sequence of arginine, glycine, and aspartic acid (RGD in one-letter code; Ruoslahti & Pierschbacher, 1987). However, since integrins from different cells interact with the RGD ligands with differing specificities, other factors must also be important for the recognition. The specificity might reside in the particular conformation of the RGD sequence in the individual proteins, in the amino acids adjacent to the RGD sequence, or in a more remote part of the ligand sequence. Elucidation of the conformation of the integrin-binding ligands, or at least of the RGD sequence each contains, should help to answer this question. Unfortunately, it will not be straightforward to solve the structures of such complex proteins.

Recently a number of potent integrin inhibitors ("disintegrins"; Gould et al., 1990) have been isolated from the venom of various vipers. They are all cysteine rich and highly homologous, and each contains an RGD sequence. However, they differ somewhat in their activity. As they are relatively small (ca. 50–80 amino acids), they provide a unique opportunity to gain an insight into the three-dimensional structure of RGD active proteins and the factors that are important in controlling specificity. Although the primary sequences of almost 20 different disintegrins have been reported, there is at present no information on their conformation or on the arrangement of disulfide bridges. Here we report the three-dimensional structure of echistatin (49 amino acids, first isolated from *Echis carinatus*; Gan et al., 1988), one of the smallest known natural disintegrins and also one of the smallest known RGD active proteins.

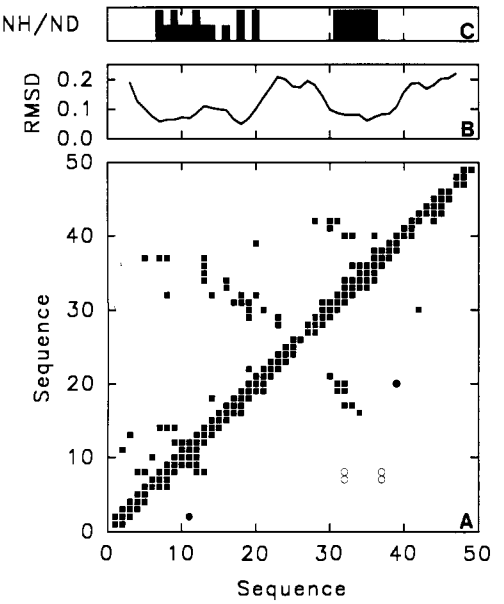

FIGURE 1: NOE map for echistatin and assessment of the definition of local conformation. (A) Map of the NOE connectivities detected in echistatin. Each square indicates at least one NOE between the residues in the sequence, plotted on each axis. NOEs involving side-chain resonances are plotted above the diagonal; those involving only main-chain resonances are plotted below the diagonal. Circles indicate disulfide bridges: filled circles denote those that can be unambiguously defined, while open circles denote those that remain uncertain (see text). (B) The mean pairwise RMSD (in nanometers) for pentapeptide backbone segments of the calculated structures, plotted aginst the sequence number of the central residue in the segment. (C) Rates of NH/ND exchange: bars indicate residues with backbone amide protons which were detected after 12 h (short bars) and after 1 week (tall bars) in D_2O. Sequence of echistatin α1 (Gan et al., 1988): ECESG PCCRN CKFLK EGTIC KRARG DDMDD YCNGK TCDCP RNPHK GPAT.

Materials and Methods

Echistatin α1 was synthesized by standard solid-phase synthesis methods similar to those described previously (Garsky et al., 1989), purified by a combination of gel-filtration and HPLC techniques, and characterized by electrospray mass spectrometry, amino acid analysis, gas-phase sequencing, and optical spectroscopic techniques. The CD spectrum of the synthetic echistatin was identical with that previously published for the peptide (Garsky et al., 1989), consistent with an identical pairing of the disulfide bonds and backbone conformation. The biological activity of the synthetic peptide, assessed in a standard human platelet aggregation assay (Sigma Diagnostics platelet aggregation assay kit), was 34.1 ± 4.5 nM, in excellent agreement with that found for the native peptide (Garsky et al., 1989). The primary sequence is given in the legend to Figure 1. NMR[1] spectra were

* Author to whom correspondence should be addressed.

