3D structure of bovine pancreatic ribonuclease A in aqueous solution: An approach to tertiary structure determination from a small basis of ^{1}H NMR NOE correlations

Manuel Rico, Jorge Santoro, Carlos González, Marta Bruix, José Luis Neira, José Luis Nieto and José Herranz

Instituto de Estructura de la Materia, Consejo Superior de Investigaciones Científicas (CSIC), Serrano 119, E-28006 Madrid, Spain

SUMMARY

A method is proposed to generate initial structures in cases where the distance geometry method may fail, such as when the set of ^{1}H NMR NOE-based distance constraints is small in relation to the size of the protein. The method introduces an initial correlation between the φ and ψ backbone angles (based on empirical observations) which is relaxed in later stages of the calculation. The obtained initial structures are refined by well-established methods of energy minimization and restrained molecular dynamics. The method is applied to determine the solution structure of Ribonuclease A (124 residues) from a NOE basis consisting of 467 NOE cross-correlations (97 intra-residue, 206 sequential, 23 medium-range and 141 long-range) obtained at 360 MHz. The global shape and backbone overall fold of the eight final refined structures are close to those shown by the crystal structure. A meaningful difference in the positioning of the catalytically important His[119] side chain in the solution and crystal structures has been detected.

INTRODUCTION

Sequence-specific assignments for backbone proton resonances of most of the 124 residues of bovine pancreatic Ribonuclease A* have been recently reported by two independent groups, our own (Rico et al., 1989) and that of Scheraga and coworkers (Robertson et al., 1989). Assignments are practically coincident. Secondary structure elements were identified in both works from observed sequential, and medium- and long-range NOE effects, together with data from amide proton exchange rates. Thus, helical regions, β-strands in regular antiparallel β-sheet structures, β-

---

*Enzyme: RNase A, bovine pancreatic ribonuclease A (EC 3.1.27.5).

turns and bulges in β-strands were identified, all of which showed a close correspondence with those observed in the crystal structure (Borkakoti et al., 1982; Wlodawer et al., 1988).

Since our work was carried out at 360 MHz, we had to deal with severe problems of signal overlap and sensitivity. In fact, we obtained a relatively small basis of NOE correlations (467 total NOEs, 97 intra-residue, 206 sequential, 23 medium-range and 141 long-range). The problem was whether it would be possible to arrive at an initial structure susceptible of being refined in a later stage by well-known methods of restrained molecular dynamics. Attempts to generate an initial structure by the distance geometry method (Havel et al., 1983) failed. Thus, we devised a special strategy involving fundamentally the introduction in the 'variable target' method (Braun and Gō, 1985; Braun, 1987) of (a) an initial restriction between φ and ψ torsion angles (a 'most populated' ellipse relationship in the Ramachandran plot is accepted) and (b) the use of well-defined elements of secondary structures. By driving the computation in a stepwise manner (starting from a short peptide and lengthening it) we were able to generate a number of substructures from which we selected one (on the basis of minimum violation of distance constraints) to be refined by well-established methods of dynamical simulated annealing. Final refined structures show a global shape and backbone overall fold close to those obtained for the crystal structure. They will be susceptible to checking against a most refined solution structure on the basis of 2D spectra at 600 MHz, which will be available shortly.

In this paper, we describe in detail the method and its application to the determination of the tertiary structure of RNase A in solution from the set of NOEs observed at 360 MHz. The method may be of general applicability in those cases where, for some reason (poor solubility, aggregation or other), the NOE basis to impose distance constraints is small in relation to the number of residues of the protein.

MATERIALS AND METHODS

Samples and NMR spectra

RNase A was purchased from Sigma Chemical Co. and used without further purification. As stated in Rico et al. (1989), this preparation contains phosphate in sufficient amounts (as shown by ^{31}P NMR) to saturate the enzyme. All results in this paper refer to phosphate-bound RNase A in H_2O or D_2O solution, 8 mM, 0.2 M NaCl, uncorrected pH or pD 4.0 and 35°C, unless otherwise stated. ^{1}H NMR 2D COSY and NOESY spectra were as described in Rico et al. (1989).

The set of NOE cross-correlations used in the calculation of tertiary structures of RNase A (included as supplementary material) was drawn from NOESY spectra in H_2O and D_2O obtained at 360 MHz with a mixing time τ_m of 150 ms. NOESY spectra were also obtained with τ_m of 100 and 70 ms to evaluate qualitatively the contribution of spin diffusion to the intensity of NOE cross-correlation peaks. The effect of spin diffusion was not high as judged by the absence of marked

changes in the relative intensity of NOE cross-correlations at different mixing times. Also, different intensities were observed for NOEs to methylene proton pairs or *ortho-meta* aromatic protons, which is often taken as a sign of a low degree of spin diffusion.

Computational method

There is not at present a single generally accepted method for biopolymer structure determination from NMR data. Restrained energy minimization (REM) and restrained molecular dynamics (RMD) have shown to be extremely useful to that end, although they may not be cost-effective in terms of computer time. CPU times can be strongly reduced by starting from plausible initial structures, such as those generated by applying distance geometry calculation. Hybrid methods in which initial structures are refined by REM and RMD are at present a suitable compromise to the goal of computing tertiary structures at a reasonable cost (Kaptein et al., 1988; Nilges et al., 1988).

