The EMBO Journal vol.6 no.2 pp.529–537, 1987

The conformations of hirudin in solution: a study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics

G.Marius Clore, Dinesh K.Sukumaran, Michael Nilges, Jutta Zarbock and Angela M.Gronenborn

Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, FRG

Communicated by R.Huber

The solution conformations of the protein hirudin have been investigated by the combined use of distance geometry and restrained molecular dynamics calculations. The basis for the structure determination comprised 359 approximate interproton distance restraints and 10 ϕ backbone torsion angle restraints derived from n.m.r. measurements. It is shown that hirudin is composed of three domains: a central core made up of residues 3–30, 37–46 and 56–57; a protruding 'finger' (residues 31–36) consisting of the tip of an antiparallel β sheet, and an exposed loop (residues 47–55). The structure of each individual domain is relatively well defined with average backbone atomic r.m.s. differences of <2 Å between the final seven converged restrained dynamic structures and the mean structure obtained by averaging their coordinates. The orientation of the two minor domains relative to the central core, however, could not be determined as no long-range ($|i-j|>5$) interdomain proton–proton contacts could be observed in the two-dimensional nuclear Overhauser enhancement spectra. From the restrained molecular dynamics calculations it appears that the two minor domains exhibit large rigid-body motions relative to the central core.

Key words: hirudin/solution conformations/nuclear Overhauser effect/interproton distances/distance geometry/restrained molecular dynamics

Introduction

The thrombin inhibitor hirudin from the leech *Hirudo medicinalis* is the most powerful natural anticoagulant known (Haycraft, 1884; Markwardt, 1970). It has been characterized as a polypeptide containing 65 residues (Bagdy *et al.*, 1976; Dodt *et al.*, 1984, 1985; Harvey *et al.*, 1986) which exhibits its anticoagulant properties by binding tightly ($K_{ass} \sim 2 \times 10^{10}$ M^{-1}) and specifically to α-thrombin, thereby preventing the cleavage of fibrinogen (Magnusson, 1972; Markwardt, 1985). In contrast to other anticoagulants, hirudin neither interferes with the biosynthesis of clotting factors nor affects additional blood enzymes in the human coagulation cascade (Markwardt, 1985). Furthermore, hirudin does not induce antigenic reactions when administered to either animals or humans (Markwardt *et al.*, 1982, 1984). These remarkable properties of hirudin have generated considerable interest with respect to its potential clinical use as an anticoagulant (Nowak and Markwardt, 1980; Ishikaw *et al.*, 1980; Walsmann and Markwardt, 1981; Koss and Mittman, 1982; Lent, 1986).

Knowledge of the structure of hirudin is a prerequisite for understanding the mechanism of thrombin inhibition. As hirudin itself fails to crystallize and the crystals of the hirudin–thrombin complex obtained to date are unsuitable for X-ray diffraction studies (W.Bode and R.Huber, personal communication) we have initiated a ^{1}H-n.m.r. study of the solution conformations of hirudin. In a recent paper (Sukumaran *et al.*, 1986) we presented the complete assignment of the ^{1}H-n.m.r. spectrum of hirudin and delineated secondary structure elements on the basis of a qualitative interpretation of nuclear Overhauser effects (NOE). In this paper we extend our studies to the investigation of the solution conformations of hirudin based on interproton distance and dihedral angle restraints derived from NMR measurements and calculations combining metric matrix distance geometry (Crippen and Havel, 1978; Havel *et al.*, 1983; Harvel and Wüthrich, 1984,1985) and restrained molecular dynamics (Kaptein *et al.*, 1985; Clore *et al.*, 1985,1986a,b,c; Brünger *et al.*, 1986; Nilsson *et al.*, 1986).

Results and Discussion

Interproton distance and dihedral angle restraints

A set of 359 interproton distance restraints, comprising 139 short-range ($|i-j| \leq 5$) and 63 long-range ($|i-j| > 5$) inter-residue distances and 157 intra-residue distances, were derived from pure phase absorption two-dimensional NOE (NOESY) spectra (Jeener *et al.*, 1979; Macura *et al.*, 1981; Marion and Wüthrich, 1983) recorded in D_2O and H_2O, with mixing times of 100 ms and 200 ms. Examples demonstrating the quality of the NOESY spectra are given in Sukumaran *et al.* (1986). The intra-residue and short-range inter-residue distance restraints were classified into three ranges, 1.8–2.8 Å, 1.8–3.3 Å and 1.8–5.0 Å, corresponding to strong, medium and weak NOEs, respectively (Williamson *et al.*, 1985; Clore *et al.*, 1985, 1986b,c; Kline *et al.*, 1986), while the long-range inter-residue distance restraints were grouped into a single-distance range of 1.8–5.0 Å.

A set of 10 ϕ backbone torsion angle restraints were derived from $^3J_{HN\alpha}$ coupling constants (Pardi *et al.*, 1984) measured by two-dimensional ω_1-scaled double quantum-filtered homonuclear correlated (DQF–COSY) spectroscopy in H_2O (Rance *et al.*, 1983; Brown, 1984). These comprised the ϕ angles of residues 13, 15, 19, 21, 33, 35 and 36 which were restrained to a range of $-80°$ to $-180°$ on account of apparent values of $^3J_{NH\alpha} > 9$ Hz (see Sukumaran *et al.*, 1986).

The NOE distance and ϕ backbone torsion angle restraints were supplemented by 9 distance restraints for the three disulphide bonds between Cys 6 and Cys 14, Cys 16 and Cys 28, and Cys 22 and Cys 39 (Dodt *et al.*, 1985). [Note that for each disulphide bridge the distances S_i-S_j, and $S_i-C_j^\beta$ and $S_j-C_i^\beta$ are restrained to values of 2.02 ± 0.05 Å, 2.99 ± 0.05 Å and 2.99 ± 0.05 Å, respectively.]

