Volume 243, number 2, 223–233 FEB 06669 January 1989

The influence of stereospecific assignments on the determination of three-dimensional structures of proteins by nuclear magnetic resonance spectroscopy

Application to the sea anemone protein BDS-I

Paul C. Driscoll, Angela M. Gronenborn and G. Marius Clore

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Received 10 November 1988

The influence of the stereospecific assignments of β-methylene protons and the classification of χ_1 torsion angles on the definition of the three-dimensional structures of proteins determined from NMR data is investigated using the sea anemone protein BDS-I (43 residues) as a model system. Two sets of structures are computed. The first set comprises 42 converged structures (denoted STEREO structures) calculated on the basis of the complete list of restraints derived from the NMR data, consisting of 489 interproton and 24 hydrogen bonding distance restraints, supplemented by 23 φ backbone and 21 χ_1 side chain torsion angle restraints. The second set comprises 31 converged structures (denoted NOSTEREO structures) calculated from a reduced data set in which those restraints arising from stereospecific assignments, and the corresponding χ_1 torsion angle restraints, are explicitly omitted. The results show that the inclusion of the stereospecific restraints leads to a significant improvement in the definition of the structure of BDS-I, both with respect to the backbone and the detailed arrangement of the side chains. Average atomic rms differences between the individual structures and the mean structures for the backbone atoms are 0.67 ± 0.12 Å and 0.93 ± 0.16 Å for the STEREO and NOSTEREO structures, respectively; the corresponding values for all atoms are 0.90 ± 0.17 Å and 1.17 ± 0.17 Å, respectively. In addition, while the overall fold remains unchanged, there is a small but significant atomic displacement between the two sets of structures.

Protein structure; NMR; Stereospecific assignment; NOE; Interproton distance; Simulated annealing

1. INTRODUCTION

A detailed knowledge of the three-dimensional structures of biological macromolecules is an essential prerequisite for a complete understanding of their function and of structure-function relationships. The use of ^{1}H-NMR spectroscopy and in particular nuclear Overhauser effect (NOE) experiments is now recognized as a useful method for determining low resolution structures of small (M_r < 10000) proteins in solution [1–3]. For the majority of NMR-derived protein structures to date, it is generally the case that the backbone fold is relatively well defined, with backbone atomic rms distributions about the mean coordinate positions of 1–2 Å, whilst the definition of the side chains is significantly poorer. To obtain a deeper understanding of the function of these molecules it is

Correspondence address: G.M. Clore and A.M. Gronenborn, Laboratory of Chemical Physics, Building 2, Room 123, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; E-COSY, exclusive two-dimensional correlated spectroscopy; HOHAHA, two-dimensional homonuclear Hartmann-Hahn spectroscopy; rms, root mean square; SA, simulated annealing; $d_{XY}(i,j)$, NOE or distance between proton X on residue i and proton Y of residue j

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

necessary to look in more detail at the structural characteristics of the side chains.

In two recent papers we presented the determination of the solution structure of the 43 residue anti-hypertensive and anti-viral protein BDS-I from the sea anemone *Anemonia sulcata*, using NMR spectroscopy and hybrid metric matrix distance geometry-dynamical simulated annealing calculations [4,5]. The restraints used for the calculations were derived in part from NOEs involving stereospecifically assigned β-methylene groups in conjunction with approximate ranges for some χ_1 side chain torsion angles. In this paper we compare these previously reported structures with new ones calculated from a shorter list of restraints that makes use of center averaging for the β-methylene groups with suitably modified NOE restraints, and omits the χ_1 torsion angle restraints. We show that the inclusion of the additional restraints relating to stereospecific assignments of β-methylene protons leads to a significant improvement in the definition of the structure, not only for the side chains but also for the backbone. Indeed, in the absence of the additional stereospecific restraints, it is considerably more difficult to draw conclusions concerning the interactions of functional groups within the protein. Further, the inclusion of the stereospecific restraints leads to a small but significant shift in the overall structure.

2. EXPERIMENTAL

2.1. *Sample preparation*

The isolation of BDS-I from the sea anemone *Anemonia sulcata* and the preparation of samples for NMR spectroscopy were as described in [4].

2.2. *NMR spectroscopy*

All NMR spectra were recorded with a Bruker AM series 600 MHz spectrometer. Accurate measurements of $^3J_{\alpha\beta}$ spin-spin coupling constants were obtained from E-COSY experiments at 27 and 40°C, utilising a 64 step phase-cycling scheme [6]. Digital resolution in the transformed spectrum was 0.78 Hz/point in F_2 and 3.13 Hz/point in F_1. Intraresidue NOE measurements were taken from 50 ms NOESY spectra recorded in H_2O and D_2O. There is no experimental scheme that completely eliminates the effects of zero quantum coherence for all spin systems in the spectrum. Therefore these NOE experiments were recorded under three different sets of conditions: without any variation of the mixing time, with a 5% random variation of the mixing time between t_1 increments, and with a three step systematic variation of the mixing time similar to the method proposed in [7]. In this way it was possible to account for the effects of zero quantum coherence for many of the amino acid residues.

2.3. *Structure calculations*

Structure calculations were carried out using the hybrid metric matrix distance geometry-dynamical simulated annealing approach [8]. The distance geometry part of the calculation was carried out using the program DISGEO [9], while dynamical simulated annealing was carried out with the program XPLOR [10–12], which is derived originally from the program CHARMM [13]. The calculations were performed on MicroVAX II, VAX 8550, VAX 8530 and CONVEX C1-XP computers.