0006-2960/91/0430-7369$02.50/0

Table I: Structural Statistics for DISMAN-Calculated Echistatin Structures[a]

	root mean square violation (nm)								
	all (490)		intraresidue (215)		medium range (184)		long range (91)		RMSD
structure	A	B	A	B	A	B	A	B	(A/B)
1	0.0118	0.0115	0.0059	0.0060	0.0159	0.0162	0.0124	0.0123	0.102
2	0.0117	0.0119	0.0047	0.0044	0.0151	0.0152	0.0148	0.0156	0.109
3	0.0126	0.0125	0.0085	0.0089	0.0147	0.0152	0.0155	0.0134	0.072
4	0.0119	0.0119	0.0050	0.0054	0.0147	0.0150	0.0163	0.0155	0.108
5	0.0122	0.0126	0.0078	0.0074	0.0163	0.0162	0.0111	0.0142	0.086
6	0.0124	0.0129	0.0063	0.0083	0.0165	0.0161	0.0134	0.0145	0.068
7	0.0128	0.0125	0.0066	0.0066	0.0143	0.0157	0.0191	0.0155	0.077
8	0.0131	0.0131	0.0068	0.0060	0.0167	0.0173	0.0158	0.0155	0.116
mean	0.0123	0.0124	0.0065	0.0067	0.0155	0.0159	0.0148	0.0146	0.092

[a] Data are shown for the eight best structures calculated with each of the possible arrangements of disulfide bridges (A, Cys 7/Cys 32 and Cys8/Cys 37; B, Cys 7/Cys 37 and Cys 8/Cys 32). The root mean square violation of experimental distance constraints was calculated for all constraints and for subclasses according to the separation of the residues $(i, i + j)$ in the sequence (medium range, $1 \leq j \leq 4$; long range, $j \geq 5$). The number of constraints in each class is indicated in parentheses. RMSD values between corresponding pairs of structures with different arrangements of the disulfide bridges were calculated by using the backbone atoms of residues Gly 5–Ile 19 and Asp 29–Pro 40.

acquired on a Bruker AM500 spectrometer under the following conditions: 2 mM echistatin in H_2O or D_2O containing 0.05 M deuterated acetic acid, pH 3.2, 25 °C. Full experimental details are described elsewhere (Saudek et al., 1991).

Distance constaints were derived from NOESY spectra recorded with mixing times between 40 and 200 ms. Interproton distances were assigned to the NOE cross peaks by using an empirical calibration based upon the minimum and maximum possible distances between the CHα and the backbone amide protons of sequential pairs of residues (Wüthrich et al., 1983; Wüthrich, 1986; Widmer et al., 1989). Cross peaks detected in spectra recorded with a mixing time of 40 ms were divided into three groups according to their intensities and assigned distances of 0.27, 0.32, and 0.37 nm. NOEs first detected at mixing times of 60, 80, 100, or 200 ms were assigned distances of 0.40, 0.42, 0.45, or 0.50 nm, respectively. In the absence of stereospecific assignments, pseudoatoms were used with the appropriate corrections to interproton distances (Wüthrich et al., 1983).

Structures were calculated with the distance geometry program DISMAN (Braun & Go, 1985) by using the list of distances as upper-limit constraints on interproton distances and the sum of van der Waals radii as lower-limit constraints on the distances between nonbonded atoms. Disulfide bridges were also specified as distance constraints (Braun & Go, 1985). After preliminary calculations to identify inconsistencies in the input, DISMAN was run as described by the authors with the following modifications to accommodate the uncertainty in the pattern of the disulfide bridges (see Results and Discussion). At each level, only interactions between residue pairs i,j were considered for which $|i - j| \leq n$. Initially, 1500 structures were generated by progressively including distance constraints for which $n \leq 23$ ("level 23"). (The two disulfide bridges that are uncertain are between residue pairs for which $n > 23$.) Van der Waals contributions to the target function were scaled down by a factor of 0.1 until level 23, at which the weighting was restored to 1.0. Two sets of calculations were then performed in parallel for each of the possible patterns of disulfide bridges. The 64 structures that had the lowest values of the target function at level 23 were taken further by minimizing the target function at the levels of the disulfide bridges and then at level 49. The van der Waals contributions were again scaled down during the intermediate steps. The eight best structures from each set of calculations were selected for analysis.

Pairwise RMSD values were generated within the SYBYL molecular modeling package (Tripos Associates, St. Louis, MO, version 5.4) and averaged. The most variable regions were identified and omitted from the final fitting. RMSD values over pentapeptide segments were generated and averaged in a similar manner (Widmer et al., 1989).