Calculational strategy

The calculational strategy used to generate a set of three-dimensional structures satisfying the experimental restraints followed that used previously by Clore *et al.* (1986b,c) for α1-purothionin and phoratoxin. Namely, a set of DG structures were generated using the metric matrix distance geometry program DISGEO (Havel, 1986) and then subjected to refinement using a combination of restrained energy minimization and

Table I. Atomic r.m.s. distributions and shifts

	Atomic r.m.s. difference (Å)							
	All residues		Residues 3–30, 37–46, and 56–57		Residues 30–37		Residues 46–56	
	Backbone atoms	All atoms	Backbone atoms	All atoms	Backbone atoms	All atoms	Backbone atoms	All atoms
(A) R.m.s. distributions								
<DG> vs $\overline{DG}$	1.7 ± 0.2	2.4 ± 0.3	1.6 ± 0.2	2.3 ± 0.3	0.9 ± 0.2	2.0 ± 0.3	1.3 ± 0.2	2.1 ± 0.3
<DGm> vs $\overline{DGm}$	1.8 ± 0.2	2.4 ± 0.2	1.7 ± 0.3	2.3 ± 0.2	1.0 ± 0.2	2.0 ± 0.3	1.4 ± 0.2	2.1 ± 0.3
<RDDG> vs $\overline{RDDG}$	2.6 ± 0.5	3.0 ± 0.5	1.9 ± 0.2	2.4 ± 0.2	1.3 ± 0.2	2.1 ± 0.5	1.9 ± 0.1	2.7 ± 0.3
(B) R.m.s. shifts								
<DG> vs <DGm>	1.1 ± 0.1	1.3 ± 0.1	1.1 ± 0.2	1.2 ± 0.1	0.8 ± 0.2	1.1 ± 0.1	0.9 ± 0.2	1.1 ± 0.2
<DGm> vs <RDDG>	2.4 ± 0.8	2.7 ± 0.8	1.4 ± 0.5	1.7 ± 0.5	1.2 ± 0.2	1.8 ± 0.4	1.8 ± 0.4	2.7 ± 0.5
<DG> vs <RDDG>	2.8 ± 0.8	3.1 ± 0.8	1.8 ± 0.4	2.1 ± 0.4	1.3 ± 0.1	2.0 ± 0.4	2.0 ± 0.5	2.5 ± 0.4
$\overline{DG}$ vs $\overline{DGm}$	0.7	0.7	0.6	0.7	0.5	0.6	0.6	0.6
$\overline{DGm}$ vs $\overline{RDDG}$	1.2	1.4	0.6	0.7	0.8	0.9	0.9	1.2
$\overline{DG}$ vs $\overline{RDDG}$	1.6	1.8	1.0	1.1	0.9	1.1	1.2	1.4
$(\overline{DG})$m vs $\overline{DG}$	1.3	1.6	1.4	1.6	0.8	1.4	1.1	1.4
$(\overline{DGm})$m vs $\overline{DGm}$	1.2	1.5	1.3	1.5	0.9	1.5	1.1	1.3
$(\overline{RDDG})$m vs $\overline{RDDG}$	1.7	1.9	1.6	1.8	1.0	1.4	1.2	1.7
$(\overline{DG})$m vs $(\overline{DGm})$m	1.4	1.7	1.4	1.6	1.0	1.6	1.2	1.7
$(\overline{DGm})$m vs $(\overline{RDDG})$m	1.6	1.9	1.3	1.5	0.9	1.5	1.4	1.8
$(\overline{DG})$m vs $(\overline{RDDG})$m	2.1	2.4	1.8	2.0	1.2	1.7	1.5	2.0
(C) Atomic r.m.s. standard errors								
$\overline{DG}$	0.6	0.9	0.6	0.9	0.3	0.8	0.5	0.8
$\overline{DGm}$	0.7	0.9	0.6	0.9	0.4	0.8	0.5	0.8
$\overline{RDDG}$	1.0	1.1	0.7	0.9	0.5	0.8	0.7	1.0

The notation of the structures is as follow: <DG> comprise the seven converged distance geometry structures, <DGm> the structures derived from the DG structures by restrained energy minimization, and <RDDG> the structures derived from the DGm structures by restrained molecular dynamics (see text). $\overline{DG}$, $\overline{DGm}$ and $\overline{RDDG}$ are the mean structures obtained by averaging the coordinates of the DG, DGm and RDDG structures, respectively. The standard atomic r.m.s. error of these mean structures is given by $\sim$rmsd/$\sqrt{n}$ where rmsd is the average atomic r.m.s. difference between the n structures and the average structure. $(\overline{DG})$m, $(\overline{DGm})$m and $(\overline{RDDG})$m are the structures obtained by restrained energy minimization of the mean $\overline{DG}$, $\overline{DGm}$ and $\overline{RDDG}$ structures, respectively.

Table II. Interproton distance deviations and radii of gyration

Structure	R.m.s. difference between calculated and target interproton distance restraints (Å)				Radii of gyration (Å)
	All (359)	Inter-residue		Intra-residue (157)	
		Short range ($\|i-j\| \leq 5$) (139)	Long range ($\|i-j\| > 5$) (63)		
<DG>	0.70 ± 0.06	0.67 ± 0.07	1.21 ± 0.15	0.36 ± 0.03	11.16 ± 0.15
<DGm>	0.18 ± 0.02	0.18 ± 0.02	0.22 ± 0.04	0.15 ± 0.02	10.91 ± 0.18
<RDDG>	0.16 ± 0.01	0.16 ± 0.02	0.21 ± 0.04	0.14 ± 0.02	10.38 ± 0.20
$\overline{DG}$	0.50	0.52	0.80	0.28	10.90
$\overline{DGm}$	0.27	0.28	0.29	0.23	10.63
$\overline{RDDG}$	0.22	0.23	0.22	0.21	9.91
$(\overline{DG})$m	0.19	0.21	0.24	0.15	11.17
$(\overline{DGm})$m	0.17	0.18	0.20	0.14	11.04
$(\overline{RDDG})$m	0.16	0.17	0.17	0.15	10.65

The notation of the structures is the same as that in Table I. The r.m.s. difference (rmsd) between the calculated (r_{ij}) and target restraints is calculated with respect to the upper (r_{ij}^u) and lower (r_{ij}^l) limits such that

$$\text{rmsd} = \begin{cases} [\Sigma(r_{ij} - r_{ij}^u)^2/n]^{1/2} & \text{if } r_{ij} > r_{ij}^u \\ 0 & \text{if } r_{ij}^l \leq r_{ij} \leq r_{ij}^u \\ [\Sigma(r_{ij} - r_{ij}^{cl})^2/n]^{1/2} & \text{if } r_{ij} < r_{ij}^l \end{cases}$$

Table III. Energies of the structures

Structure	Energy (kcal/mol)									
	Total	Bond (786)	Angle (1414)	Dihedral (364)	Improper (173)	Van der Waals	Electro-static	H-bond	NOE restraints (359)	ϕ torsion angle restraints (10)
<DG>	5507 ± 779	130 ± 8	722 ± 151	406 ± 21	0.02 ± 0.01	375 ± 183	−1 ± 31	−5 ± 2	3548 ± 626	333 ± 257
<DGm>	421 ± 154	46 ± 7	461 ± 47	313 ± 24	33 ± 2	−38 ± 27	−605 ± 37	−23 ± 6	234 ± 51	2 ± 2
<RDDG>	−90 ± 96	38 ± 6	410 ± 28	280 ± 30	30 ± 2	−97 ± 30	−884 ± 43	−50 ± 5	182 ± 26	1 ± 1
$\overline{DG}$	$>10^6$	34 000	8171	580	0.8	$>10^6$	1466	2	1812	1707
$\overline{DGm}$	$>10^6$	35 000	7914	684	23	$>10^6$	233	−2	504	118
$\overline{RDDG}$	$>10^6$	40 000	7417	641	38	$>10^6$	−853	−6	340	0.0
($\overline{DG}$)m	2502	1337	562	316	40	538	−544	−18	263	8
($\overline{DGm}$)m	2123	1190	539	325	43	432	−601	−19	210	3
($\overline{RDDG}$)m	184	45	428	303	37	−91	−701	−37	200	0.2

