Journal of Biomolecular NMR, 1 (1991) 71–82
ESCOM

J-Bio NMR 007

The structure of ColE1 rop in solution

W. Eberle[a], A. Pastore[a], C. Sander[a] and P. Rösch[b,*]

[a]*European Molecular Biology Laboratory, Meyerhofstr. 1, D-6900 Heidelberg 1, Germany*
[b]*Max-Planck-Institute for Medical Research, Department of Biophysics, Jahnstr. 29, D-6900 Heidelberg 1, Germany*

Received 6 December 1990
Accepted 29 January 1991

Keywords: rop; Protein conformation; 2D NMR; Molecular dynamics; Repressor proteins

SUMMARY

The structure of the ColE1 repressor of primer (rop) protein in solution was determined from the proton nuclear magnetic resonance data by a combined use of distance geometry and restrained molecular dynamics calculations. A set of structures was determined with low internal energy and virtually no violations of the experimental distance restraints. Rop forms homodimers: Two helical hairpins are arranged as an antiparallel four helix bundle with a left-handed rope-like twist of the helix axes and with left-handed bundle topology. The very compact packing of the side chains in the helix interfaces of the rop coiled-coil structure may well account for its high stability. Overall, the solution structure is highly similar to the recently determined X-ray structure (Banner, D.W., Kokkinidis, M. and Tsernoglou, D. (1987) *J. Mol. Biol.*, **196**, 657–675), although there are minor differences in regions where packing forces appear to influence the crystal structure.

INTRODUCTION

It is generally accepted that protein structures can tolerate even large changes in the amino acid sequence while maintaining the overall fold and specific function. Although numerous examples supporting this hypothesis are known (Lesk and Chothia, 1980), a general theory capable of predicting the effects of sequence changes on structure and function is still not available. The combined use of structural techniques such as X-ray crystallography and proton nuclear magnetic resonance (^{1}H NMR) on the one hand and site-directed mutagenesis and peptide synthesis on the other hand allows the direct test of structure predictions (Markley et al., 1986; Eijsink et al., 1990; Schlichting et al., 1990).

Abbreviations: rop, repressor of primer; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; RAN Set, Structures generated from random choice of the dihedral angles; HEL Set, Structures generated from random choice of the dihedral angles restricted to ranges allowed for helices; MD, molecular dynamics; EM, energy minimization; RMSD, root-mean-square deviation of atomic positions.

* To whom correspondence should be addressed at University of Bayreuth, Department of Structure and Chemistry of Biopolymers, P.O. Box 10 12 51, D-8580 Bayreuth, Germany.

0925-2738/$ 5.00 © 1991 ESCOM Science Publishers B.V.

One of the smaller proteins for which a great number of mutants are available is the ColE1 repressor of primer (rop) (Twigg and Sherrat, 1980). In vivo, rop controls the replication of plasmid pBR322 by regulation of RNA-RNA interactions (Cesareni and Banner, 1985; Helmer-Citterich et al., 1988; Castagnoli et al., 1989). Rop is a small protein of 63 amino acids. It is known to form homodimeric four helix bundles both in the crystal according to crystallographic X-ray analysis (Banner et al., 1987) as well as in solution according to ^{1}H NMR studies (Eberle et al., 1990). This four helix bundle structural motif is present in other proteins and combines very high stability with compact packing (Cohen and Parry, 1986; Weber and Salemme, 1980; Ho and DeGrado, 1987). Because of the small size and the high temperature and pH stability, rop represents an ideal candidate for solution structural studies of protein folding. Indeed, rop is currently being used as a suitable scaffold to design proteins with functional units, e.g. metal binding and antigenic binding proteins (Sander, 1987; Emery et al., 1990). Laying the groundwork for further studies of rop mutants was one of the goals we had in mind when we started the structure determination of the wildtype rop protein in solution.