3. RESULTS AND DISCUSSION

3.1. *Stereospecific assignments*

Stereospecific assignments of prochiral β-methylene protons were based on $^3J_{\alpha\beta}$ spin-spin coupling constants and intraresidue $d_{\alpha\beta}(i,i)$ and $d_{N\beta}(i,i)$ NOEs. These NMR parameters have a characteristic pattern for each of the three energetically preferred staggered rotamer positions ($\chi_1 = 60°$, 180° or $-60°$) [14]. A survey of high resolution protein crystal structures indicates that 95% of all side chains lie within $\pm 15°$ of the preferred rotamer positions [15]. Consequently, it should be possible to obtain stereospecific assignments for a large number of the β-methylenes. Limitations in the assignment arise from chemical shift degeneracy of the $C^\beta H$ resonances, and intrinsic mobility of the side chain. The latter will manifest itself by NMR data that are inconsistent with a single preferred rotamer conformation [16]. For BDS-I the former problem was minimized by the presence of two isoforms of the protein in the same sample. The substitution of a Phe residue for a Leu at position 18 in the sequence resulted in a number of resolved spin systems with different chemical shifts for each form of the protein. This allowed the analysis of many spin systems which possessed overlapping $C^\beta H$ resonances in one form of the protein or the other. In all respects the structure of the two forms of the protein was found to be the same within the error limits of the NOE and coupling constant data [4,5]. Indeed, residue 18 is found to be in a completely solvent exposed position where it has little influence on the packing of the main chain or other side chains [5]. For residues which exhibited $^3J_{\alpha\beta}$ coupling constants in the range 5.0–9.0 Hz or

NOEs inconsistent with a preferred rotamer conformation no stereospecific assignment was made. This occurred for three residues which were found in the final structures to have their side chains exposed to the solvent where motional averaging can be envisaged, namely, Ile-17, Phe-18, and Arg-19 [5]. For proline residues, stereospecific assignments of the $C^{\beta}H$ protons were based only on the relative magnitude of the two $d_{\alpha\beta}(i,i)$ NOEs (i.e. a strong $d_{\alpha\beta}(i,i)$ NOE arises from the close proximity of the $C^{\alpha}H$ and $C^{\beta 3}H$ protons in proline with χ_1 close to 0° [17]).

The 43 amino acid protein BDS-I has five prolines and 25 non-proline residues with prochiral β-methylene groups as well as two threonine and two isoleucine residues (each with a β-methine proton) for which χ_1 torsion angle restraints can in principle be obtained from the analysis of intraresidue NMR data. The criteria for obtaining the stereospecific assignment of β-methylene groups on the basis of a semi-quantitative analysis of intraresidue NOEs and coupling constants has been described in [14,18]. Experimentally we were able to determine stereospecific assignments for the β-methylene groups of 17 of the 25 non-proline residues and two of the prolines in BDS-I. Note that the complete set of NOEs and coupling constants is not necessarily required to make the stereospecific assignment or indeed the classification of the χ_1 torsion angle. For example the χ_1 =

Table 1

Intraresidue coupling constant and NOEs pertinent to the stereospecific assignment and classification of χ_1 torsion angles for β-methylene and β-methine containing residues in BDS-I

β-Methylene groups

Residue	Coupling constants[a]		NOEs[b]				χ_1 classification
	$^3J_{\alpha\beta 2}$	$^3J_{\alpha\beta 3}$	$d_{\alpha\beta 2}(i,i)$	$d_{\alpha\beta 3}(i,i)$	$d_{N\beta 2}(i,i)$	$d_{N\beta 3}(i,i)$	
Pro-3	not stereo assigned		ZQC[c]				
Cys-4	<5.0[d]	<5.0[d]	strong	strong	absent	medium	60° ± 60°
Phe-5	11.5	2.7	weak	medium	medium	absent	−60° ± 60°
Cys-6	11.3	3.3	weak	medium	strong	medium	−60° ± 60°
Ser-7	degenerate $C^{\beta}H$ chemical shifts						
Lys-9	11.8	3.8	weak	strong	medium	weak	−60° ± 60°
Pro-10	e	e	medium	strong	–	–	–
Arg-12	overlap in spectrum						
Asp-14	11.6	1.7	weak	strong	strong	absent	−60° ± 60°
Leu-15	3.0	>10.0	strong	weak	strong	strong	180° ± 60°
Trp-16	degenerate $C^{\beta}H$ chemical shifts						
Leu-18	degenerate $C^{\beta}H$ chemical shifts						
Phe-18	5.5 and 9.5 not stereo assigned (J values inconsistent with preferred rotamer)						
Arg-19	6.3 and 6.8 not stereo assigned (J values inconsistent with preferred rotamer)						
Cys-22	<5.0[d]	>10.0[d]	ZQC[c]		strong	strong	180° ± 60°[f]
Pro-23	degenerate $C^{\beta}H$ chemical shifts						
Tyr-26	12.1	2.7	weak	strong	strong	weak	−60° ± 60°
Tyr-28	12.5	3.4	weak	strong	strong	weak	−60° ± 60°
Ser-30	<5.0[d]	<5.0[d]	strong	strong	absent	weak	60° ± 60°
Asn-31	3.6	4.2	strong	strong	absent	weak	60° ± 60°
Cys-32	3.0	3.0	strong	strong	absent	medium	60° ± 60°
Tyr-33	degenerate $C^{\beta}H$ chemical shifts						
Lys-34	overlap in spectrum						
Trp-35	6.3	10.4	medium	absent	strong	strong	180° ± 60°
Pro-36	not stereo assigned		ZQC[c]				
Asn-37	12.5	4.8	ZQC[c]		medium	absent	−60° ± 60°
Cys-39	11.0	3.4	absent	medium	medium	absent	−60° ± 60°
Cys-40	11.4	2.6	weak	strong	medium	absent	−60° ± 60°
Tyr-41	>10.0[d]	<5.0[d]	weak	strong	strong	weak	−60° ± 60°
Pro-42	e	e	weak	medium	–	–	–
His-43	6.3	11.3	medium	weak	strong	strong	180° ± 60°

(continued overleaf)

Table 1 (continued)
β-Methine groups[g]