The method used in this work is a hybrid one. Because of the small number of restrictions obtained from the NMR spectra compared to the size of the protein, distance geometry methods were found to be inappropriate to generate initial structures. In fact, all our trials to compute structures with this method failed. To overcome this problem, we have developed a new program, on the basis of the variable target function method described in Braun and Gō (1985), that performs all calculations in real space:

$$T_k = T_k^{dc} + T_k^{so} \qquad (1)$$

The first term on the right depends on the distance constraints, and is defined as

$$T_k^{dc} = \sum_{i < j \in F_k} w_{ij}(k) \cdot (r_{ij}^2 - u_{ij}^2) \cdot \theta(r_{ij}^2 - u_{ij}^2) \qquad (2)$$

where the summation extends over all distance constraints between atoms belonging to the atom set F_k; F_k is the set of atoms considered in step k; r_{ij} is the distance between atoms i and j; u_{ij} is the upper bound of this distance; θ is the step function:

$$\theta(y - x) = \begin{array}{ll} 1; & \text{if } y \geqslant x \\ 0; & \text{otherwise} \end{array} \qquad (3)$$

and $w_{ij}(k)$ is the weight of the distance constraint between atoms i and j in the optimization step k. The weights are initially defined as

$$w_{ij}(1) = \max\left(\frac{10}{(N_j - N_i + 1)^2}, 0.5\right) \qquad (4)$$

where N_i and N_j are the sequential numbers of the residues to which atoms i and j belong, and they are updated recursively

$$w_{ij}(n+1) = \frac{min(2W_{ij}(n), 10); \quad if \ i,j \in F_k}{w_{ij}(n) \qquad ; \quad otherwise} \tag{5}$$

The terms T_k take into account the presence of wrong sterical overlaps, and are defined as

$$T_k^{so} = \sum_{m < n \in F_k} (1_{mn}^2 - r_{mn}^2) \cdot \theta(1_{mn}^2 - r_{mn}^2) \tag{6}$$

where the summation extends over all atom pairs separated by three or more bonds, and 1_{mn} is the lower bound of the distance between atoms m and n, defined as the distance at which the repulsive energy of the non-bonded interaction m and n reaches 2.5 kcal.

The efficiency of the minimization process depends critically on how the variation of the target function is carried out. In our protocol, we start with a very short peptide and we slowly lengthen it until the whole chain is considered. Typically, each step includes four residues more than the previous one. Each time that a new fragment is included, the weights for the constraints are updated and 25 steps of conjugate gradient minimization are carried out.

In our method, as in Braun and Gō (1985), the rigid geometry approximation is used and the minimization is performed in the torsional angle space. To restrict as much as possible the dimensionality of the problem, we have further reduced the accessible conformational space by making use of the fact that extensive regions of the conformational space are sparsely populated. Inspection of the computed distribution of the backbone dihedral angles, φ and ψ, shows a high population in the right-handed α-helix and β-sheet regions, as well as significant population in the connecting area between them and in the left-handed α-helix region (Fig. 1). All these regions,
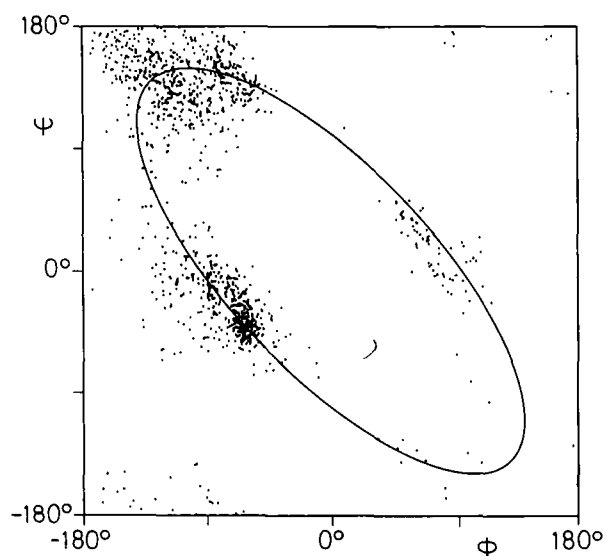

Fig. 1. Plot of φ, ψ values for a representative sample of highly refined X-ray structures of six proteins. Best-fit ellipse covering the observed φ, ψ points.

together with those corresponding to different kinds of turns (Richardson, 1981), can be visited by walking along the ellipse

$$\varphi = -141 \cos(\gamma - 22.5)$$
$$\psi = 149 \cos(\gamma + 20.0)$$

(7)

We have replaced the two backbone dihedral angles for each residue by the single variable γ. This new reduction of the conformational space, including the γ parameter and the side-chain dihedral angles, strongly improves the convergence of the minimization process. It must be pointed out here that a similar ellipse has been considered (Pohl, 1980) as an average pathway for conformational changes in proteins.

The resulting structures of this minimization were used as starting coordinates in a restrained molecular dynamics simulation with the GROMOS package (Van Gunsteren and Berendsen, 1987). First, the structures were submitted to 300 steps of REM to bring the energy down to acceptable values. Strong gradients were promptly reduced so that the simulation can start from relatively relaxed structures. The second step consists of 20 ps of RMD per structure. The simulations were first carried out at 300 K. The temperature was kept constant by rescaling it every 50 steps. The integration step was 2 fs, and the SHAKE algorithm was used to keep the bond lengths in standard values. The first 15 ps of the simulations were considered as an equilibrium period and the last 5 ps were used for averaging. The averaged structures were finally submitted to a REM. During all the calculations the force field used was the same. The NOE term was:

$$U_{NOE} = \begin{array}{ll} K_{NOE}(r_{ij}^2 - u_{ij}^2); & \text{if } r_{ij} > u_{ij} \\ 0 & ; \quad \text{otherwise} \end{array}$$

(8)

with $K_{NOE} = 40.0 \text{ kJ mol}^{-1} \text{ Å}^{-2}$.

The lowest-energy member of the set of structures was then subjected to a more elaborate refinement procedure consisting of an extensive high-temperature RMD simulation following an annealing strategy (Vlieg et al., 1988). The structure was first slowly heated during 10 ps up to 1000 K. At this temperature, 40 ps of RMD were carried out. Eight structures were extracted from the trajectory, one every other 5 ps, and were then submitted to a cooling strategy consisting of 5 ps at 750, 5 ps at 500 and 10 ps at 300 K. The last 5 ps were used for averaging. The final structures were energy-minimized.