The sequence-specific resonance assignments of nearly all peaks in the ^{1}H NMR spectrum and the determination of the secondary structure of the protein were presented earlier (Eberle et al., 1990). Here we present the determination of the three-dimensional structure of rop in solution from the NMR data by a combined use of distance geometry and restrained molecular dynamics calculations.

MATERIAL AND METHODS

Sample preparation

The protein was purified by the standard procedure (Banner et al., 1983). This procedure yielded protein at roughly pH 7. The protein resulting from the preparation was lyophilized and subsequently redissolved in 1 M phosphate buffer, pH 2.3. It was then dialyzed against 5 mM phosphate buffer, pH 2.3. The protein was lyophilized and redissolved either in D_2O or in 90% H_2O/10% D_2O. The samples in D_2O were again dialyzed and redissolved in 99.96% D_2O ('100%' from Sigma). This procedure resulted in a final protein concentration of approximately 10 mM in the NMR sample. NMR measurements were usually performed at 308 K at resonance frequencies of 500 and 600 MHz. Details of the experimental procedures have been published elsewhere (Eberle et al., 1990).

Calculation procedures

Distance restraints were obtained from nuclear Overhauser enhancement spectroscopy (NOESY) data acquired with 75 ms mixing time. Cross-peak intensities were classified into three different categories according to a calibration against the peak intensity of the δ and ε protons of the aromatic rings (up to 3.0 Å, 3.6 Å, and 4.5 Å for strong, medium and weak nuclear Overhauser enhancement (NOE) intensities; Wüthrich, 1986). The usual corrections for pseudo-atoms (Wüthrich et al., 1983) were added where no stereo-specific assignment was possible (Eberle et al., 1990). An additional correction of 0.5 Å was added to distances involving methyl groups to take into account the multiplicity. In this way 397 upper bounds were obtained for the monomer and 30 additional distance bounds between the two monomers.

We compared this calibration approach with two other methods, namely comparison of peak

volumes instead of peak heights and the calibration based on the $\beta-\beta'$ cross-peak heights. The calibration relying on the $\beta-\beta'$ cross-peak intensities led to strong inconsistencies in the distance patterns. The procedure using cross-peak volume calculation as performed with the AURELIA program package (P. Neidig, Bruker) did not yield results significantly different from the peak height calibration procedure.

Information about intra-residue distances was incorporated into the calculations only for some of the HN/Hβ pairs and indirectly as restraints on the dihedral angles. Hydrogen bonds were explicitly included only when NOE peaks between both the HN(i)-Hα (i−3) and HN(i)-Hα (i−4) could be observed. In addition, the deuterium exchange rate of these hydrogens was observed to be slow, as evidence for the presence of hydrogen bonds (Wagner et al., 1987). Each hydrogen bond was translated into two distance constraints: 1.8 Å $\leqslant d_{HN(i)-O(i-4)} \leqslant$ 2.2 Å and 2.7 Å $\leqslant d_{N(i)-O(i-4)} \leqslant$ 3.2 Å. Eleven hydrogen bonds were introduced involving the amide protons of residues 9, 13, 16, 27, 36, 37, 39, 41, 50, 54, and 56. A list of the distance restraints is shown in Table 1.

The structure of the monomer was calculated by the distance geometry approach using the DISMAN program (Braun and Go, 1985). The variable target function was changed so that the ranges of the distance restraints and van der Waals constraints were increased in steps of one residue for peptide segments of lengths from 1 to 6 and in steps of 4 afterwards. Up to 500 cycles of minimization were performed for each level. Ten starting structures were generated from random choices of the dihedral angles (RAN1-10 set). Ten more were obtained by choosing the ϕ and ψ angles of residues from 3 to 28 and from 32 to 56 randomly within the region allowed for helical conformation (-57 ± 30, -47 ± 30; HEL1-10 set). The best ten structures in terms of distance violations (five from the RAN set and five from the HEL set) were chosen for further refinement.