The notation of the structures is the same as that in Table I. The number of terms for the bond, angle, dihedral and improper dihedral potentials and for the effective NOE interproton distance and ϕ backbone torsion angle restraints potentials are given in parentheses. The effective restraints potentials are represented by a square well potential (see Clore *et al.*, 1986b) with restraints force constants of 20 kcal/mol/Å^2 and 20 kcal/mol/rad^2 for the NOE distance and ϕ torsion angle restraints, respectively.

restrained molecular dynamics with the program CHARMM (Brooks *et al.*, 1983) in which the NOE interproton distance and ϕ backbone torsion angle restraints were incorporated into the total energy function of the system in the form of effective potentials (Kaptein *et al.*, 1985; Clore *et al.*, 1985, 1986a,b,c; Brünger *et al.*, 1986). The empirical energy function consisted of bond, angle, torsion, planarity and non-bonding (i.e. van der Waals, electrostatic and hydrogen bonding) potentials (Karplus and McCammon, 1983; Brooks *et al.*, 1983). The calculations were restricted to residues 1−57 as no intermediate-range ($|i-j| \geq 2$) or long-range ($|i-j| > 5$) NOEs involving residues 58−65 could be detected, the only observable NOEs in this region being of the sequential type ($|i-j| = 1$). Thus, although the local structure of residues 58−65, as deduced from the sequential NOEs, is that of an irregular extended strand (Sukumaran *et al.*, 1986), the location of residues 58−65 with respect to the rest of the protein cannot be ascertained on the basis of the present data.

In the structure determination stage all restraints were explicitly included in the calculations and distances involving methyl and methylene protons were corrected for the pseudo-atom representation used by DISGEO as described by Wüthrich *et al.* (1983). In the refinement stage, distances involving methyl and methylene protons were referred to a single $(<r^{-6}>)^{-1/6}$ average distance (Clore *et al.*, 1986a) and the form of the effective restraints potential was a square well (Clore *et al.*, 1986b,c) with force constants of 20 kcal/mol/Å^2 for the NOE restraints and 20 kcal/mol/rad^2 for the ϕ dihedral angle restraints. Further the disulphide bridge distance restraints were no longer included explicitly as they were implicitly contained in the bond, angle and dihedral potentials of the total empirical energy function.

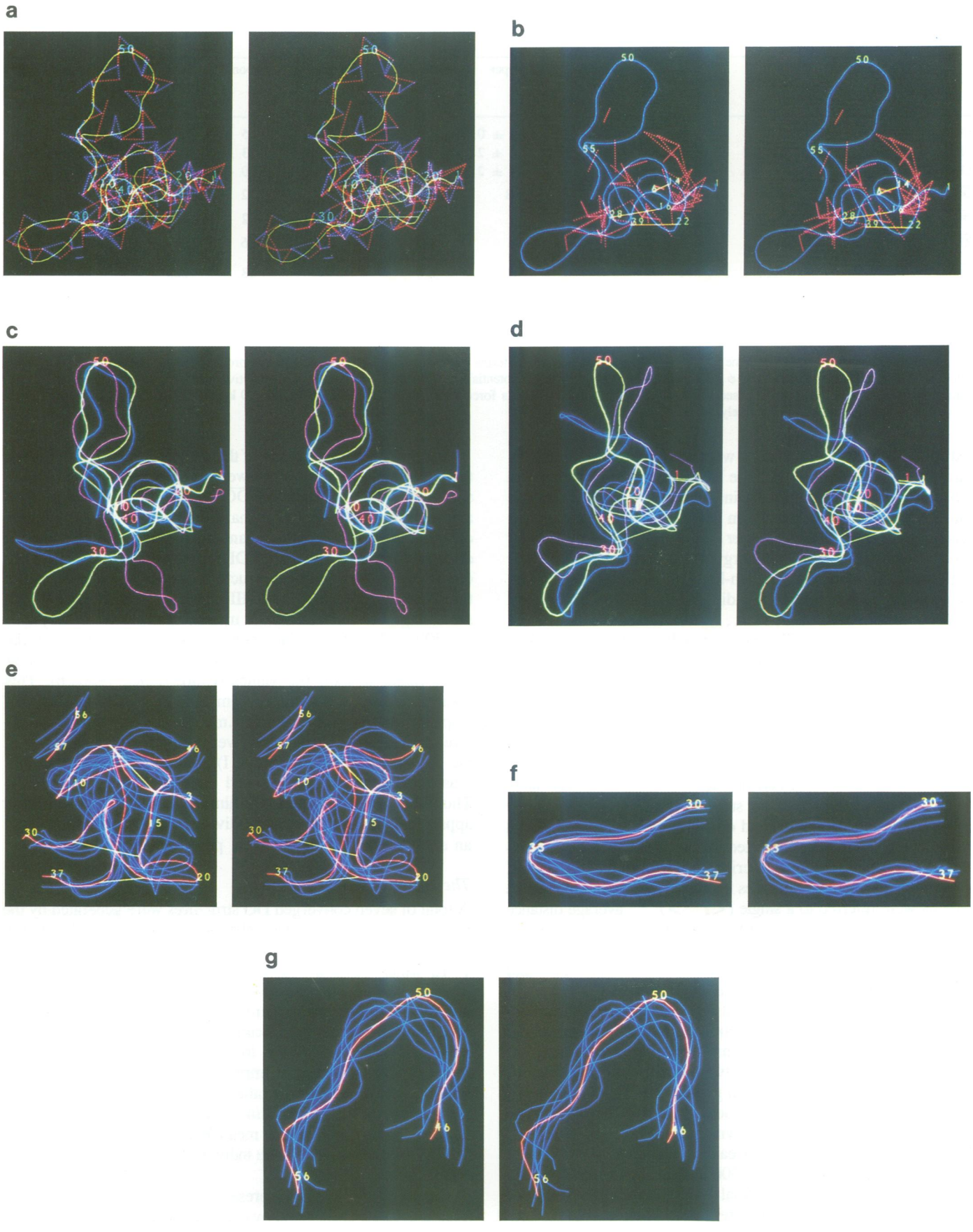
The refinement proceeded in two phases: (i) 500 cycles of restrained energy minimization to generate the DGm structures; and (ii) 1 ps of equilibration (Brooks *et al.*, 1983) followed by 12 ps of restrained molecular dynamics at 380 ± 20 K. The coordinates for the last 8 ps of each dynamics trajectory were averaged and then subjected to 500 cycles of restrained energy minimization to generate the final RDDG structures. This last minimization step is essentially a regularization procedure to correct for minor distortions in covalent structure produced by the averaging procedure and results in atomic r.m.s. shifts of <0.2 Å for all atoms.