Residue	Coupling constant[a] $^3J_{\alpha\beta}$	NOEs[b]		χ_1 classification
		$d_{\alpha\beta}(i,i)$	$d_{N\beta}(i,i)$	
Ile-17	7.0	(J inconsistent with preferred rotamer)		
Thr-21	3.7	overlap	strong	120° ± 120°[h]
Thr-29	3.2	strong	absent	60° ± 60°
Ile-39	3.8	strong	overlap	120° ± 120°[h]

[a] Coupling constants were measured from E-COSY spectra recorded in D_2O at 27°C and 40°C with presaturation of the residual HOD solvent resonance. The final digital resolution in F_2 after zero-filling was 0.78 Hz/point

[b] NOEs were taken from one of three 50 ms NOESY experiments as described in section 2

[c] The evaluation of the relative magnitude of the $d_{\alpha\beta}(i,i)$ NOEs was obscured by zero quantum coherence effects in all of the NOESY experiments

[d] Coupling constants not accurately measurable because of slight non-degenerate overlap of cross-peaks for the same residue for each form of the protein. Coupling constants are estimated on the basis of characteristic cross-peak profiles observed for three spin systems in a 55 ms WALTZ-17 HOHAHA experiment [4]

[e] Stereospecific assignments of proline $C^{\beta}H$ resonances made on the basis of the relative magnitude of the $d_{\alpha\beta}(i,i)$ NOEs alone [17]

[f] For residue Cys-22 classification of the χ_1 torsion angle was possible without obtaining stereospecific assignments of the $C^{\beta}H$ protons

[g] For threonine and isoleucine, the β-methine proton has the equivalent position as the $C^{\beta 2}H$ proton in β-methylene containing residues

[h] For Thr-21 and Ile-39 the data are only sufficient to rule out the rotamer in which the β-methine proton is *trans* to the $C^{\alpha}H$ proton

Table 2

Structural statistics[a]

	⟨SA⟩	$\overline{SA}$	$(\overline{SA})r$
Rms deviations from experimental restraints (Å)[b]			
STEREO structures (42) with STEREO restraints			
All (513)	0.085 ± 0.002	0.059	0.079
Interresidue short range ($\|i-j\| \leq 5$) (150)	0.083 ± 0.005	0.053	0.086
Interresidue long range ($\|i-j\| > 5$) (105)	0.111 ± 0.007	0.082	0.100
Intraresidue (234)	0.074 ± 0.006	0.053	0.065
Hbond (24)[c]	0.051 ± 0.011	0.038	0.043
NOSTEREO structures (31) with NOSTEREO restraints			
All (433)	0.062 ± 0.004	0.026	0.065
Interresidue short range ($\|i-j\| \leq 5$) (132)	0.056 ± 0.007	0.005	0.063
Interresidue long range ($\|i-j\| > 5$) (87)	0.094 ± 0.009	0.055	0.104
Intraresidue (190)	0.045 ± 0.007	0.010	0.039
Hbond (24)[c]	0.041 ± 0.014	0.021	0.056
STEREO structures with NOSTEREO restraints			
All (433)	0.065 ± 0.002	0.044	0.064
Interresidue short range ($\|i-j\| \leq 5$) (132)	0.065 ± 0.007	0.043	0.071
Interresidue long range ($\|i-j\| > 5$) (87)	0.095 ± 0.011	0.071	0.094
Intraresidue (190)	0.043 ± 0.005	0.025	0.041
NOSTEREO structures with STEREO restraints[d]			
All (513)	0.288 ± 0.029 [34 ± 5]	0.188 [19]	0.267 [13]
Interresidue short range ($\|i-j\| \leq 5$) (150)	0.326 ± 0.052 [13 ± 3]	0.207 [9]	0.271 [10]
Interresidue long range ($\|i-j\| > 5$) (105)	0.445 ± 0.059 [13 ± 3]	0.313 [9]	0.434 [13]
Intraresidue (234)	0.158 ± 0.016 [8 ± 2]	0.080 [1]	0.157 [8]

(continued overleaf)

Table 2 (continued)

	⟨SA⟩	$\overline{SA}$	($\overline{SA}$)r
STEREO structures			
F_{NOE} (kcal·mol^{-1})[e]	187 ± 11	92	160
F_{tor} (kcal·mol^{-1})[f]	26 ± 6	13	24
F_{repel} (kcal·mol^{-1})[g]	91 ± 7	2417	72
$E_{L\text{-}J}$ (kcal·mol^{-1})[h]	−103 ± 11	$>10^6$	−117
Deviations from idealized geometry[i]			
Bonds (Å) (646)	0.014 ± 0.006	0.329	0.013
Angles (°) (1157)	2.910 ± 0.363	28.373	2.517
Impropers (°) (242)	0.830 ± 0.06	2.629	0.797
NOSTEREO structures			
F_{NOE} (kcal·mol^{-1})[e]	79 ± 8	15	92
F_{tor} (kcal·mol^{-1})[f]	5.5 ± 3.5	0.6	13
F_{repel} (kcal·mol^{-1})[g]	46 ± 6	2906	41
$E_{L\text{-}J}$ (kcal·mol^{-1})[h]	−107 ± 12	$>10^6$	−102
Deviations from idealized geometry[i]			
Bonds (Å) (646)	0.012 ± 0.004	0.367	0.012
Angles (°) (1157)	3.064 ± 0.504	28.429	2.549
Impropers (°) (242)	0.711 ± 0.071	0.725	0.738

[a] The notation of the structures is as follows: ⟨SA⟩ are the 42 STEREO or 31 NOSTEREO dynamical simulated annealing structures; $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the individual SA structures of each set best fitted to each other; ($\overline{SA}$)r is the structure obtained by restrained minimization of $\overline{SA}$

[b] The rms deviations from the experimental restraints are calculated with respect to the upper and lower limits of the distance restraints [19]. None of the structures exhibited violations greater than 0.5 Å. The number of distances in each category is given in parentheses next to the category name