The starting structure of the dimer was then obtained by docking two monomers using the molecular graphics program QUANTA (Polygen Corp.) and refined by molecular dynamics calculations using the XPLOR package (Brünger, 1987). The standard XPLOR force field was used, with the exception of the force constants for the torsional potential, which was switched off for ϕ and ψ angles and set to 300 kcal mol^{-1} rad^{-1} for the ω angles.

After 50 cycles of unrestrained energy minimization, about 3 ps of MD at high temperature (600 K) were calculated basically using the heating stage of previously described protocols (Nilges et al., 1988) (Fig. 1). The temperature was then linearly decreased to 300 K during an interval of 4.8 ps. During these intervals, a simplified repulsive ('repel') function was used for both the electrostatic and van der Waals interactions. A quartic function was used as distance restraint poten-

TABLE I
NUMBER AND DISTRIBUTION OF DISTANCE RESTRAINTS

Number of NOE restraints	
total for the dimer	824
inter-molecular	30
total for the monomer	397
inter-residue	310
intra-residue	65
hydrogen bonds	2*11

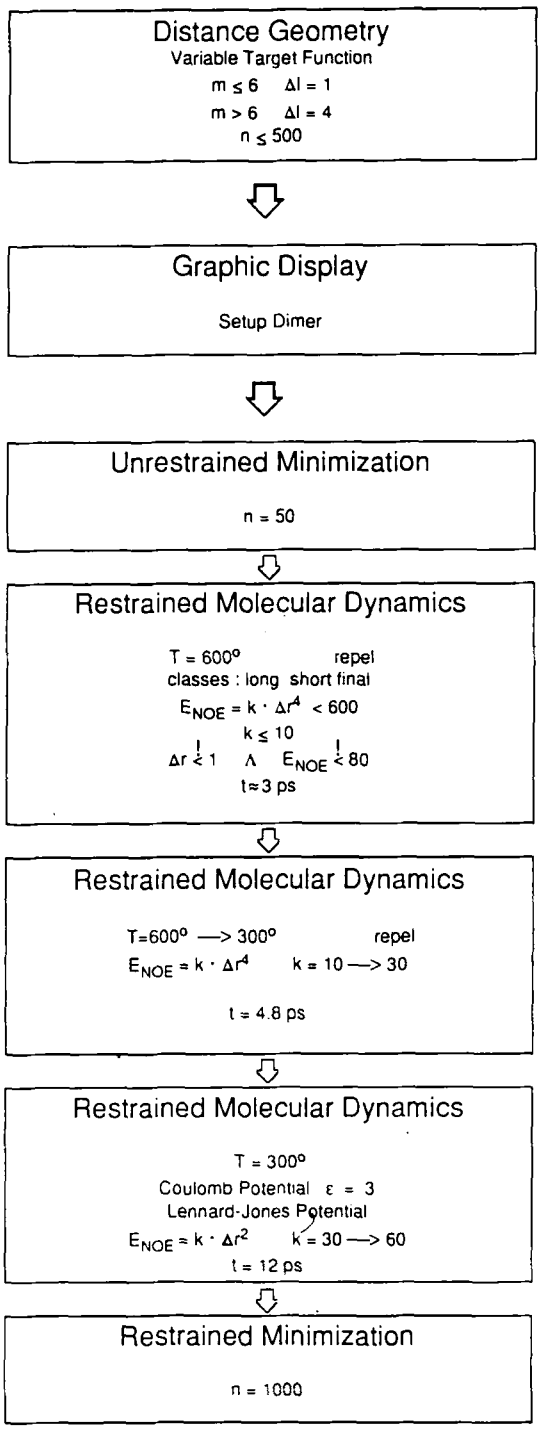


Fig. 1. Flowchart of the Distance Geometry/Molecular Dynamics protocol. The units of energies, lengths, temperature and time are in terms of kcal, Å, K and ps. The following symbols were used: m is the window length of the DISMAN variable target function; Δl is the increment of m; n is the number of minimization steps; E_{NOE} is the pseudopotential energy term taking into account the NOE restraints; repel stands for a step function simplifying the description of the van der Waals interactions.