In addition, the coordinates of the DG, DGm and RDDG structures best fitted to each other were averaged to yield the mean structures $\overline{DG}$, $\overline{DGm}$ and $\overline{RDDG}$ (Clore *et al.*, 1986a,b). The average structures themselves have no physical significance except that they represent the mean structures about which the individual DG, DGm and RDDG structures are randomly distributed. Indeed the mean structures are very poor with respect to all energy terms (see Table III) and are stereochemically bad structures. For this reason the mean structures were subjected to 1500 cycles of restrained energy minimization in which the van der Waals radii were slowly increased from a quarter of their usual values to their full values (Clore *et al.*, 1986a,b). This resulted in the structures ($\overline{DG}$)m, ($\overline{DGm}$)m and ($\overline{RDDG}$)m, respectively. In terms of atomic r.m.s. displacements these structures are closer to their respective mean structures than any of the individual structures (Table I). At the same time they are reasonable in sterochemical and energetic terms (Table III). Thus, the restrained energy minimized structure ($\overline{RDDG}$)m is approximately as good as the individual RDDG structures from an energetic and stereochemical point of view (Table III).

The converged structures

A total of seven converged DG structures were generated by the metric matrix distance geometry calculations and subjected to refinement. The course of the refinement is summarized in Tables I−III which detail the atomic r.m.s. differences and distributions, the interproton distance deviations and radii of gyration, and the energies of the structures. Note that the energies of the RDDG structures are significantly lower than those of the DGm structures which are in turn lower than those of the DG structures, the major source of improvement coming from the non-bonding terms and the restraints energies (Table III). Further, the energy of the ($\overline{RDDG}$)m structure derived by restrained energy minimization of the mean $\overline{RDDG}$ structure is approximately as good as those of the individual RDDG structures (Table III).

For the purpose of the present structure determination, the results are best understood by considering the protein to be made up of three domains: (i) the central core comprising residues 3−30, 37−46 and 56−57 which is tightly folded and held together by the three disulphide bridges; (ii) an exposed 'finger'


a
b
c
d
e
f
g

like segment of the 27−40 antiparallel β-sheet comprising residues 31−36; and (iii) an exposed loop comprising residues 47−55. Inspection of Figure 1a and b clearly shows that there are numerous interproton distance restraints within each domain but no long range ($|i-j| > 5$) restraints between the domains (Figure 1b). Given that all cross-peaks in the NOESY spectra were assigned unambiguously, the absence of NOEs indicates that there are no long-range interdomain proton−proton distances less than 5 Å present. Consequently, the orientation of the two minor domains with respect to the central core cannot be determined. This is easily appreciated from an examination of Figure 1c and d which show the superposition of the RDDG structures showing the largest displacements of the 31−36 and 47−55 domains, respectively, on the (RDDG)m structure, best fitted to the central core. These features are also clearly seen in the plots of atomic r.m.s. distributions of the RDDG structures about the mean RDDG structure, best fitted to either all residues or to the central domain residues only (Figure 2).

Interestingly, the size of the conformational space sampled by the two minor domains in the DG and DGm structures is much smaller than that sampled by the RDDG structures (see Table I). This is due to the fact that when no distances are present to constrain the spatial orientation of one domain with respect to another, the metric matrix distance geometry algorithm systematically places the residues of the two domains as far apart as possible, such that the two domains lie approximately in a single plane. It is this same feature which is responsible for the systematic expansion of the DG structures relative to X-ray structures (Havel and Wuthrich, 1985). The effect of restrained energy minimization is simply to take the DG structures into the closest local subminimum energy region. This is achieved by only small atomic r.m.s. shifts of the order of ~1 Å (Table I). Consequently, the atomic r.m.s. distributions of the DGm structures about their mean DGm structure are very similar to those of the DG structures. Restrained molecular dynamics, on the other hand, explores a much larger region of conformational space compatible with the experimental restraints and locates the lowest energy subminima within the global minimum energy region (Clore *et al.*, 1986a,b,c). This is manifested by the large region of conformational space sampled by the two minor domains relative to the central domain in the RDDG structures.

The backbone atomic positions of the individual domains are, however, much better determined. Thus, the atomic r.m.s. distribution of the RDDG structures about their mean structure within each domain are relatively small (≤2 Å) and only slightly larger than those for the DG and DGm structures (Table I). This can be appreciated visually from the best fit superpositions of the central domain (Figures 1e), the 30−37 region (Figure 1f) and the 46−56 region (Figure 1g) of the RDDG structures on the (RDDG)m structure, and from the atomic r.m.s. distribution plots for the three domains shown in Figure 2.

That the atomic backbone positions of the RDDG structures are relatively well defined locally can also be seen from the local backbone atomic r.m.s. distributions of tripeptide segments of the RDDG structures about the mean RDDG structure (Figure 3). These have values of ≤1 Å. In terms of the ϕ and ψ backbone torsion angles, however, the variation is somewhat larger (Figure 4) with an average angular r.m.s. difference of 60 ± 20° between the RDDG structures and the mean RDDG structure. This value is comparable to those found in model calculations on bovine pancreatic trypsin inhibitor (Havel and Wuthrich, 1985).

Not surprisingly the atomic positions of most of the sidechains are rather poorly defined. Nevertheless, it is apparent from the plot of local side chain atomic r.m.s. distributions (Figure 3) that the side chains of four segments are relatively well defined, namely residues 13−15, 27−29, 38−42 and 46−48. These regions are principally located within the central core and consequently exhibit a higher degree of order due to packing requirements.