[c] For each backbone hydrogen bond there are two restraints: $r_{NH\text{-}O} < 2.3$ Å and $r_{N\text{-}O} < 3.3$ Å. The lower limits are given by the sum of the van der Waals radii of the relevant atoms. 12 backbone hydrogen bonds within regular elements of secondary structure were identified on the basis of the NOE and NH exchange data [5]

[d] Numbers in square parentheses are the numbers of interproton distance violations of the NOSTEREO structures against the STEREO distance restraints that are greater than 0.5 Å

[e] The total target function, representing the effective potential energy of the system, is made up of covalent, van der Waals, torsion angle and NOE terms: $F_{tot} = F_{covalent} + F_{repel} + F_{tor} + F_{NOE}$. See ref. [8] for a full description. The values of the square-well NOE potential F_{NOE} are calculated with a force constant of 50 kcal·mol^{-1}·Å^{-2}. F_{tor} comprises ϕ and χ_1 torsion angle restraints derived from the NMR data

[f] The values of F_{tor} are calculated with a force constant of 200 kcal·mol^{-1}·rad^{-2}. F_{tor} is a square-well dihedral potential which is used to restrict the ranges of 23 ϕ and 22 χ_1 torsion angles and the ω peptide bond torsion angles of the five proline residues (Pro 36 and 42 being restrained to the *cis* conformation and the others to the *trans*)

[g] The values of the van der Waals repulsion term F_{repel} are calculated with a force constant of 4 kcal·mol^{-1}·Å^{-4} with the hard sphere van der Waals radii set to 0.8 times the standard values used in the CHARMM empirical energy function [13]

[h] $E_{L\text{-}J}$ is the Lennard-Jones van der Waals energy calculated using the CHARMM empirical energy function [13]. This is not used in the structure calculation but only as a check on the quality of the structures

[i] The number of bond, angle and improper terms is given in parentheses. The improper terms serve to maintain planarity and appropriate chirality; they also maintain the peptide bond of all residues (with the exception of prolines) in the *trans* conformation. In the dynamical simulated annealing calculations, the restraints for the disulfide bridges are implicitly included in the bond and angle terms

60° rotamer is uniquely defined by two strong $d_{\alpha\beta}(i,i)$ NOEs or two $^3J_{\alpha\beta}$ coupling constants smaller than 5.0 Hz. One $^3J_{\alpha\beta}$ coupling constant greater than 10.0 Hz automatically implies that the remaining one is smaller than 7.0 Hz, and eliminates the possibility of the $\chi_1 = 60°$ rotamer position. Some assistance in the estimation of the relative sizes of $^3J_{\alpha\beta}$ coupling constants was ob-

tained from qualitative correlations of the $C^{\alpha}H$-$C^{\beta}H$ cross peak patterns observed in the D_2O HOHAHA and E-COSY spectra [4]; further, a small active $^3J_{\alpha\beta}$ coupling constant gives rise to an intrinsically weak E-COSY cross peak due to overlap of antiphase cross peak multiplet components. In the case of β-type secondary structure, it is also possible to distinguish the $\chi_1 = -60°$ and 180° rotamer positions on the basis of sequential $d_{\beta N}(i,i+1)$ NOEs [18]. For the present study of BDS-I the latter was not used since it requires the assumption or some prior knowledge of the local secondary structure based on either a qualitative analysis of the pattern of sequential backbone NOEs and $^3J_{NH\alpha}$ coupling constants or preliminary structure calculations. Table 1 documents all the relevant NMR parameters that were used for the identification of stereospecific assignments and classification of the χ_1 torsion angle ranges for both the β-methylene and β-methine containing residues of BDS-I. When a preferred rotamer position of a side chain was identified the torsion angle restraint used in the calculation of the structures was described by a square well effective potential function [19] with width ± 60° about the rotamer positions, i.e. $\chi_1 = 60 \pm 60°$, $180 \pm 60°$ and $-60 \pm 60°$. This represents a conservative restraint designed to ensure that each χ_1 torsion angle in the calculated structures remains in a range consistent with the experimentally determined rotamer position.

3.2. *Structure calculations*

Two sets of structures of BDS-I were calculated using different sets of input NOE and torsion angle restraints. In all cases where stereospecific assignments of β-methylene groups were not included in the description of NOE restraints, the upper limits of the input interproton distance restraints were corrected for center averaging as described in [20]. This amounts effectively to the addition of 1.0 Å to the upper distance bound for each $C^{\beta}H_2$ group involved. Center averaging of other groups (aromatic $C^{\delta}H$ and $C^{\epsilon}H$ protons, glycine $C^{\alpha}H_2$ protons, leucine $C^{\delta}H_3$ groups, etc.) was also carried out with appropriate distance corrections [20]. The calculation of a set of STEREO structures derived from the full set of 489 interproton distance, 24 hydrogen bonding, and 23 ϕ backbone and 21 χ_1 side chain torsion angle restraints is described fully in [5]. A total of 42 STEREO structures with no NOE violations greater than 0.5 Å was obtained. In the same manner a second set of 31 NOSTEREO structures was calculated using a reduced list of restraints in which all restraints involving distinction of stereospecifically assigned β-methylene groups were replaced with restraints based on center averaging, and χ_1 torsion angle restraints were removed. This reduced restraints list comprised 409 interproton distance restraints along with the same 24 hydrogen bonding and 23 ϕ backbone dihedral angle restraints used previously. Again none of these 31 structures contained NOE violations greater than 0.5 Å. For each set of converged structures ⟨SA⟩, the individual SA structures were best fitted to each other and averaged to yield a mean structure, denoted as $\overline{SA}$. The mean structures exhibit poor stereochemistry and non-bonded contacts, which was improved by 1000 cycles of restrained minimization to give structures denoted as $(\overline{SA})r$. These restrained minimized average structures are closer to the respective mean structure than any of the individual SA structures, while satisfying the experimental restraints as well as any of the individual structures.