tial function. At thermal equilibrium conditions, 12 ps were calculated changing the potential function for the electrostatic and van der Waals forces to the Coulomb and Lennard-Jones potentials. A dielectric constant of 3 was used for the calculations. During this stage, a harmonic potential was used for distance restraining. Different increasing values of the force constants were chosen at different stages of the calculation, ending with a value of 60 kcal mol^{-1} $Å^{-2}$. Finally, 1000 cycles of energy minimization were run to locate the closest energy minimum.

The approach of combining structure calculation methods to get valid solution structures of proteins is now a fairly widely accepted procedure. Indeed, although XPLOR has proven to be a very flexible and highly efficient MD program, it still needs a suitable starting structure. We were not able to obtain convincing results with this program alone when starting from extended structures. In these cases, the folding and the secondary structure of the molecule were basically correct, but localized distortions in the helical parts were observed with large NOE violations (more than 1.0 Å). As the violations always occurred in different regions and no NOE violations occurred when starting from the crystallographic structure, this behaviour cannot be explained by inconsistency of the NMR data. Rather, it is caused by the function being trapped in local minima.

To overcome this problem, the combined use of the distance geometry algorithm and of XPLOR was necessary. Although DISMAN alone was not able to reach the global minimum of the target function, it could approach it very closely and generated suitable input structures for further refinement with the XPLOR program.

Root-mean-square deviations of atomic positions between structure pairs and solvent accessibility of atoms and residues were calculated according to Kabsch (1976) and Kabsch and Sander (1983).

The coordinates and the list of NOE restraints will be deposited in the Protein Data Library.

RESULTS AND DISCUSSION

Monomer

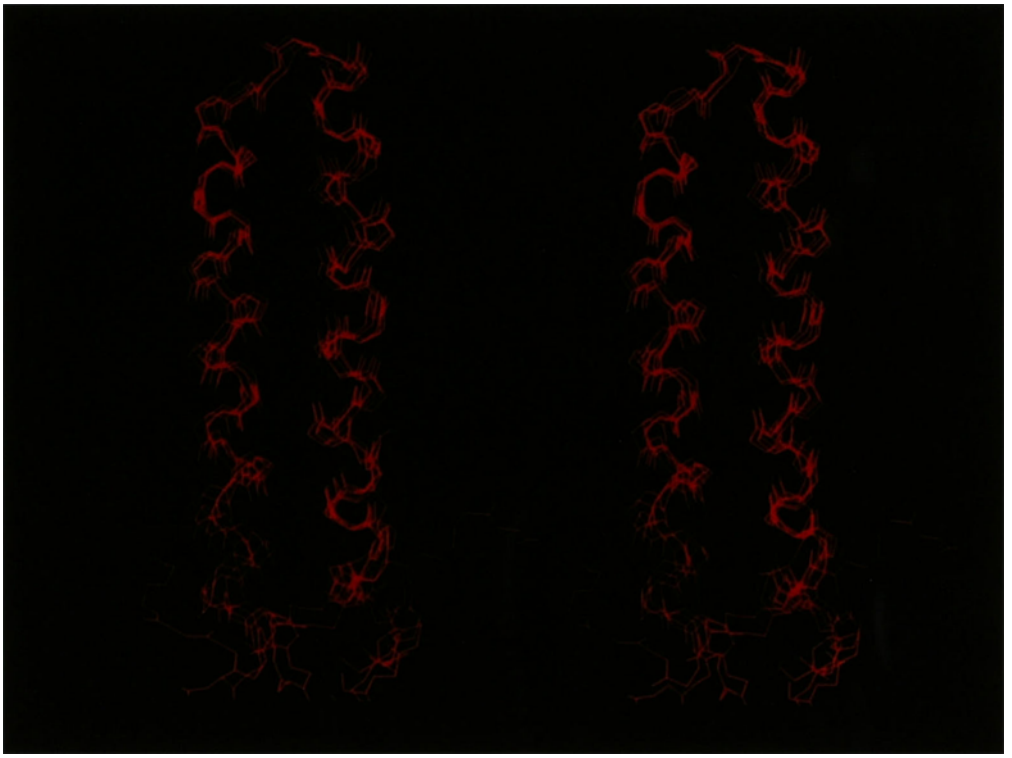
Ten final structures for the monomer were calculated. Two regions of α-helix span from residue 3 to 28 and 32 to 56. One 3_{10} helical turn is localised between residues 28 to 30. A tight hairpin loop extends from 30 and 32 with Ala^{31} in a left-handed helix conformation. The C-terminal region is in random conformation. The RMSD of the backbone atoms and of all atoms is shown in Table 2. Figures 2a and b show different representations of the results obtained for the monomer.