RESULTS

Assignments

The two reported assignments for RNase A coincide practically for 119 out of 124 residues, once small differences in referencing and external conditions are duly taken into account. Delta-(δ)-values in Robertson et al. (1989) are measured with reference to DSS, whereas those in Rico et al. (1989) are given with respect to TSP, so that a -0.05 ppm correction must be applied to the former for the two to be comparable. The measuring temperature was 30°C in the former, whe-

reas that in the latter was 35°C, thus causing an average downfield shift for the former NH δ values of ~0.01 ppm. The reported δ values in Robertson's work are for pH 3, whereas those in our study are for pH 4. A number of resonances are affected by titrating groups in this range, the most significant of which are, for $C_\alpha H$ resonances, 38, 39, 72, 90 and 95, and for NH amide resonances, 45, 49, 50, 66, 67, 72, 80, 86, 91, 92, 96, 105, 120 and 121.

On the basis of spectra at a higher field, we have confirmed as correct amide and $C_\alpha H$ resonances assigned as probable and highly probable, as well as those for Pro[114] and Asp[121] assigned as tentative. The tentative assignment of Pro[42] has to be changed to the one given in Robertson's work. Conversely, the assignment given by them for Asp[121] as the cross-correlation peak in the COSY fingerprint region at (8.82, 4.87 ppm) is in fact that corresponding to Val[108], the one arising from Asp[121] being lost because of saturation at the water peak frequency. Table 1 collects definitive assignments for the five unassigned, tentative or conflicting residues. Altogether, the almost total coincidence of both assignments performed in a fully independent way is a sign of the high logical consistency of the sequence-specific assignment as elaborated by Wüthrich and co-workers (Wüthrich, 1986).

Application of the method to the determination of the solution structure of RNase A

Extensive analysis of 2D NOE spectra of RNase A at 360 MHz provided a list of 467 unambiguously assigned NOE proton correlations, 97 of which were intra-residue, 206 sequential, 23 medium range (2–5 residues) and 141 long range (≥ 5 residues). The observed NOEs were translated into upper-limit distance constraints according to their intensity, by using the following qualitative criteria: strong NOEs were set to distances lower than 2.5 Å; medium, lower than 3.5 Å; and weak, lower than 4.5 Å. Threonines 45, 82, 99 and 100, valines 47, 54 and 57 and isoleucine 107, all showing $^3J_{\alpha\beta} \geq 8$ Hz, were fixed to the conformations where the bonds $C_\alpha H$ and $C_\beta H$ are in a *trans-periplanar* orientation by adding an extra term to the torsional potential. Local NOEs were used to assign stereo-specifically the methyl groups of valine residues. The pseudoatom approach was used in all other cases, to allow for the lack of stereospecific resonance assignments. Bond distance constraints were imposed between sulfide atoms in cysteine residues intervening in the native disulfide bridges.

During the sequential assignment, the following elements of secondary structure were determined: α-helices, 4–12, 23–33 and 51–59 and β-strands, 43–49, 72–75, 80–85, 94–111 and 116–124.

TABLE 1

DEFINITIVE ASSIGNMENTS OF ^1H RESONANCES (δ, PPM FROM TSP) FOR CONFLICTING RESIDUES IN BOVINE PANCREATIC RNASE A (pH 4.0, 35°C)

Amino acid	NH	$C_\alpha H$	$C_\beta HH'$		$C_\gamma HH'$		$C_\delta HH'$	
Pro[42]	—	4.53	2.48	2.13	2.25	2.15	4.09	3.92
Val[108]	8.73	4.84	2.03		0.78[a]	0.64[a]		
Pro[114]	—	4.66	2.33	1.88	1.88	1.77	3.62	3.46
Pro[117]	—	4.47	1.03	0.23	1.55		3.65	3.41
Asp[121]	8.71	4.71	2.23	2.07				

[a] Methyl groups.

In all initial structures, these elements were conserved by setting the γ parameter in Eq. (2) to 1.57 and −0.30 for residues in α-helices and β-strands, respectively. For the remaining regions (loops and bulges), several starting structures were generated from a combination of residue backbone conformations compatible with backbone sequential NOEs. Thus, strong sequential αN or NN NOEs were accepted to fix points along the ellipse for residue i corresponding, respectively, to the standard right-handed α-helix and β-strand backbone conformations. In cases where both NOEs are observed with enough intensity, a further point in the ellipse is admitted. When no sequential NOE was detected (residue 67) three initial conformations were accepted. The initial conformations generated in this way for the segments with non-regular structures (13–22, 34–43, 60–71, 76–80, 86–93 and 112–115) were subjected separately to minimization of the target function in Eq. (1). Many of the initial structures were trapped in local minima with exceedingly high relative values for the target function, and were not further considered. About two or three structures for each segment were conserved. Even after the above reduction of starting structures, the total number of remaining combinations was exceedingly large, therefore a merging of initial structures was carried out by successive cycles of minimization after a stepwise incorporation of further structural elements in the sequence. In each minimization cycle, a number of structures were rejected on the basis of their value of the target function.

After concluding this process, we found eight independent substructures for which the minimization of the target function was successfully accomplished, in the sense that relatively low values of distance-constraint violations and sterical overlaps could be reached. Table 2 lists the total sums of the violations of NOE distance constraints and the total sum of atomic overlaps for the eight substructures.

As may be seen in Table 2, all eight substructures show extremely high values for the sum of atomic overlaps, and consequently have unacceptable energy values. They were then subjected to several steps of restrained energy minimization by using the GROMOS force field, with the result

TABLE 2

SUM OF ATOMIC OVERLAPS, Σ_{AO}(Å), AND SUM OF DISTANCE-CONSTRAINT VIOLATIONS, Σ_{dviol} (Å), FOR THE EIGHT INITIAL SUBSTRUCTURES, AND POTENTIAL ENERGY, E_{pot} (kJ mol^{-1}), NOE ENERGY TERM, E_{NOE} (KJ mol^{-1}), AND SUM OF DISTANCE-CONSTRAINT VIOLATIONS, Σ_{dviol} (Å), OF THE EIGHT SUBSTRUCTURES, ONCE RELAXED FROM THE ELLIPSE CONSTRAINT BY EM AND MD

Substructure	Initial		After EM + MD		
	Σ_{AO}[a]	Σ_{dviol}	E_{pot}[b]	E_{NOE}[c]	Σ_{dviol}
1	296	40.6	$-7.4 \cdot 10^3$	360	34.6
2	274	39.5	$-7.2 \cdot 10^3$	360	32.0
3	262	35.1	$-8.0 \cdot 10^3$	140	19.3
4	290	39.6	$-6.1 \cdot 10^3$	340	28.3
5	337	49.5	$-7.9 \cdot 10^3$	460	34.7
6	452	75.9	$-6.5 \cdot 10^2$	520	41.6
7	276	38.7	$-6.6 \cdot 10^3$	320	31.4
8	274	33.9	$-7.4 \cdot 10^3$	100	14.5

[a] Exceedingly high energy.
[b] Parametrization according to the GROMOS package.
[c] A force constant $K_{NOE} = 40.0$ kJ mol^{-1} Å^{-2} was used.