The structural statistics for the two sets of structures, as well as for the respective mean and restrained minimized average structures, are listed in table 2. The individual structures and the two $(\overline{SA})r$ structures exhibit very small deviations from idealized covalent geometry and have good non-bonded contacts, evidenced by small values of the van der Waals repulsion potential F_{repel} and a negative Lennard-Jones van der Waals energy.

The rms deviations from the experimental data are slightly higher for the STEREO structures versus the STEREO restraints than for the NOSTEREO structures versus the NOSTEREO restraints (see table 2). This simply reflects the fact that as one tightens the restraints and increases their number, it becomes more difficult to generate structures that satisfy the tighter restraints as well as the looser ones. Given the potential errors in the estimation of distances from NOEs in protein spectra, the interproton distance rms deviations for both sets of structures can be regarded as small, and the difference between them is not significant.

The high computational efficiency of the hybrid distance geometry-dynamical simulated annealing

approach (~1 h per structure on a VAX 8550) permits the calculation of a relatively large number of structures for each set of restraints in a reasonably short time. As a result we are able to draw statistically meaningful conclusions about the differences in the structures obtained with and without stereospecific assignments.

3.3. *Comparison of the STEREO and NOSTEREO structures*

Fig.1 illustrates the superposition of the individual STEREO and NOSTEREO ⟨SA⟩ structures. The overall fold of the protein is clearly similar for the two sets of structures. Fig.2 shows both a histogram representing the distribution of

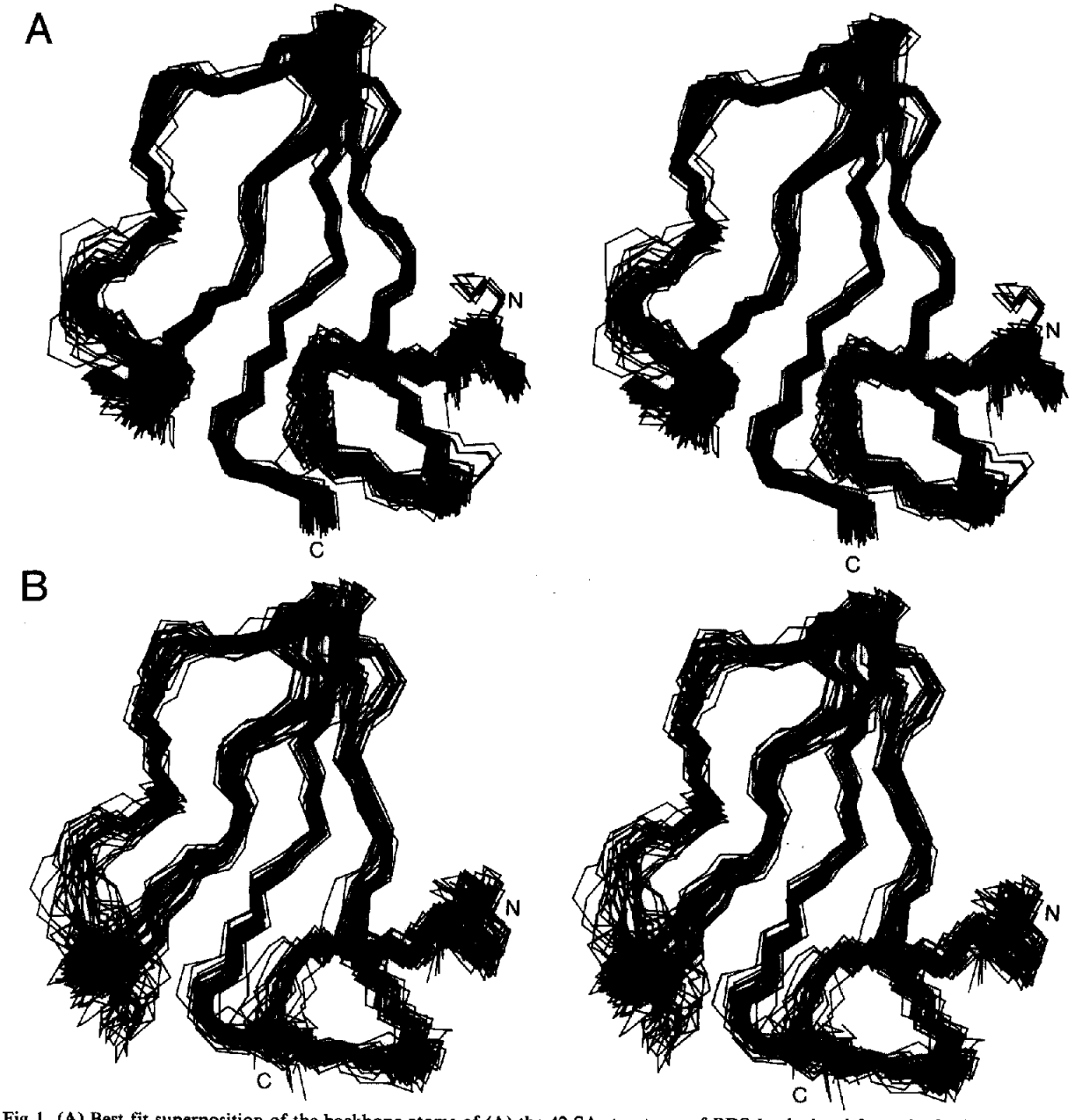


Fig.1. (A) Best fit superposition of the backbone atoms of (A) the 42 SA structures of BDS-I calculated from the STEREO restraint list, and (B) the 31 SA structures calculated from the NOSTEREO restraint list.