Dimer

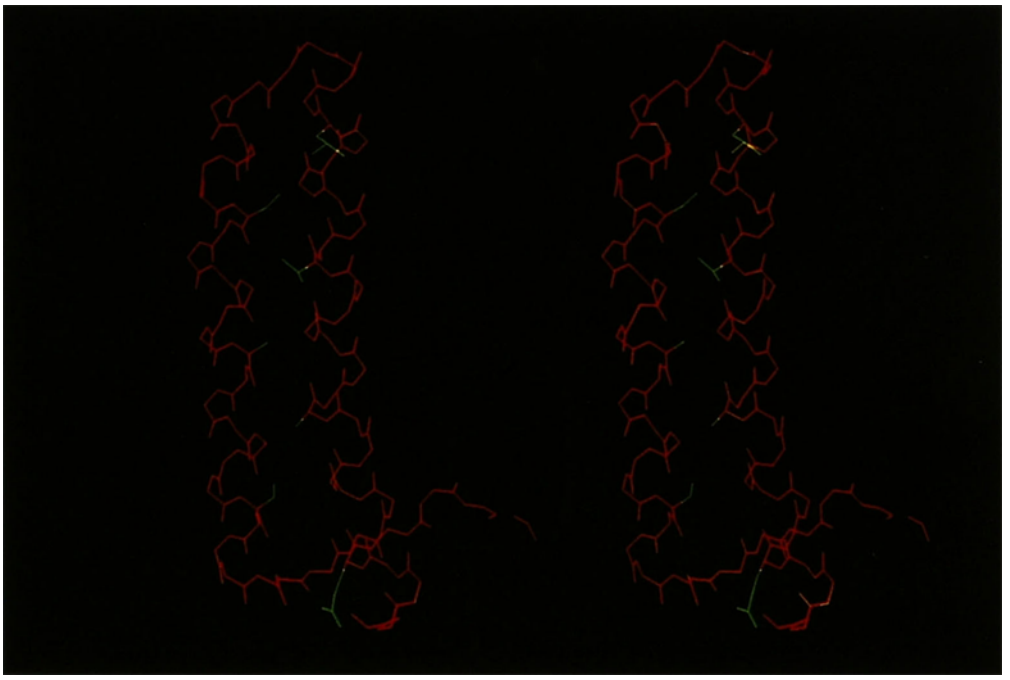
The dimer was first superficially modelled by docking two identical copies of the monomer with the QUANTA graphics program and subsequently refined by MD calculations.