Structural features

The structural features of hirudin are represented as two stereoviews of the structure (RDDG)m in the form of a ribbon drawing (Figure 5). The central core is held together by three disulphide bridges between Cys 6 and Cys 14, Cys 16 and Cys 28, and Cys 22 and Cys 39. The first six residues form an irregular strand leading into a short loop closed off by the first disulphide bridge at its base. This loop is followed by a short double-stranded antiparallel β-sheet (residues 15−22) connected by a β-turn (residues 17−20). The first strand of the β-sheet has a β-bulge at residue 16. As the size of this first antiparallel β-sheet is so small, the conformational space that it can sample is rather limited. Furthermore, its location is stabilized by hydrophobic interactions between Tyr 3 and Thr 4 on the one hand and Leu 13, Leu 15 and Val 21 on the other. The second strand of the first β-sheet leads into the first strand of the second double-stranded antiparallel β-sheet (residues 27−40) via a tight turn at residues 24−27. Residues 31−36 of the second sheet form an exposed 'finger' which constitutes one of the minor domains. The turn at residues 32−35 is a classical type II turn with a Gly at position 3 in the turn. The two disulphide bridges between Cys 16 and Cys 28 and between Cys 22 and Cys 39 act

Fig. 1. Smoothed backbone (N, C^{α}, C) atom representation of the final restrained molecular dynamics structures of hirudin. (**a**) Superposition of the short-range ($|i-j| \leq 5$) inter-residue (red) and intra-residue (blue) interproton distances on (RDDG)m (yellow). (**b**) Superposition of the long-range ($|i-j| > 5$) inter-residue distances (red) on (RDDG)m (blue) with the disulphide bridges shown in yellow. (**c**) Superposition of the two individual RDDG structures (blue and mauve) showing the largest displacement of the minor domain comprising residues 31−36 on (RDDG)m (green) best fitted to the central core. (**d**) Superposition of the two individual RDDG structures (blue and mauve) showing the largest displacement of the loop domain comprising residues 47−55 on (RDDG)m (green) best fitted to the central core. (**e**) Best fit superposition of the central core (3−30, 37−46, 56−57) of the individual RDDG structures (blue) on (RDDG)m (red) with the disulphide bridges of (RDDG)m in yellow. (**f**) Best fit superposition of the 30−37 region of the individual RDDG structures (blue) on (RDDG)m (red). (**g**) Best fit superposition of the 46−56 region of the individual RDDG structures (blue) on (RDDG)m (red). The use of the restrained energy minimized structure (RDDG)m as a reference structure in Figure 1a−d does not imply in any way that the orientation of the 31−36 and 47−55 regions have been determined relative to the central core. Rather, (RDDG)m is simply used as it represents the structure closest to the mean structure RDDG about which the individual RDDG structures are randomly distributed and at the same time is approximately as good in energetic and stereochemical terms as the individual RDDG structures (see Table III). The superpositions of the experimental distances on (RDDG)m shown in Figure 1a and b show clearly, in a way that could not be conveyed by a linear or square matrix diagram of distances versus residue number, that although the local structure of the two minor domains (regions 31−36 and 47−55) can be determined (Figure 1f and g) by the short range ($|i-j| \leq 5$) interproton distances (Figure 1a), the orientation of the minor domains relative to the major domain cannot be determined (Figure 1c and d) as there are no long-range ($|i-j| > 5$) interproton distances between the domains (Figure 1b). (See Materials and methods section for a description of the way in which the backbone atoms were smoothed to give the continuous space curve representation shown here.)

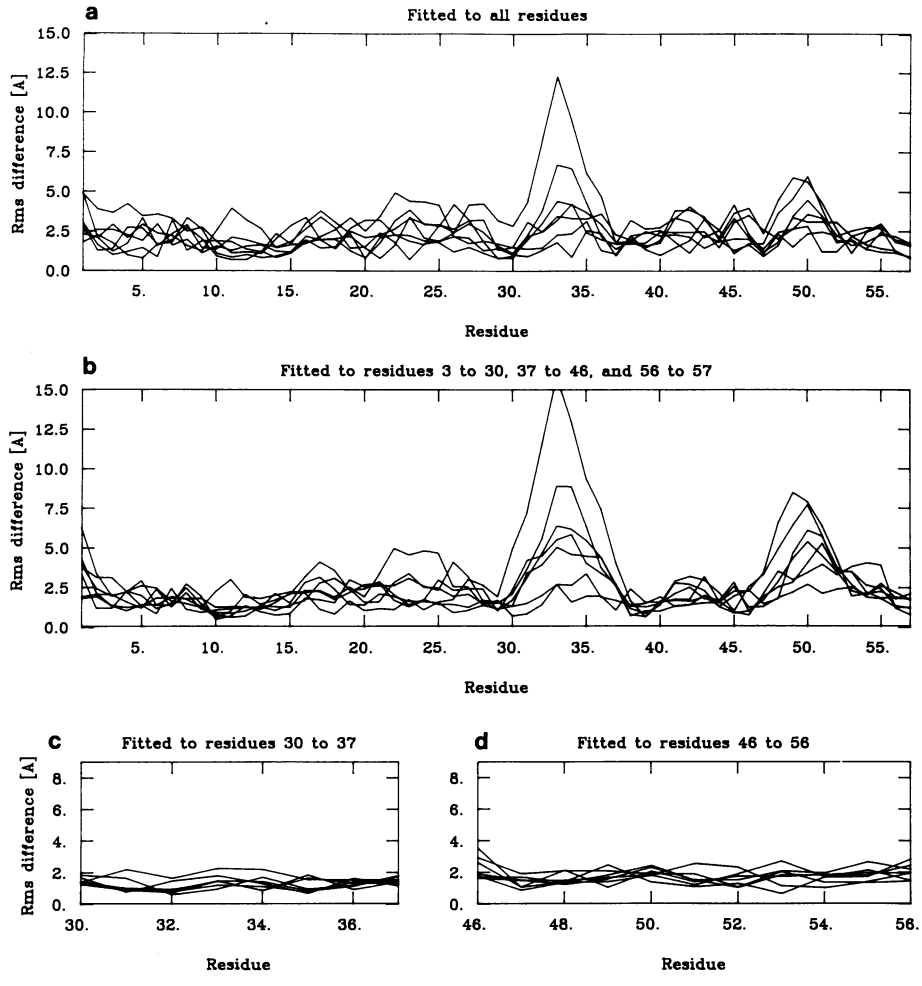


Fig. 2. Atomic r.m.s. distributions of the backbone atoms (N, C^{α}, C, O) atoms of the seven RDDG structures about the mean structure $\overline{RDDG}$ best fitted to (a) all residues, (b) the central core (residues 3–30, 37–46, 56–57), (c) the 30–37 region comprising the 31–36 minor domain, and (d) the 46–56 region comprising the 47–55 minor domain.

as clamps fixing the ends of the two β-sheets to the central core. Interestingly, the first residue of the 'finger' is a Gly at position 31 which may act as a hinge between this domain and the centre core. All other residues of the 'finger' (with the exception of Gly 34 in the turn) are hydrophilic in character. Residues 41–47 form the connection across the central core to the second minor domain consisting of the exposed loop comprising residues 47–55. Although probably flexible, the local conformation of this loop is partially restricted by the two prolines at positions 46 and 48 and stabilized by electrostatic interactions involving the $N^{\epsilon}NH_3$ of Lys 47 and the backbone carbonyl oxygen atoms of residues 52 and 53. As in the case of the 'finger', two glycines at positions 42 and 54 are found near the beginning and end of this loop, allowing hinge-like movements around them. The last two residues of the structure, Phe 56 and Glu 57, comprise part of the central core and their positions are fixed by hydrophobic and electrostatic interactions, respectively. In particular, Phe 56 is in close contact with Asn 12 and Glu 57 with Ser 9.