Results and Discussion

The complete assignment of the NMR spectrum of echistatin (Saudek et al., 1991) made possible the compilation of a list of 490 interproton distances from the NOESY spectra (Table I, Figure 1). The disulfide cross-links between the cysteine pairs 2/11 and 20/39 were established by the detection of CHβ_i/CHβ_j and CHβ_i/CHα_j NOEs detected for both pairs and are in agreement with the results of degradation studies (P. Lepage, unpublished results). The four remaining cysteine residues were found to be located in close spatial proximity (Figure 1). The results of the calculations run in parallel with the 7/32–8/37 or 7/37–8/32 pattern of disulfide bridges led to essentially identical three-dimensional structures (Table I). (The third possibility, 7/8–32/37, was excluded, as this pattern was found to be incompatible with the NOEs). Although the results of the calculations did not allow the identification of the correct set of cross-links, the uncertainty did not impair the structure determination at the present level of resolution. In the following, only the 7/32–8/37 pattern are displayed, but all our data and conclusions are fully compatible with the 7/37–8/32 pattern.

All calculations converged to the same global fold. The superposition of the eight structures with the lowest residual violations of experimental constraints is shown in Figure 2 and a numerical analysis is presented in Table I. Although all structures satisfy the experimental constraints within their errors, they diverge in certain regions. The local variability of the backbone conformation is displayed in Figure 1. This local variability reflects a lower number of distance constraints in certain regions (Figure 1) and is a consequence of a greater conformational space being available for these segments.

The structure of echistatin might be described as a bundle of loops (Figure 3) containing no regular secondary structure. The disulfide bridges stabilize the central rigid part from which an irregular β-hairpin and the C-terminus protrude. The most mobile parts of the protein are clearly the termini and the RGD-containing loop (Figure 2). The hairpin is formed by

[1] Abbreviations: NMR, nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser effect; RGD, Arg-Gly-Asp sequence; RMSD, root mean square deviation.