Although there are at least two fundamentally different ways of arranging the monomers, with different handedness of the bundle, only one was consistent with the experimental NOE data. Thirty additional inter-molecular distance restraints located the relative position of the monomers. The dimer has left-handed bundle topology according to the convention of Presnell and Cohen (1989), generalized to dimeric bundles. As shown previously (Banner et al., 1987), the four helices in rop wind around each other as a so-called coiled-coil. The superhelix parameters for the solution structure are similar to those obtained for the crystallographic structure (pitch,

a



b



c

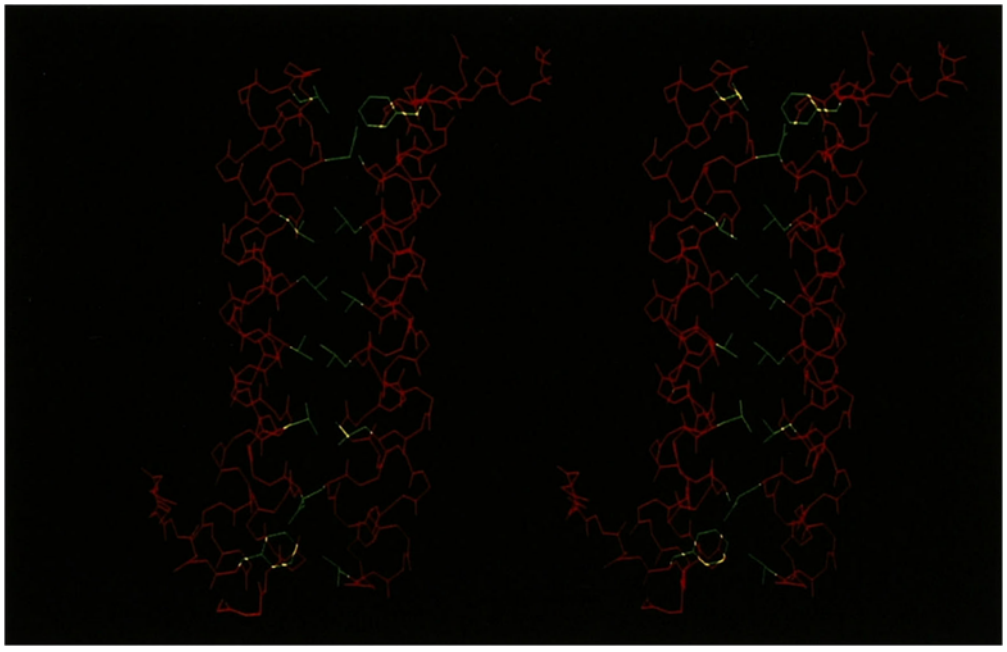


Fig. 2. Stereo views of the representative ColE1 rop structures. (a) Superposition of 5 monomer structures from the HEL set. The backbone atoms from residues Gln^4 to Arg^{50} were used in the fitting procedure. (b) Helix-helix packing within the monomer (structure HEL1). The backbone atoms and the side chains of the residues occurring at the heptad position 'a' (according to Banner et al., 1987) are shown. (c) Helix-helix packing at the dimer interface (structure HEL1). The backbone atoms and the side chains of the residues at heptad position 'd' are shown.

172.5 Å; radius, 7.0 Å and radius of curvature, 114.5 Å). A slight curvature of the helix axes previously observed in the monomer disappears during the generation of the dimer: The additional interaction between the two hairpins contributes to regularization of the structure.

The twofold symmetry of the dimer did not persist during the refinement procedure, since no

TABLE 2
COMPARISON BETWEEN STRUCTURE PAIRS: MEAN VALUES AND STANDARD DEVIATIONS OF THE RMSD (Å)

Structures	RMSD	RMSD
	backbone	all heavy atoms
DISMAN vs DISMAN (monomer)	2.08(76)	3.21(65)
NMR vs NMR (monomer)	0.92(19)	1.89(20)
NMR vs NMR (dimer)	1.14(21)	2.01(21)
NMR vs NMR (symmetry)[a]	0.81(15)	1.76(23)
NMR vs X-ray (monomer)	1.20(13)	2.33(16)
NMR vs X-ray (dimer)	1.35(12)	2.41(11)

The backbone atoms of both monomers were used in the comparison. To eliminate effects of the chain ends only residues Gln^4 to Arg^{50} were considered. The standard deviations are indicated in parentheses.