The outer surface of the two minor domains and a small loop (residues 7–13) in the central core form a contiguous surface of hydrophilic and charged residues. Further, the C-terminal residues 58–65 which form an irregular strand (Sukumaran *et al.*, 1986) also contain a number of acidic residues which could easily lie on this same surface. It is therefore tempting to speculate that the segmental motion of the two minor domains relative to the central core suggested in this study is of functional significance and that the two minor domains act as tentacles embracing the hirudin binding surface on α-thrombin. This proposal is consistent with kinetic and equilibrium measurements showing that a single hirudin molecule occupies two or more sites on α-thrombin simultaneously, one of which is the catalytic site, and that binding sites distinct from the catalytic site are crucial in the binding of hirudin to α-thrombin (Fenton *et al.*, 1979; Fenton, 1981; Walsmann and Markwardt, 1981; Chang, 1983; Stone and Hofsteenge, 1986). Finally it is likely that the intrinsic higher degree of internal flexibility of the two exposed minor domains coupled with interdomain motion is responsible for the failure of hirudin to crystallize.

Materials and methods

Samples for n.m.r. spectroscopy contained 8 mM hirudin (purified from the whole body of leeches as described by Bagdy *et al.*, 1976) in either 90% H_2O/10% D_2O or 99.96% D_2O, pH 3.0. Hirudin was a gift from Dr R.Maschler and Prof. E.Fink (Plantorgan Werk) and Dr D.Tripier (Hoechst). NOESY spectra (Jeener *et al.*, 1979; Macura *et al.*, 1981) were recorded in the pure-phase absorption mode (Marion and Wuthrich, 1983) using the experimental conditions reported previously (Sukumaran *et al.*, 1986).

Metric matrix distance geometry calculations were carried out using the program DISGEO (Havel and Wuthrich, 1984; Havel, 1986). All energy minimization and restrained molecular dynamics calculations were carried out as described previously (Clore *et al.*, 1986a,b; Brunger *et al.*, 1986) on a CRAY-XMP using a CRAY version (A.T.Brunger, unpublished data) of the program CHARMM

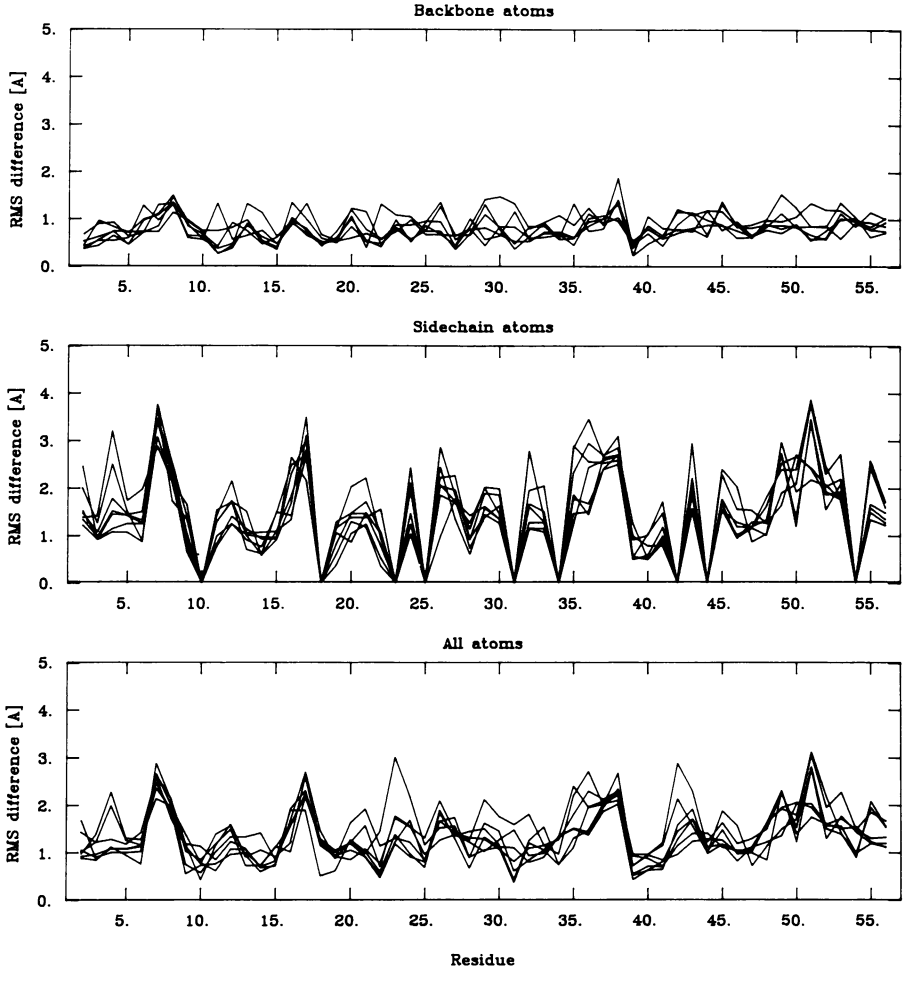


Fig. 3. Local atomic r.m.s. distributions of the seven RDDG structures about the mean structure $\overline{\text{RDDG}}$ for the backbone atoms, sidechain atoms and all atoms. The values plotted represent best fit atomic r.m.s. differences between tripeptide segments along the chain as a function of the sequence number of the middle residue.