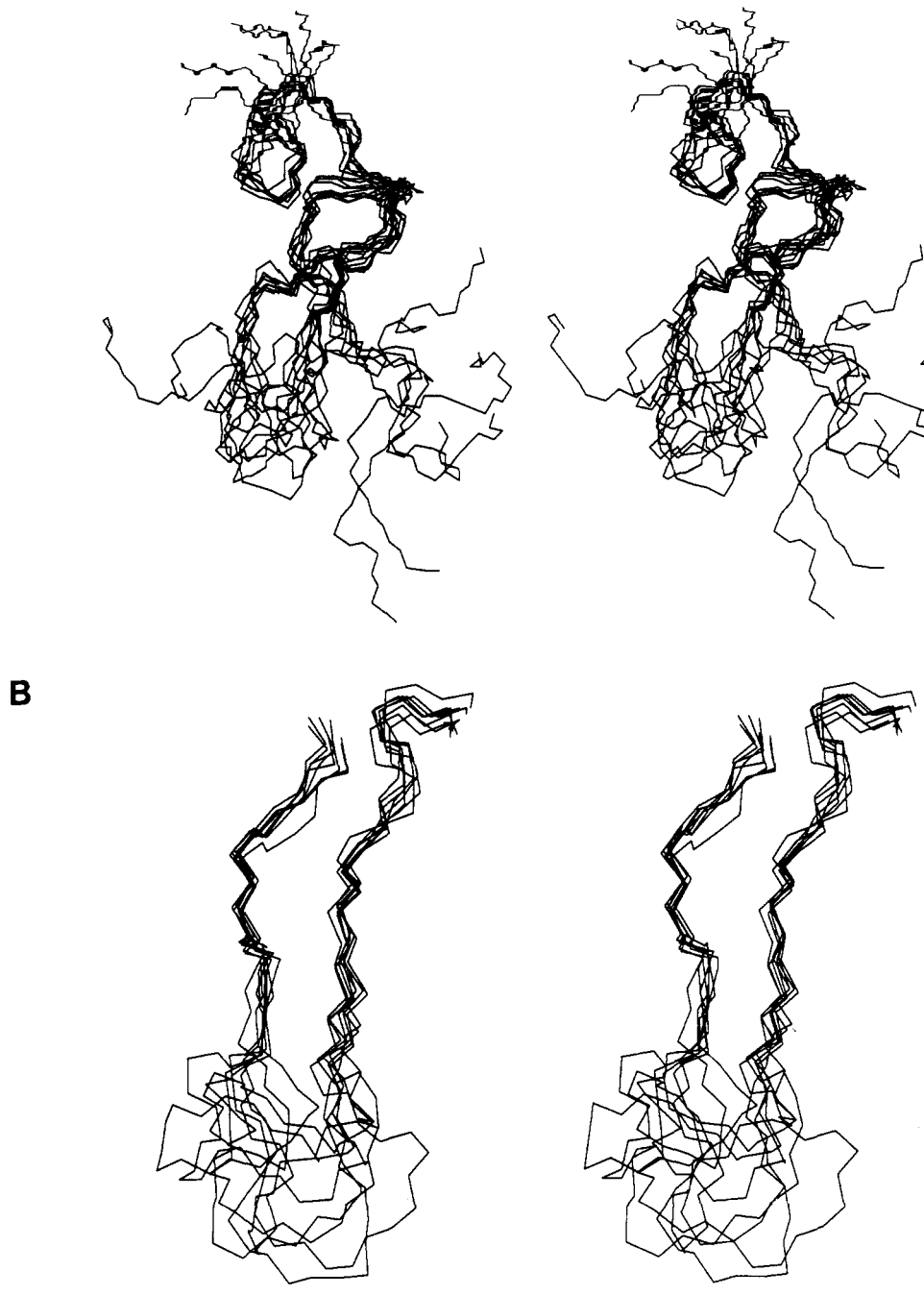

FIGURE 2: Stereoviews of the structure of echistatin calculated by DISMAN. Eight structures are shown. (A) Full backbone structures: the backbone atoms of residues 5–19 and 29–40 were fitted to structure 1A (see Table I). (B) The irregular hairpin containing the RGD sequence, shown from a different angle: the backbone atoms of residues 16–21 and 28–35 were fitted to structure 1A.

two extended antiparallel strands, but their arrangement does not correspond to a regular antiparallel β-sheet. The RGD sequence is forced into a loop at the tip of the hairpin. Furthermore, the chemical shifts of the resonances of the RGD sequence are almost identical with those of an isolated RGD peptide (Saudek et al., 1991), indicating that there is not a significant difference between the environments of the sequence in the two molecules.

The differences in mobility are indicated by the line shapes of some of the resolved CH_3 groups in the one-dimensional spectrum (e.g., A23, A48, and T49), the local RMSD values, and above all, by the rates of NH/ND exchange (Figure 1). A prolonged stability for 38% of the backbone amide protons of echistatin in D_2O indicates an important network of hydrogen bonds. These are all located in the cysteine cross-linked part of the molecule and in the antiparallel strands of the hairpin. In contrast, all the backbone amide hydrogens of the termini and the RGD loop are in rapid NH/ND exchange. The resolution of the structure so far achieved allows the identification of some of the CO–NH pairs, e.g. residues 8–5, 12–8, and 20–30. Remarkably, one of the most stable hydrogen bonds, CO_{20}–NH_{30} (detectable in D_2O for more than 1 week) is located at the border between the well-defined strands of the hairpin and the much more mobile RGD loop (Figure 2B).

Apart from the restriction imposed by the strands of the hairpin, the RGD recognition site is very mobile and exposed. Such behavior is typical of small segments of proteins whose function is fast recognition and fitting (Williams, 1989). The "hand in glove" model describes how the speed is paid for by somewhat reduced specificity (Williams, 1978). It is clear that the RGD sequence alone does not account for all the specificity