Fig.2. (A) Histogram illustrating the distribution of the additional NOEs derived from the stereospecific assignment of β-methylene groups. Each bar represents the difference in the number of NOE restraints between the STEREO and NOSTEREO restraint lists for each residue. The black portion of the bar indicates the number of additional NOEs involving the stereospecifically assigned β-methylene protons of that residue. The white portion of the bar indicates the number of additional NOEs from the rest of the amino acid to β-methylene protons of other residues. (B) The atomic rms difference between the backbone atoms of the mean structures calculated from the STEREO ($\overline{SA}$st) and NOSTEREO ($\overline{SA}$nst) restraint lists. (C–E) Atomic rms differences between the individual structures and the respective mean structures for the STEREO (open circles and dashed lines) and NOSTEREO (closed circles and solid lines) structures for backbone atoms, all atoms and side chain atoms, respectively.

additional NOEs included in the list of restraints for the STEREO structures relative to the NOSTEREO structures, and the statistical atomic rms difference data relating to the two sets of structures. In fig.2A the black bars give the number of additional NOEs involving each residue's own β-methylene group (i.e. for residues whose $C^{\beta}H_2$ protons were stereospecifically assigned), whereas the open bars indicate the number of additional NOEs arising from short distances between other parts of the amino acid and the stereospecifically assigned β-methylenes of other residues. The majority of the additional NOEs correspond to interresidue rather than intraresidue distances. The inclusion of stereospecifically assigned NOE restraints in the calculations results in a significant improvement in the overall definition of the backbone structure (i.e. the atomic rms distribution about the mean coordinate positions is reduced). It is important to note that the structure is actually slightly altered by the inclusion of additional restraints, in so far that the atomic rms difference between the two mean structures is larger than the atomic rms difference between the individual structures and their respective mean structure (see table 3). The change in structure reflects the fact that by removing the center averaging approximation for several of the β-methylene groups, the target function that is minimized in the simulated annealing calculations is altered. A shift in the protein fold is evident from fig.2B which shows the atomic rms difference between backbone atom positions of the best-fitted STEREO and NOSTEREO mean structures as a function of residue number.

The NOSTEREO structures do not satisfy the STEREO restraints at all well, as shown by the large interproton distance rms deviations and the large numbers of NOE violations in the 'NOSTEREO structures versus STEREO restraints' section of table 2. Indeed with regard to the restraints involving the side chain orientations, the mean χ_1 torsion angles of ten of the side chains in the NOSTEREO structures lie outside the ranges determined experimentally and used as restraints for the STEREO structures.

The atomic rms differences between the individual structures and the respective mean structure as a function of residue number is shown in fig.2C–E. This shows that the improvement in the definition of the structure is almost uniformly distributed along the protein chain. The most sizeable differences occur for the stretch of residues 25–30 which forms part of the long irregular loop that runs from residues 18–30, the

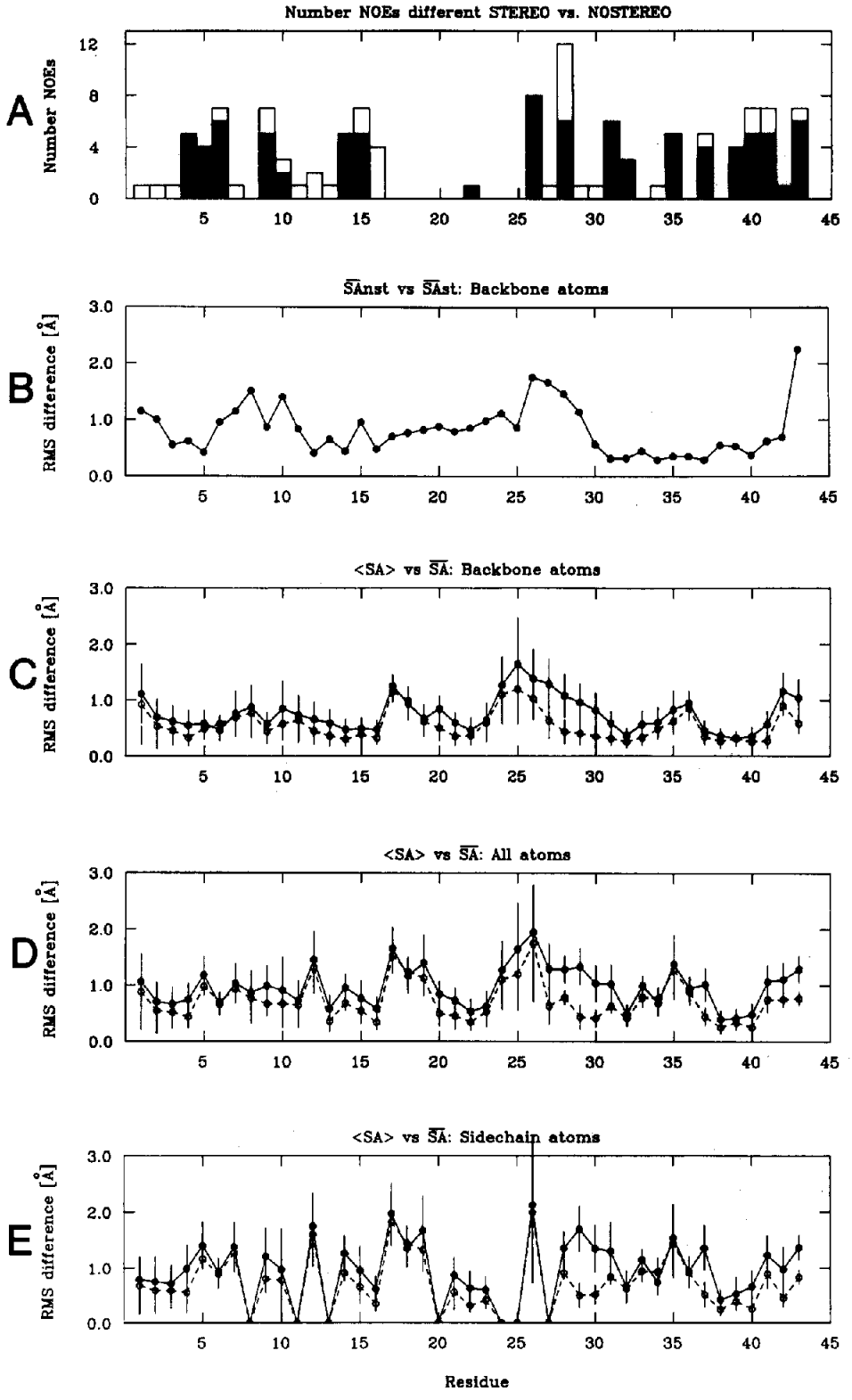
Table 3

Atomic rms differences[a]