TABLE 3

RMS DIFFERENCES[a] (Å) BETWEEN THE EIGHT OBTAINED SUBSTRUCTURES AFTER EM + MD, FOR ALL BACKBONE ATOM POSITIONS

Substructure	1	2	3	4	5	6	7	8	XR
1		2.6	3.5	2.7	4.0	3.3	2.7	3.4	3.1
2			2.7	3.1	3.7	3.7	1.3	3.5	3.3
3				3.1	3.0	4.5	2.4	3.6	3.3
4					4.4	3.8	2.9	2.9	3.0
5						5.1	3.7	4.8	4.3
6							3.7	3.6	3.4
7								3.5	3.3
8									1.8

[a] RMS differences between the eight substructures and the crystal (XR) structure.

of a quick lowering of the corresponding potential energies, without apparent global changes in their conformation. It should be noted that the high values of the potential energy are caused, in part, by the 'ellipse' constraint, and that in the REM process (Cartesian coordinates) this constraint is obviously relaxed.

The resulting strainless substructures were then subjected to a process of restrained molecular dynamics simulation by using the GROMOS package with a refinement time of 20 ps. Final energy values were remarkably lowered, and they showed more similar values between them, which is an indication that well-defined minima had been reached. In Table 2, the energy values and the distance-constraint violations for the eight substructures after the REM and RMD processes are compared. It should be noted that the sums of distance-constraint violations decrease also after the RMD process, which means that the corresponding energy term is consistent with the GROMOS force field. All eight substructures show values for the potential energy and constraint violations in the same order of magnitude. Substructure 8 shows the lowest values for both terms.

Root-mean-square (rms) differences between the various substructures for the backbone (N–Cα–C) atom positions of all 124 residues were calculated. Rms differences were also calculated between the eight substructures and that obtained for the crystal state. All values (around 3 Å) are collected in Table 3.

Figure 2 shows a plot of the root-mean-square deviations of the backbone torsional angles of the eight substructures vs. the residue number. It can be observed that maximum deviations correspond mainly to loop regions, such as 16–22, 34–38, 59–69, 87–91 and 114–116. In these regions, several factors contributing to backbone conformational undefinition coincide: absence of NOEs involving side chains; presence of αN and NN weak correlations; absence of both; presence of glycines lacking stereospecific assignment, etc. Maxima at 24–25, 41, 48–49 and 118–122 are mainly due to the sorting out of the eight substructures into two families, differing not much in the Cα location but rather in the orientation of the peptide bond planes. Inspection of local distance-constraint violations often discriminates between the two alternatives. Thus, for instance, in the region 118–122, two groups of substructures can be spotted: 1–2–6–7 and 3–4–5–8. Local distance-constraint violations averaged over the four substructures are 12.3 Å and 4.6 Å, respectively, thus favouring the second family as the right one. Likely, structures 1–2–6–7 are trapped in local minima from which the minimization process used is unable to take them out.
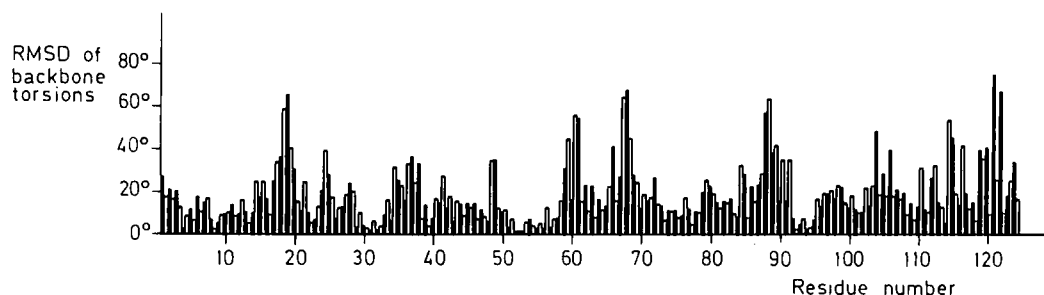
Fig. 2. Plot of the rms deviations of the backbone torsional angles of the eight resulting substructures vs. the residue number (φ, filled bar; ψ, empty bar).

Substructure 8, the lowest-energy member of the set above, was subjected to the more elaborate refinement process described above. Eight new structures were obtained which we consider equally compatible with the experimental data. A listing of atomic coordinates for all these eight final structures is given as supplementary material. Rms deviations of backbone atom positions between the eight structures are now around 2 Å. Figure 3 shows a stereoscopic view of the superposition of the eight structures and Table 4 lists values for the potential energy, the NOE energy term and the sum of distance constraint violations.

In Fig. 4, the rms deviations of the backbone torsional angles between the eight structures are plotted vs. the residue number. Inspection of Fig. 4 shows similar features to those previously observed in Fig. 2. Thus, the helix 50–60 and regions 5–13 and 25–32 of the first (3–13) and third (23–32) helices are well defined, and so are the β-strands 40–47, 70–86, 96–111 and 115–123 as well as the β-turn 111–114. Loop regions continue to show large deviations, although they have been decreased in a general way. Torsional rms deviations increase, with respect to substructures in Fig. 2, in the regions 33-39, 48–49 and 93–94. Let us consider separately the regions of largest deviations.