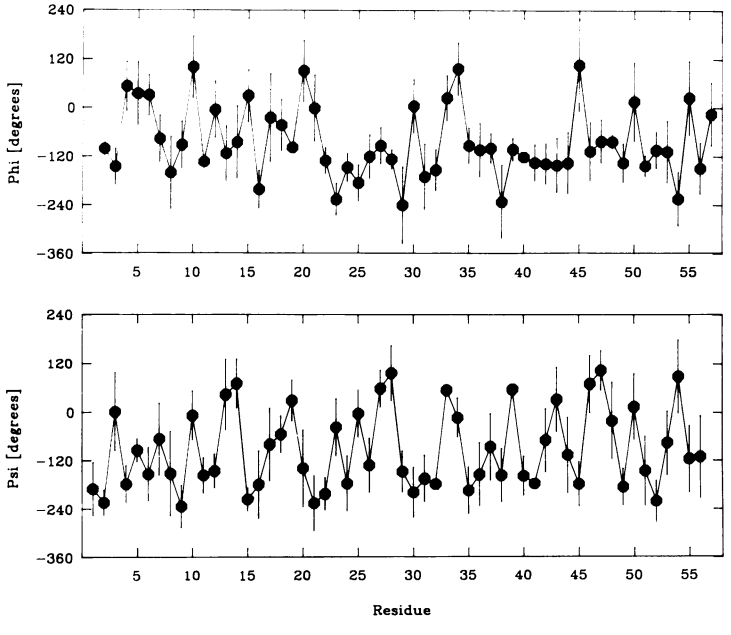


Fig. 4. Angular r.m.s. distributions of the ϕ and ψ backbone torsion angles of the RDDG structures. The filled in circles (●) are the values of the ϕ and ψ angles of the mean $\overline{\text{RDDG}}$ structure and the bars represent the average angular r.m.s deviations of the RDDG structures about the mean $\overline{\text{RDDG}}$ structure.

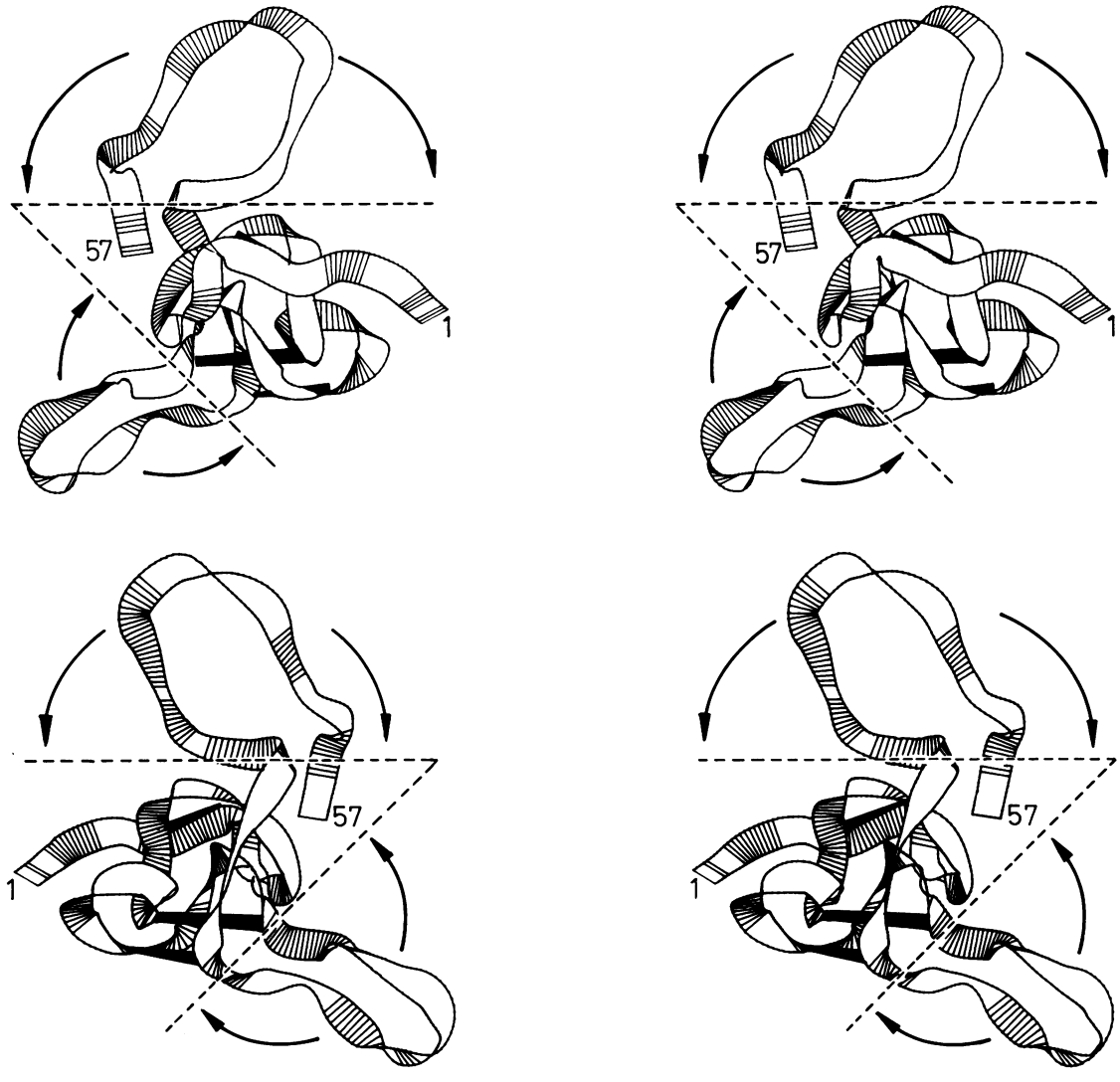


Fig. 5. Two stereoviews of the backbone atoms of the structure (RDDG)m represented as ribbon drawings illustrating the three domains of hirudin. The orientation of the two minor domains relative to the central core cannot be determined as no long range ($|i-j| > 5$) interdomain NOEs could be observed (cf. Figure 1b). As a result the individual RDDG structures exhibit large displacements in the positions of the two minor domains relative to the central one (cf. Figure 1c and d) such that they can occupy any position indicated by the arrows within the boundaries defined by the dashed lines. These pictures were produced by a computer program written by Lesk and Hardman (1982).

(Brooks *et al.*, 1983). Analysis of the structures and the molecular dynamics trajectories was carried out using a modified version of the function network of FRODO (Jones, 1978) interfaced with CHARMM on an Evens and Sutherland PS330 colour graphics system. The smooth backbone (N, C^{α}, C) atom representations shown in Figure 1 were obtained by subjecting the coordinates of the structures to 200 cycles of constrained energy minimization (Bruccoleri and Karplus, 1986) in which all backbone bond angles and bond lengths were harmonically driven to 180° and to one-third of the distance between adjacent C^{α} atoms, respectively, and the C^{α} were constrained to their original positions by weak harmonic constraints. The N, C^{α} and C atoms of the resulting peptide backbone are smoothed out to form a continuous space curve. This representation owes its origin to Feldmann *et al.* (1986).