	Atomic rms difference (Å)	
	Backbone atoms	All atoms
STEREO structures		
⟨SA⟩ vs $\overline{SA}$[b]	0.67 ± 0.12	0.90 ± 0.17
⟨SA⟩ vs ⟨SA⟩[b]	0.96 ± 0.19	1.29 ± 0.23
⟨SA⟩ vs ($\overline{SA}$)r	0.73 ± 0.13	1.00 ± 0.18
($\overline{SA}$)r vs $\overline{SA}$	0.21	0.45
NOSTEREO structures		
⟨SA⟩ vs $\overline{SA}$[b]	0.93 ± 0.16	1.17 ± 0.17
⟨SA⟩ vs ⟨SA⟩[b]	1.33 ± 0.26	1.65 ± 0.31
⟨SA⟩ vs ($\overline{SA}$)r	1.02 ± 0.17	1.29 ± 0.20
($\overline{SA}$)r vs $\overline{SA}$	0.42	0.54
STEREO vs NOSTEREO structures		
⟨SA⟩$_{STEREO}$ vs ⟨SA⟩$_{NOSTEREO}$	1.50 ± 0.19	1.85 ± 0.30
$\overline{SA}_{STEREO}$ vs $\overline{SA}_{NOSTEREO}$	0.93	1.21

[a] The notation of the structures is the same as that in table 2

[b] Note that using standard statistical theory it is easily shown that the average atomic rms difference between all pairs of SA structures is related to the average atomic rms difference between the individual SA structures and the mean $\overline{SA}$ structure by a factor of $\sim[2n/(n-1)]^{1/2}$


Number NOEs different STEREO vs. NOSTEREO
A
Number NOEs
12
8
4
0
5
10
15
20
25
30
35
40
45
SAnst vs SAst: Backbone atoms
B
RMS difference [Å]
3.0
2.0
1.0
0.0
<SA> vs SA: Backbone atoms
C
<SA> vs SA: All atoms
D
<SA> vs SA: Sidechain atoms
E
Residue

backbone at residue 20, the side chains of residues 9, 14, 15, 16, 34 and 37, and the C-terminal segment 41–43. Not surprisingly, the improvement in the definition of the local structure is correlated with the additional NOE distribution shown in fig.2A.

Differences in the definition of the structure are most clearly observed for the side chains. This is illustrated by a comparison of the side chain distributions for residues 13–16 and 28–33 shown in fig.3. The inclusion of stereospecific assignments and χ_1 torsion angle classifications into the restraints involving residues Asp-14, Leu-15, Tyr-28, Ser-30, Asn-31 and Cys-32 results in a dramatic reduction in the spread of side chain positions in the STEREO structures with respect to the NOSTEREO structures. An improvement in the definition of the position of the aromatic ring of Tyr-33, a residue for which stereospecific assignments of the $C^{\beta}H$ resonances and χ_1 classification was not possible from the NMR data because of chemical shift degeneracy, is clearly observed. This indicates that the inclusion of extra short range NOEs connecting Tyr-33 to the stereospecifically assigned β-methylene protons of Cys-32, together with the increased definition of neighbouring side chains (Tyr-28 and Asn-31), results in an improvement in the structure remote from the β-methylene groups that are stereospecifically assigned, and illustrates the cooperative nature of interatomic distances in a protein.