TABLE 4
VALUES OF THE POTENTIAL ENERGY, E_{pot} (kJ mol^{-1}), THE NOE ENERGY TERM, E_{NOE} (kJ mol^{-1}), AND THE SUM OF DISTANCE-CONSTRAINT VIOLATIONS, Σ_{dviol} (Å), FOR THE EIGHT FINAL STRUCTURES ARISING FROM SUBSTRUCTURE 8 BY DYNAMICAL SIMULATED ANNEALING

Final structure	E_{pot}[a]	E_{NOE}[b]	Σ_{dviol}
1	-7.5×10^3	81	12.5
2	-7.1×10^3	99	14.0
3	-7.8×10^3	54	9.7
4	-7.3×10^3	71	10.6
5	-8.1×10^3	59	11.2
6	-7.2×10^3	79	12.4
7	-8.1×10^3	71	12.7
8	-8.4×10^3	70	10.5

[a] Parametrization according to the GROMOS package.
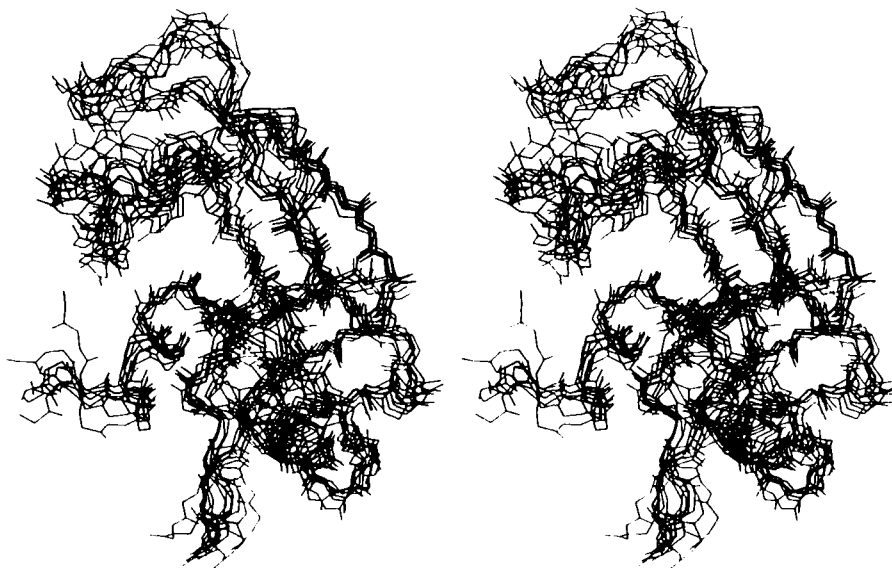[b] A force constant $K_{NOE} = 40.0$ kJ mol^{-1} Å^{-2} was used.

Fig. 3. Stereoscopic view of the superposition of the eight accepted final structures.

Region 17–24 is one of maximum variability. The largest peak in the torsional deviation plot (Fig. 4) is mainly due to the sorting out of the eight structures into two families differing in the orientation of the 18–19 peptide plane, which in turn implies large changes in the torsions $\psi 18$ and $\varphi 19$ without largely affecting the backbone conformation. Figure 5 shows a superposition of the averaged structures corresponding to the two families (structures 1–4–6 and structures 2–3–5–7–8), and illustrates this behaviour. There is a similar occurrence in the region 33–39, where the main difference between the two new families (now being 5–6–7 and 1–2–3–4–8) is the orientation of the peptide planes 33–34 and 38–39 (Fig. 6). Again, in the region 44–50 the large rms torsional deviations of residues 48 and 49 have their origin in the different orientation of the peptide plane 48–49 in structures 3–7 compared to that in the remaining ones (Fig. 7). There is a similar occurrence in the region 86–96, where the two alternative positions for the peptide plane 93–94 correspond, respectively, to structures 5–8 and structures 1–2–3–4–6–7 (Fig. 8). In all cases, the mean values of the local distance constraint violations for each family of structures have comparable values.
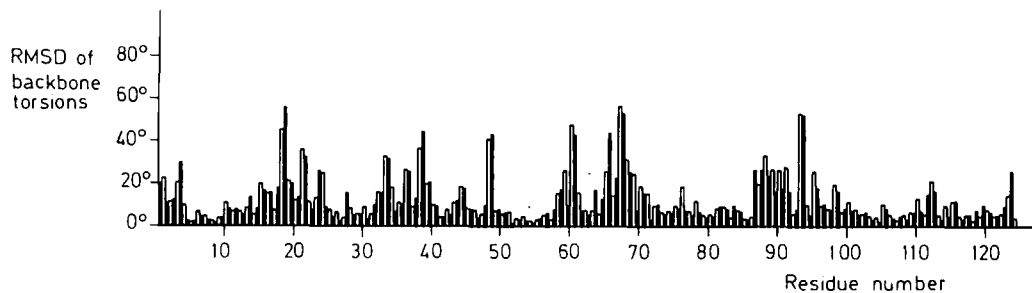

Fig. 4. Plot of rms deviations of the backbone torsional angles of the eight structures obtained from structure 8 after a more elaborated refinement process vs. the residue number (φ, filled bar; ψ, empty bar).
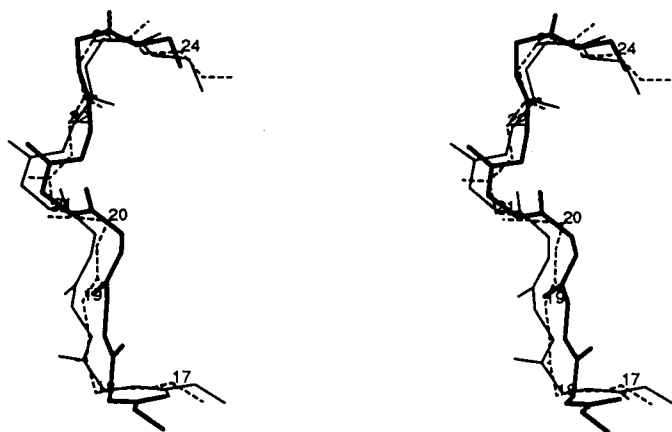
Fig. 5. Stereoscopic view of the superposition of three conformations of the loop Thr[17]-Asn[24] in RNase A: (a) solution averaged structure of final refined structures 1, 4 and 6 (thin line); (b) solution averaged structure of final refined structures 2, 3, 5, 7 and 8 (thick line); (c) X-ray diffraction structure in the crystal state (dotted line).