[a] Comparison of the monomers of the same structure.

TABLE 3
VIOLATIONS OF THE FINAL STRUCTURES

	average	range
violation/restraint (Å)	0.005	0.004–0.008
number of violations	46	33–65
maximum violation (Å)	0.060	0.041–0.095

symmetry was imposed on the two monomers during the MD calculation. The average RMSD between monomers of the same molecule was 0.82 Å for the backbone atoms, which compares well with the average RMSD between the monomer structures of different structures (0.89 Å). A slightly smaller RMSD was obtained for the HEL family over the other family, the main deviations occurring in the C-terminal region. The mean RMSD obtained after MD refinement was smaller than the one after distance geometry calculations only. In fact, in the dimer, the additional NMR restraints reduce the possibility of distortions in the helices. The list of the violations observed in the final structures is shown in Table 3. The final average violation (0.005 Å) is significantly lower than the one obtained from the structures after the DISMAN stage (0.38 Å).

Table 4 shows the average energy contributions and standard deviations for the final structures together with the deviations of bond length, bond angles and dihedral angles from standard geometry. Figure 2c shows a stereo view of the dimer with residue packing emphasized.

After the DISMAN calculation, the ϕ and ψ angles cluster mainly in the allowed region in the Ramachandran plot, although they are spread in a relatively large area (Fig. 3a). The MD refinement regularises the structure and reduces the spread of the backbone dihedrals (Fig. 3b). After refinement, the conformation of the hairpin is uniquely defined. Figure 4 shows a plot of the ϕ angles averaged over the twenty monomers in the final structures vs. the sequence number and the relative standard deviations. Very good agreement is found along the chain with the only two exceptions around the loop and around residues 49 to 52. The presence of two local conformations around Cys^{52} (Eberle et al., 1990) in slow exchange on the NMR time scale may be related to the greater variability of this region. Larger deviations are observed in the terminal regions.

Given the very large number of the NMR distance restraints and their distribution, the RMSD among the NMR structures should be mainly related to the effective motion in solution. Indeed, the residues involved in helix packing show higher rigidity (Banner et al., 1987) (Fig. 5, see also Figs. 2b and c, e.g. Ala^{8} or Ala^{12}).

TABLE 4
DEVIATIONS FROM STANDARD GEOMETRY AND ENERGIES

Deviations from idealized geometry:		Energies (kcal/mol)			
bond length (Å)	0.007(0)	total	vdW	Coulomb	dihedral
bond angles (°)	2.29(3)				
torsion angles (°)	0.18(1)	−2620(55)	−937(13)	−2069(35)	99(8)

We indicate the mean value taken over all NMR structures and, in parentheses, the standard deviation from the mean.