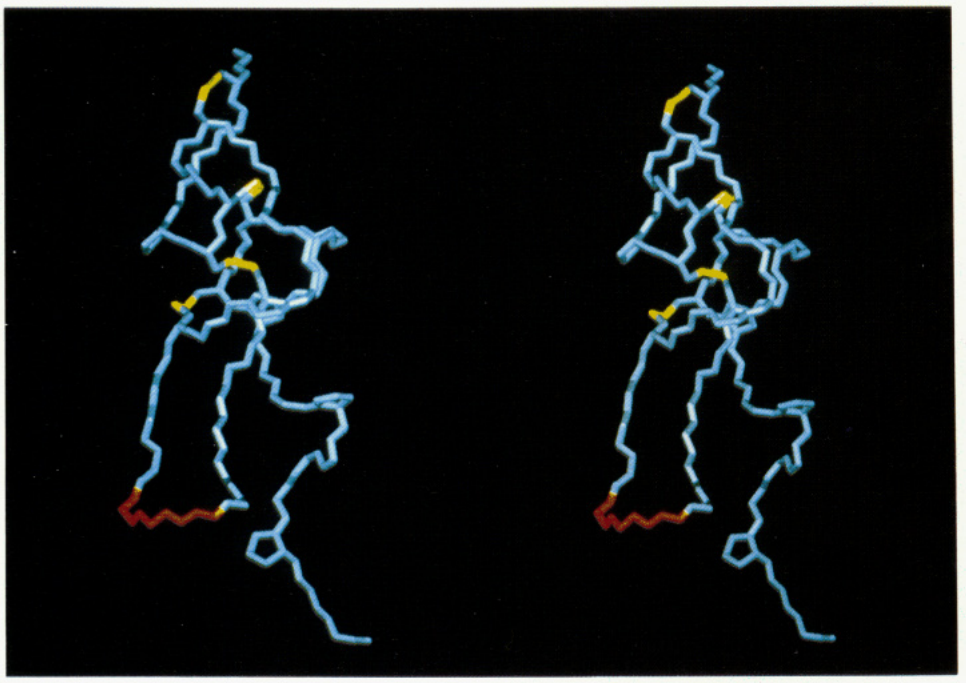

FIGURE 3: Stereoview of the three-dimensional structure of echistatin. The entire backbone and the proline side chains are shown for structure 1A (Table I). The RGD sequence is highlighted in red, and disulfide bridges are shown in yellow.

of echistatin (Ruoslahti & Pierschbacher, 1987). It is possible that some specificity resides in the extended chains of the hairpin but their main function is clearly structural, i.e., to enforce the loop and perhaps to provide a spacer between the rigid cross-linked core and the critical RGD sequence.

In the particular case of echistatin, we can speculate that the C-terminus is also important for specificity. It has been observed that echistatin loses some of its ability to inhibit platelet aggregation if this segment is deleted (Garsky et al., 1989). Its position relative to the main body of the molecule is not well-defined, but its conformation resembles a collagen-like polyproline helix containing three loose proline turns. All three prolines are seen to be stabilized in the trans conformation (Saudek et al., 1991). It contains a tetrapeptide sequence, PRNP, which is also found in fibrinogen and some other disintegrins (Gould et al., 1990).

The extent of the conservation of amino acid sequence within the disintegrin family suggests that the three-dimensional structures of the domain homologous to echistatin will be very similar. In the larger disintegrins, further specificity may be found in the domain that is absent in echistatin. For instance, some homologies between a distintegrin (trigramin), human von Willebrand factor, and collagen have been noted (Huang et al., 1989), suggesting that they may represent a common binding site. Since Cys-37 is not conserved in the larger disintegrins but all cysteines are always oxidized (Gould et al., 1990), we can also deduced that the N-terminal domain is cross-linked to the echistatin-like domain through either Cys-7 or -8.

In summary, the first three-dimensional structure of a member of the disintegrin family has been determined and the variation in mobility within the molecule has been described. The main recognition site (the RGD sequence) has been shown to be located in a mobile loop held apart from a more rigid, cross-linked, structural core by a spacer formed by the two strands of an irregular hairpin. The refinement of the structure on the basis of stereospecific assignments, the identification of hydrogen bonds, and the use of coupling constants is in progress. Given the similarity in function, it seems probable that the RGD segments of the larger integrin-binding ligands will be found in similar environments.

ACKNOWLEDGMENTS

We are grateful to C. Brockel and D. J. Cowley for the platelet aggregation assays, to M. Hibert, B. Maigret, and A. Pastore for valuable discussions, and to J. Hoflack and S. Trumpp-Kallmeyer for their help with the use of SYBYL software.

REFERENCES

Braun, W., & Go, N. (1985) *J. Mol. Biol. 186*, 611–626.
Gan, Z. R., Gould, J. R., Jacobs, J. W., Friedman, P. A., & Polokoff, M. A. (1988) *J. Biol. Chem. 263*, 19827–19832.
Garsky, V. M., Lumma, P. K., Freidinger, R. M., Pitzenberger, S. M., Randall, W. C., Veber, D. F., Gould, R. J., & Friedman, P. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4022–4026.
Gould, R. J., Polokoff, M. A., Friedman, P. A., Huang, T.-F., Holt, J. C., Cook, J. J., & Niewiarowski, S. (1990) *Proc. Soc. Exp. Biol. Med. 195*, 168–171.
Huang, T.-F., Holt, J. C., Kirby, E. P., & Niewiarowski, S. (1989) *Biochemistry 28*, 661–666.
Hynes, R. O. (1987) *Cell 48*, 549–554.
Ruoslahti, E. (1988) *Annu. Rev. Biochem. 57*, 375–413.
Ruoslahti, E., & Pierschbacher, M. D. (1987) *Science 238*, 491–497.
Saudek, V., Atkinson, R. A., Lepage, P., & Pelton, J. T. (1991) *Eur. J. Biochem.* (submitted for publication).
Widmer, H., Billeter, M., & Wüthrich, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 357–371.
Williams, R. J. P. (1978) *Angew. Chem., Int. Ed. Engl. 16* (Suppl.), 766–777.
Williams, R. J. P. (1989) *Eur. J. Biochem. 183*, 479–497.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, J. Wiley, New York.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 180*, 949–961.