The wide loop 59–74 (Fig. 9) is the region of maximum variability in the enzyme structure. The presence of Gly[68] and the absence of any sequential connection of this residue with the previous one certainly contributes to that. On the other hand, it contains the distance constraint corresponding to the disulfide bridge between residues 65 and 72. The eight obtained structures can be classified into two families, structures 2 and 5 on one side, and the remaining ones on the other. Two different routes are followed by the two families in closing the loop. Torsion values in the peptide linkage of residues 60–61, 65–66 and 67–68 are interdependent. The sums of local distance-constraint violations are equally low for the two families (0.91 and 0.84 Å, respectively). In Fig. 9, the different conformations for the two families (both compatible with observed NOEs) may be compared. We should like to underline, as a general conclusion, that even in protein segments with a small number of distance constraints the number of possible different conformations
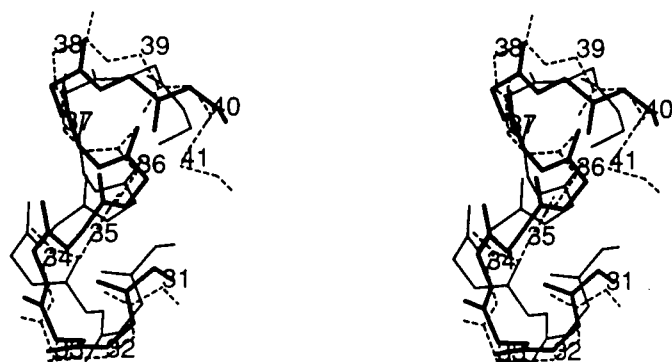

Fig. 6. Stereoscopic view of the superposition of three conformations of the loop Lys[31]-Lys[41] in RNase A: (a) solution averaged structure of final refined structures 5, 6 and 7 (thin line); (b) solution averaged structured of final refined structures 1, 2, 3, 4 and 8 (thick line); (c) X-ray diffraction structure in the crystal state (dotted line).
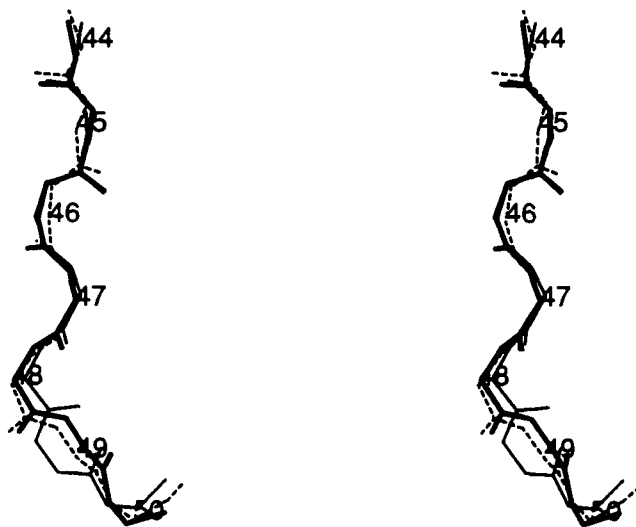
Fig. 7. Stereoscopic view of the superposition of three conformations of the segment Asn[44]–Ser[50] in RNase A: (a) solution averaged structure of final refined structures 3 and 7 (thin line); (b) solution averaged structure of final refined structures 1, 2, 4, 5, 6 and 8 (thick line); (c) X-ray diffraction structure in the crystal state (dotted line).

is not as high as may be thought in principle. In all loops of RNase A, structures can be classified into two families, differing very often in the orientation of a single peptide plane.

DISCUSSION

Although enormous progress has been achieved in the determination of solution structures of proteins from 2D-NMR data, the method is still subject to the intrinsic disadvantage of NMR spectroscopy, i.e., relatively low sensitivity, which in turn demands high protein solubility or high protein stability to stand long measuring times. Proteins showing low solubility, aggregation problems or progressive denaturation are not infrequent. In these cases, and even with the present state-of-the-art in high-field spectrometers, 2D-NMR spectra may show a poor signal-to-noise ratio. The 2D NOE data set may be too small for the usual methods of generating structures, such as distance geometry or variable target, to be effective. The method described here, with the additional constraints involving the backbone torsion angles, has been shown to be effective in gene-
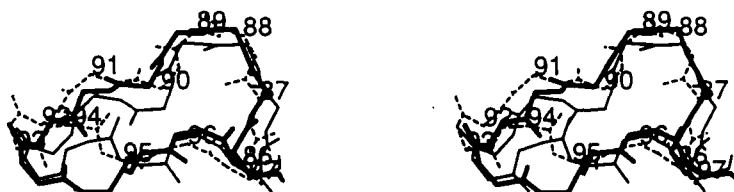

Fig. 8. Stereoscopic view of the superposition of three conformations of the loop Glu[86]–Tyr[97] in RNase A: (a) solution averaged structure of final refined structures 5 and 8 (thin line); (b) solution averaged structure of final refined structures 1, 2, 3, 4, 6 and 7 (thick line); (c) X-ray diffraction structure in the crystal state (dotted line).
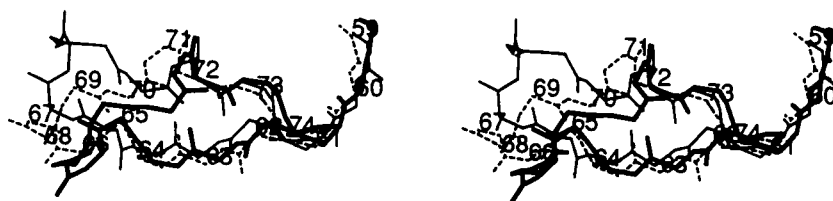
Fig. 9. Stereoscopic view of the superposition of three conformations of the loop $Ser^{59}-Asn^{74}$ in RNase A: (a) solution averaged structure of final refined structures 2 and 5 (thin line); (b) solution averaged structure of final refined structures 1, 3, 4, 6, 7 and 8 (thick line); (c) X-ray diffraction structure in the crystal state (dotted line).

rating initial structures susceptible to refining at a later stage. The reduction of the backbone conformational space to a most populated ellipse does not introduce any bias in the final $\varphi-\psi$ distribution, as may be deduced from a comparison of the starting φ and ψ angles with those obtained after REM and 20 ps of RMD simulation for each of the eight substructures.