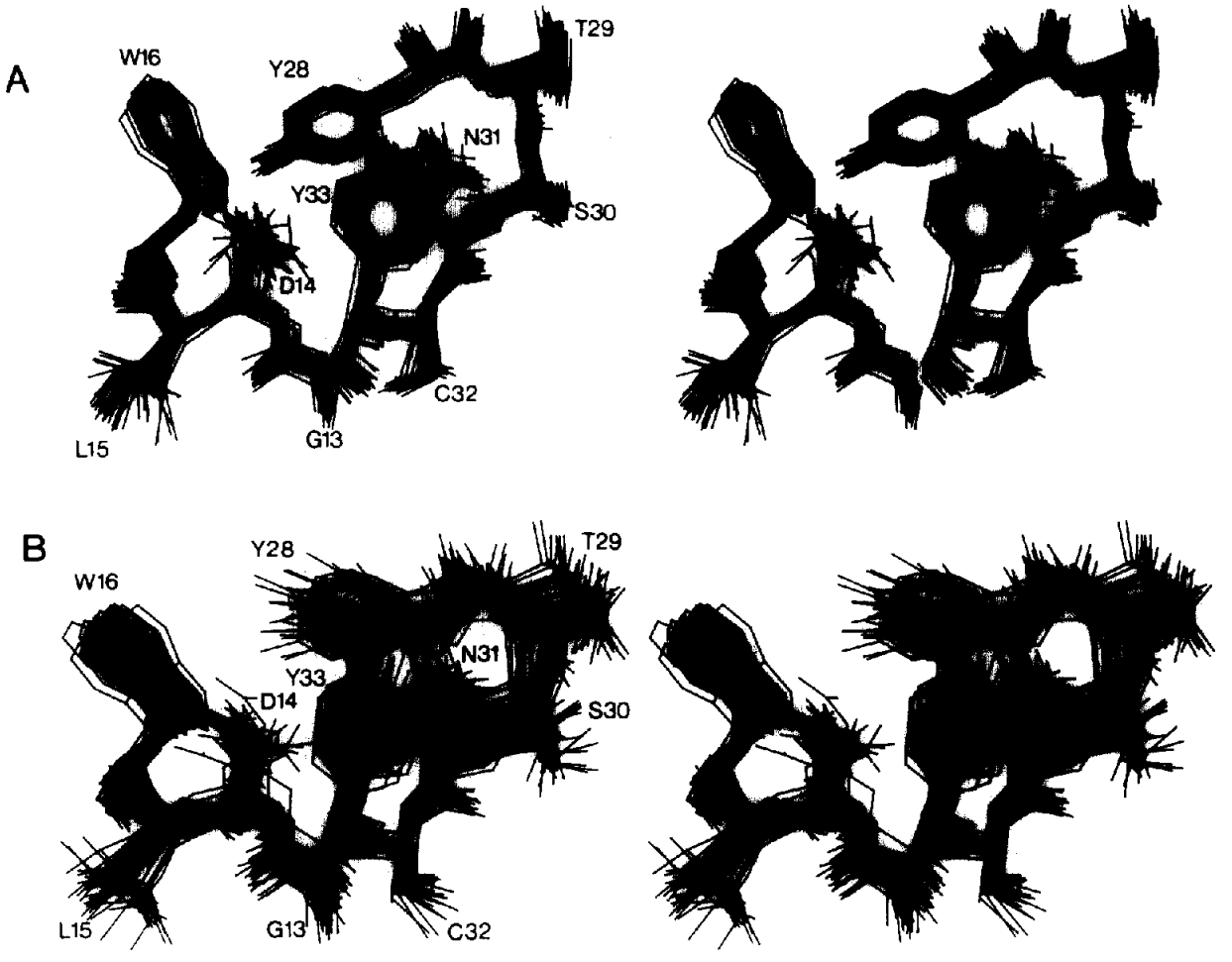


Fig.3. (A) Superposition of the heavy atoms of a representative set of side chains from (A) the 42 SA structures of BDS-I calculated from the STEREO restraint list, and (B) the 31 SA structures calculated from the NOSTEREO restraint list.

This improvement is also confirmed on a statistical basis from the atomic rms difference data presented in table 3. The atomic rms difference between the individual structures in each set and the respective mean structure upon inclusion of the STEREO restraints is reduced from 0.93 to 0.67 Å for the backbone atoms and from 1.17 to 0.90 Å for all atoms. These figures correspond to 28% and 23% improvements in the definition of the backbone fold and overall structure, respectively.

The largest improvement in definition is found for Tyr-28, and is associated with the largest number of additional NOEs arising from stereospecific assignments. In our previous paper [5] we indicated that the juxtaposition of the side chains of Trp-16 and Tyr-28 suggested the presence of a hydrogen bond between the $N^{\epsilon 1}H$ of the indole ring and the hydroxyl O^{η} of the tyrosine ring. Such a hydrogen bond might provide some restriction of the tyrosine ring consistent with its slowly flipping character on the NMR timescale [4]. Evidently, because of the poor side chain definition, such a conclusion could not be drawn from the NOSTEREO structures (see fig.3).

4. CONCLUDING REMARKS

The determination of protein structures from NMR data depends upon a large number of approximate short range (<5 Å) interproton distances. The attainable precision therefore depends to a large degree on the maximization of the number of these shorter distance estimates. In this paper we have shown that the use of stereospecific assignments of prochiral β-methylene proton resonances and the incorporation of χ_1 torsion angle restraints result in a significant improvement in the definition of the structure of BDS-I. Such highly defined structures allow the analysis of detailed side chain interactions both in the core of the protein as well as on the solvent exposed surface, for example in terms of interresidue hydrogen bonds and ring current shift effects, which otherwise might not be possible [5].

Acknowledgements: This work was support by the intramural AIDS targeted antiviral program of the Office of the Director of the National Institutes of Health. We thank Dr Laszlo Béress for the gift of BDS-I.

REFERENCES

[1] Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley, New York.
[2] Clore, G.M. and Gronenborn, A.M. (1987) Protein Engineering 1, 275–288.
[3] Clore, G.M. and Gronenborn, A.M. (1989) CRC Critical Reviews in Biochemistry, in press.
[4] Driscoll, P.C., Clore, G.M., Béress, L. and Gronenborn, A.M. (1988) Biochemistry, in press.
[5] Driscoll, P.C., Gronenborn, A.M., Béress, L. and Clore, G.M. (1988) Biochemistry, in press.
[6] Griesinger, C., Sørensen, O.W. and Ernst, R.R. (1987) J. Magn. Reson. 75, 474–492.
[7] Rance, M., Bodenhausen, G., Wagner, G., Wüthrich, K. and Ernst, R.R. (1985) J. Magn. Reson. 62, 497–510.
[8] Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) FEBS Lett. 229, 317–324.
[9] Havel, T.F. (1986) DISGEO, Quantum Chemistry Program Exchange No. 507, Indiana University.
[10] Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M. (1986) Proc. Natl. Acad. Sci. USA 83, 3201–3805.????
[11] Brünger, A.T., Kuryan, J. and Karplus, M. (1987) Science 235, 458–460.
[12] Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M. (1987) Protein Engineering 1, 399–406.
[13] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminatham, S. and Karplus, M. (1983) J. Comput. Chem. 4, 187–217.
[14] Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Go, N. and Wüthrich, K. (1987) J. Mol. Biol. 196, 611–639.
[15] Ponder, J.W. and Richards, F.M. (1987) J. Mol. Biol. 193, 775–791.
[16] Hyberts, S.G., Märki, W. and Wagner, G. (1987) Eur. J. Biochem. 164, 625–635.
[17] Clore, G.M., Gronenborn, A.M., Carlson, G. and Meyer, E.F. (1986) J. Mol. Biol. 190, 259–267.
[18] Arseniev, A., Shultze, P., Wörgötter, E., Braun, W., Wagner, G., Vasák, M., Kägi, J.H.R. and Wüthrich, K. (1988) J. Mol. Biol. 201, 637–657.
[19] Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, A.T., Karplus, M. and Gronenborn, A.M. (1986) EMBO J. 5, 2729–2735.
[20] Wüthrich, K., Billeter, M. and Braun, W. (1983) J. Mol. Biol. 169, 949–961.