Acknowledgements

This work was supported by the Max-Planck Gesellschaft and Grant No. Cl 86/1-1 from the Deutsche Forschungsgemeinschaft (GMC and AMG). We thank Dr R.Maschler and Prof. E.Fink (Plantorgan Werk AG) and Dr D.Tripier (Hoechst, FRG) for the gift of hirudin, the Max-Planck Institut fur Plasma Physik (Garching) for CRAY-XMP computing facilities, and Drs R.J.Feldman and B.R.Brooks (NIH) for suggesting the use of the smoothed peptide backbone representation.

References

Bagdy,D., Barabas,E., Graf,L., Ellebaek,T. and Magnusson,S. (1976) *Methods Enzymol.*, **45**, 669–678.

Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) *J. Comput. Chem.*, **4**, 187–217.

Brown,L.R. (1984) *J. Magn. Resonance*, **57**, 513–518.

Bruccoleri,R.E. and Karplus,M. (1986) *J. Comput. Chem.*, **7**, 165–175.

Brunger,A.T., Clore,G.M., Gronenborn,A.M. and Karplus,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3801–3805.

Chang,J.Y. (1983) *FEBS Lett.*, **164**, 307–313.

Clore,G.M., Gronenborn,A.M., Brunger,A.T. and Karplus,M. (1985) *J. Mol. Biol.*, **186**, 435–455.

Clore,G.M., Brunger,A.T., Karplus,M. and Gronenborn,A.M. (1986a) *J. Mol. Biol.*, **191**, 523–551.

Clore,G.M., Nilges,M., Sukumaran,D.K., Brunger,A.T., Karplus,M. and Gronenborn,A.M. (1986b) *EMBO J.*, **5**, 2729–2735.

Clore,G.M., Sukumaran,D.K., Nilges,M. and Gronenborn,A.M. (1986c) *Biochemistry*, in press.

Crippen,G.M. and Havel,T.F. (1978) *Acta Crystallogr.*, **A34**, 282–284.

Dodt,J., Muller,H.P., Seemuller,U. and Chang,J.Y. (1984) *FEBS Lett.*, **165**, 180–184.
Dodt,J., Seemuller,U., Maschler,R. and Fritz,H. (1985) *Biol. Chem. Hoppe Seyler*, **366**, 379–385.
Feldmann,R.J., Brooks,B.R. and Lee,B. (1986) *Tools for Each Age: Understanding Protein Architecture Through Simulated Unfolding*. Division of Computer Research and Technology, National Institutes of Health, Bethesda.
Fenton,J.W (1981) *Ann. N.Y. Acad. Sci.*, **370**, 468–495.
Fenton,J.W., Landis,B.H., Walz,D.A., Bing,D.H., Feinamn,R.D., Zabinsky,M.P., Sonder,S.A., Berliner,L.J. and Finlayson,J.S. (1979) In Bing,D.H. (ed.), *The Chemistry and Physiology of Human Plasma Proteins*. Pergamon, NY, pp. 151–185.
Harvey,R.P., Degryse,E., Stefani,L., Schanber,F., Cazenave,J.P., Courtney,M., Tolstoshev,P. and Lecocq,J.P. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1084–1088.
Havel,T.F. (1986) DISGEO, Quantum Chemistry Program Exchange Program No. 507, Indiana University.
Havel,T.F., Kuntz,I.D. and Crippen,G.M. (1983) *Bull. Math. Biol.*, **45**, 665–720.
Havel,T.F. and Wuthrich,K. (1984) *Bull. Math. Biol.*, **46**, 673–698.
Havel,T.F. and Wuthrich,K. (1985) *J. Mol. Biol.*, **182**, 281–294.
Haycraft,J.B. (1884) *Proc. Roy. Soc. London Ser. B*, **36**, 478–487.
Ishikaw,A., Hafter,R., Seemuller,U., Gohel,J.M. and Graef,M. (1980) *Thromb. Res.*, **19**, 351–358.
Jeener,J., Meier,B.H., Bachmann,P. and Ernst,R.R. (1979) *J. Chem. Phys.*, **71**, 4546–4553.
Jones,T.A. (1978) *J. Appl. Crystallogr.*, **11**, 268–272.
Kaptein,R., Zuiderweg,E.R.P., Scheek,R.M., Boelens,R. and van Gunsteren,W.F. (1985) *J. Mol. Biol.*, **182**, 179–182.
Karplus,M. and McCammon,J. (1983) *Annu. Rev. Biochem.*, **53**, 263–300.
Kline,A.D., Braun,W. and Wuthrich,K. (1986) *J. Mol. Biol.*, **189**, 377–382.
Kloss,T. and Mittman,U. (1982) *Longenbech's Arch. Chir.*, **358**, 548.
Lent,C. (1986) *Nature*, **323**, 494.
Lesk,A.M. and Hardman,K.D. (1982) *Science*, **216**, 539–540.
Macura,S., Huang,Y., Suter,D. and Ernst,R.R. (1981) *J. Magn. Resonance*, **43**, 259–281.
Magnusson,S. (1972) In Boyer,P.D. (ed.), *The Enzymes*. Academic Press, NY, 3rd edn., Vol. 3, pp. 227–321.
Marion,D. and Wuthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.
Markwardt,F. (1970) *Methods Enzymol.*, **19**, 924–932.
Markwardt,F. (1985) *Biomed. Biochim. Acta*, **44**, 1007–1013.
Markwardt,F., Hauptmann,J., Nowak,G., Klessen,C. and Walsmann,P. (1982) *Tromb. Haemostasis*, **47**, 226–229.
Markwardt,F., Nowak,G., Sturzebecher,J., Griessbach,U. and Vogel,G. (1984) *Haemostasis*, **53**, 2521–2524.
Nilsson,L., Clore,G.M., Gronenborn,A.M., Brunger,A.T. and Karplus,M. (1986) *J. Mol. Biol.*, **188**, 455–475.
Nowak,G. and Markwardt,F. (1980) *Exp. Pathol.*, **18**, 438–443.
Pardi,A., Billeter,M. and Wuthrich,K. (1984) *J. Mol. Biol.*, **180**, 741–751.
Rance,M., Sorensen,O.W., Bodenhausen,G., Wagner,G., Ernst,R.R. and Wuthrich,K. (1984) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.
Stone,S.R. and Hosteenge,J. (1986) *Biochemistry*, **25**, 4622–4628.
Sukumaran,D.K., Clore,G.M., Preuß,A., Zarbock,J. and Gronenborn,A.M. (1986) *Biochemistry*, in press.
Walsmann,P. and Markwardt,F. (1981) *Pharmazie*, **36**, 653-660.
Williamson,M.P., Havel,T.F. and Wuthrich,K. (1985) *J. Mol. Biol.*, **182**, 295–315.
Wuthrich,K., Billeter,M. and Braun,W. (1983) *J. Mol. Biol.*, **160**, 949–961.

Received on 6 November, 1986; revised on 15 December, 1986