The sampling efficiency of the process has been kept high by (a) the admission of a large number of initial structures, (b) a relatively long time for RMD simulation at low temperature, and (c) the application of an RMD simulation at high temperature. Large rms torsion differences (or their equivalent rms backbone atom positions) in loops or other regions where distance constraints are scarce or loose can be taken as a check of the method's sampling efficiency.

All eight structures arising from the high-temperature RMD simulation are considered acceptable on the basis of their corresponding values for internal energy and for the energetic term corresponding to distance violations. All of them are around the corresponding values of the initial structure. A comparison of a weighted average of all these structures (taking as weights the inverses of the corresponding total energetic terms) with the crystal structure is given in Fig. 10. Inspection of that figure shows that the global shape and overall fold of the solution-averaged struc-
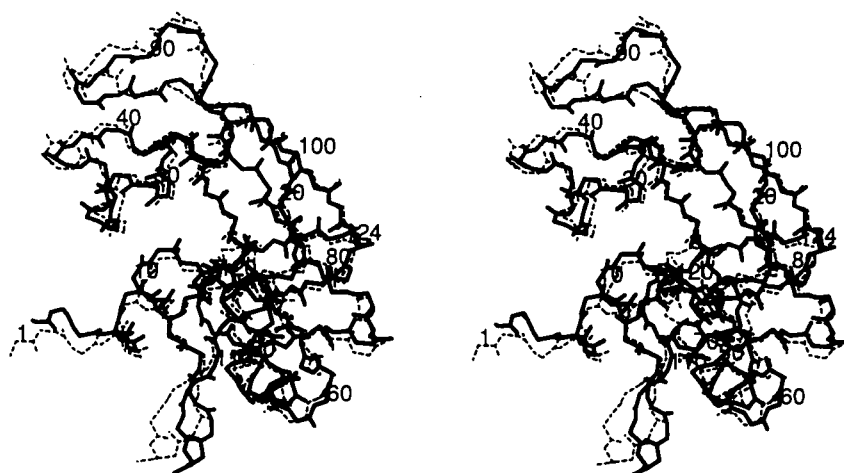

Fig. 10. Stereoscopic view of the superposition of two global conformations of RNase A: (a) solution averaged structure of the eight final refined structures (continuous line); (b) X-ray diffraction structure in the crystal state (dotted line).

TABLE 5

RMS DIFFERENCES (Å) BETWEEN THE EIGHT FINAL STRUCTURES AND THE CRYSTAL STRUCTURE (XR) FOR ALL BACKBONE ATOM POSITIONS

Final structure	1	2	3	4	5	6	7	8	XR
1		2.6	1.4	2.1	1.6	1.8	2.3	1.8	1.6
2			2.3	1.9	2.9	2.1	2.0	2.2	2.1
3				1.9	1.8	1.2	2.0	1.3	1.6
4					2.2	1.8	1.9	1.7	1.8
5						2.1	2.2	2.0	1.8
6							1.8	1.1	1.7
7								1.9	2.2
8									1.6

ture closely matches the one in the crystal state. In fact, the rms deviations of backbone atom positions is only 1.8 Å (Table 5).

Largest deviations of backbone torsional angles between the averaged solution structure and that corresponding to the crystal state are found in loop regions 17–24 (Fig. 5), 31–40 (Fig. 6), 59–72 (Fig. 8) and 86–97 (Fig. 9). Close inspection of local distance-constraint violations of either families in the solution structures or an EM relaxed crystal structure show low and comparable values, thus indicating that all of them are equally compatible with the observed basis of NOE-based distance constraints.

A global topological difference is observed (Fig. 11) in the orientation of the loop 109–117 with respect to the N-terminal α-helix and the loop 59–74. No major differences are found, however, in the torsion angles of this segment between the solution and crystal structures. A mere change of
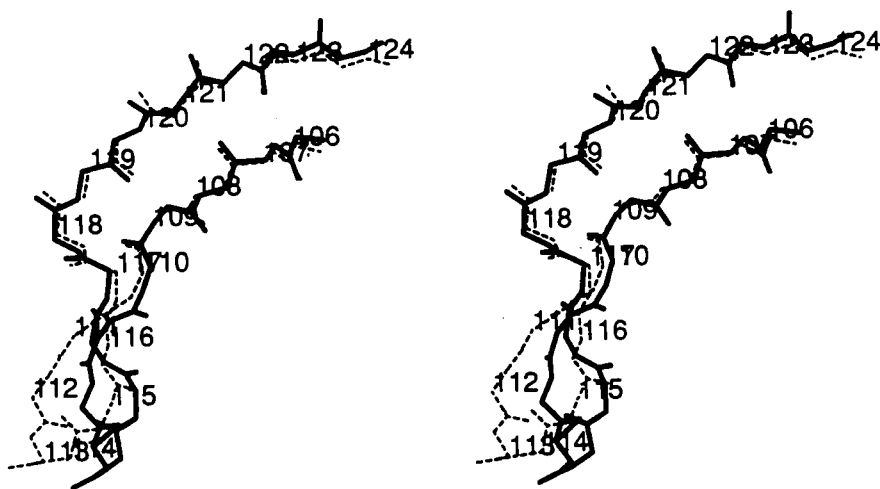

Fig. 11. Stereoscopic view of the superposition of two conformations of the loop $Ala^{109}-Val^{124}$ in RNase A: (a) solution averaged structure of the eight final refined structures (continuous line); (b) X-ray diffraction structure in the crystal state (dotted line).
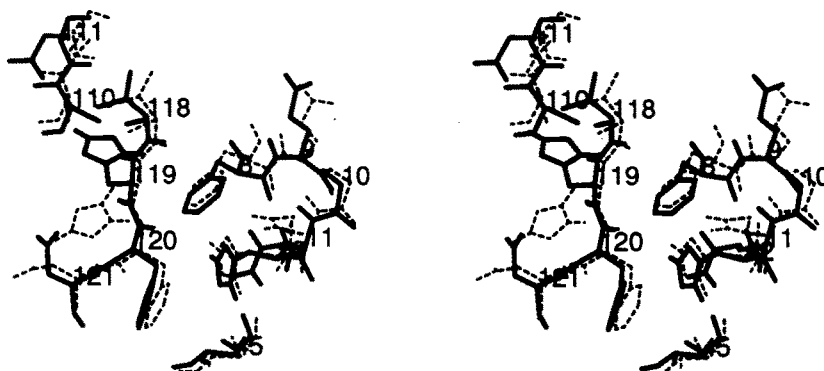
Fig. 12. Stereoscopic view of the solution (thick line) and crystal (dotted line) conformations of side chains in the active center, showing the alternative positionings of Val[118] and His[119] side chains.

$30°$ in the φ torsional angle of Glu[111] is responsible for this effect. More NOEs in the vicinity of the cross-point of the two strands (Fig. 11) are needed to determine whether this difference is meaningful.

Well-defined positions for side chains (χ_1 in the same staggered conformational region for at least six of eight structures) are found for: valines 43, 47, 54, 57, 63, 116 and 118; the three isoleucines; threonines 3, 36, 45, 70, 78, 82, 87 and 99; the three phenylalanines; aspartics 38 and 53; tyrosines 25, 73, 92, 97 and 115; cysteines 26, 58, 72 and 84; histidines 12, 48 and 105; glutamines 11, 28, 60, 74 and 101; glutamics 9, 49 and 111; asparagines 24 and 62; leucine 51; lysines 7, 66, 91, 98 and 104 and none of the four methionines. They add up to 53 residues against 52 with an ill-defined position for the side chain. Ser, Asn and Met residues are the more abundant among the latter. Side-chain locations for 40 of the 53 well-defined side-chain residues are the ones found in the crystal state. Further assignments of proton resonances and NOE cross-correlations are needed to make a definite statement on locked or flexible side chains. Conflicting NOE-based distance constraints are found for the methyl signal in methionines 13 and 79 (highly flexible side chain), H2 and H4 proton resonances in His[105] (attributable to free ring flipping) and the very special case of the methyl signals of Val[118] and H4 of His[119] which will be commented on in some detail due to its participation in the active center.

Under the conditions where the present NOE data were obtained (pH 4.0, 35°C, phosphate-bound enzyme) a strong NOE is observed between one of the methyl signals of Val[118] and H2 of His[119], which definitely is not predicted by the crystal structure, thus suggesting a different positioning of these two side chains in solution. The free enzyme at slightly higher pH shows, in addition to the previous NOE, a new correlation between H4 of His[119] and the $H_\beta H_{\beta'}$ resonances of Asp[121]. There is no plausible location of the His[119] side chain able to account for the two observed NOE effects. Most likely, a dynamic equilibrium is taking place between two allowed positions for the His[119] side chain, one analogous to that in the crystal state which would be responsible for the NOE with Asp[121] $H_\beta H_{\beta'}$ and a second one, similar to an alternative position found by Borkakoti et al. (1982) in their X-ray study, which would account for the NOE with Val[118].

The dynamic equilibrium is affected by pH and phosphate binding. High pH and absence of phosphate favour the major conformer observed in the crystal study. Under the conditions of our

study, the alternative positions of Val[118] and His[119] side chains prevails. Figure 12 shows a stereoscopic view of the solution and crystal structures of the active center, illustrating the two alternative positions of these side chains. The locations of other side chains thought to participate in substrate binding and catalysis, such as Phe[8], His[12], Thr[45] and Phe[120], are well-defined and coincide closely with their positioning in the crystal state.

In summary, a method is proposed to generate initial substructures susceptible to be refined by well-established RMD methods. A set of eight structures has been obtained for RNase A in solution, compatible with the 467 NOE-based distance-constraint set, which show an average rms deviation of 2 Å. The solution structures show a close resemblance to that corresponding to the crystal state. A meaningful difference between the solution and crystal structures in the positioning of the catalytically important His[119] side chain has been spotted.

ACKNOWLEDGEMENTS

REFERENCES

Borkakoti, N., Moss, D.S. and Palmer, R.A. (1982) *Acta Crystallogr.*, **B38**, 2210-2217.
Braun, W. (1987) *Q. Rev. Biophys.*, **19**, 115-157.
Braun, W. and Gō, N. (1985) *J. Mol. Biol.*, **186**, 611-626.
Havel, T.F., Kuntz, I.W. and Crippen, G.M. (1983) *Bull. Math. Biol.*, **45**, 665-720.
Kaptein, R., Boelens, R., Scheek, M.R. and van Gunsteren, W.F. (1988) *Biochemistry*, **27**, 5289-5395.
Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317-324.
Pohl, F.M. (1980) In *Protein Folding* (Ed., Jaenicke, R.), Elsevier, Amsterdam, pp. 183-196.
Richardson, J. (1981) *Adv. Prot. Chem.*, **34**, 167-339.
Rico, M., Bruix, M., Santoro, J., González, C., Neira, J.L., Nieto, J.L. and Herranz, J. (1989) *Eur. J. Biochem.*, **183**, 623-638.
Robertson, A.D., Purisima, E.O., Eastman, M.A. and Scheraga, H.A. (1989) *Biochemistry*, **28**, 5930-5938.
Van Gunsteren, W.F. and Berendsen, H.J.C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, BIOMOS, Groningen, The Netherlands.
Vlieg, J., Scheek, R.M., van Gunsteren, W.F.; Berendsen, H.J.C., Kaptein, R. and Thomason, J. (1988) *Proteins*, **3**, 209-218.
Wlodawer, A., Svensson, L.A., Sjölin, L. and Gilliland, G.L. (1988) *Biochemistry*, **27**, 2705-2717.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.