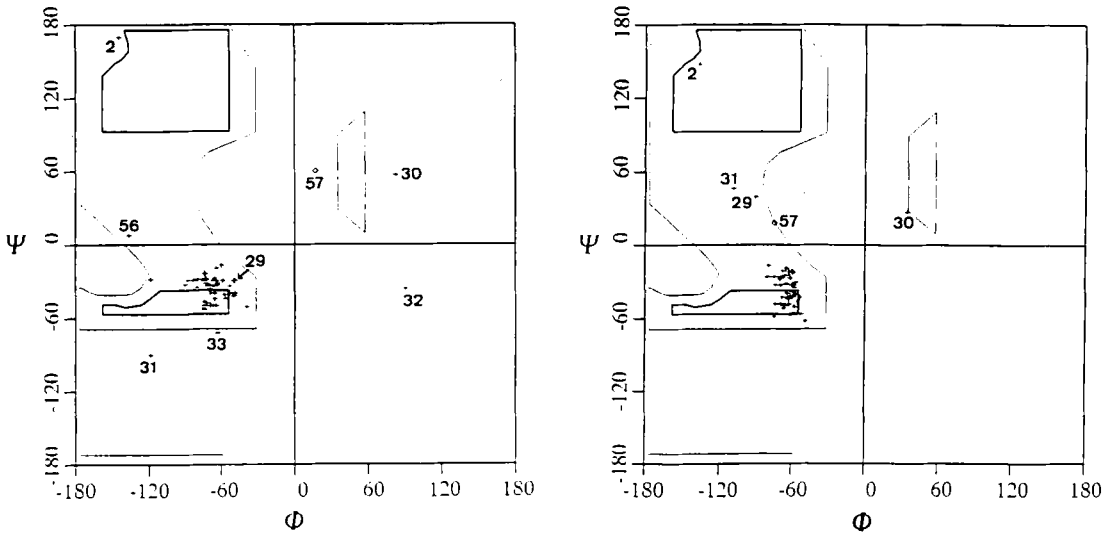


Fig. 3. Ramachandran plots of the RANI structure (one monomer): (a) after DISMAN calculations; (b) after EM and MD refinement.

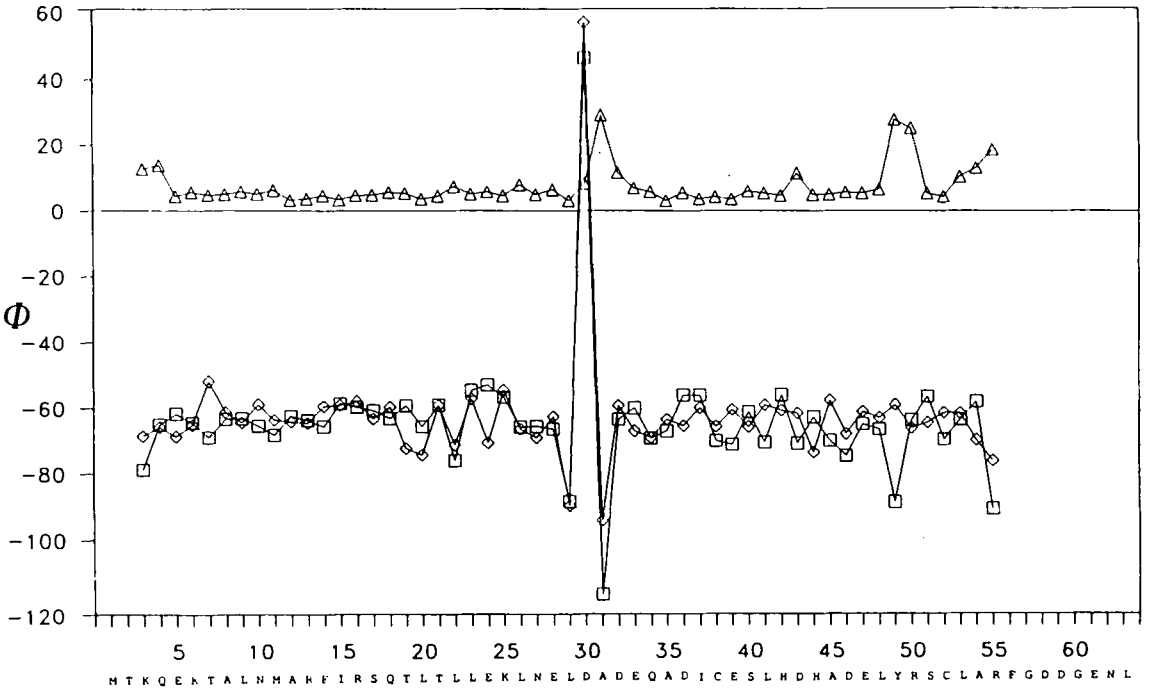


Fig. 4. Plot of the ϕ angles of the X-ray structure (◇) and the average values of the NMR structures (□) with their standard deviations (△). The amino acid sequence is shown at the bottom.

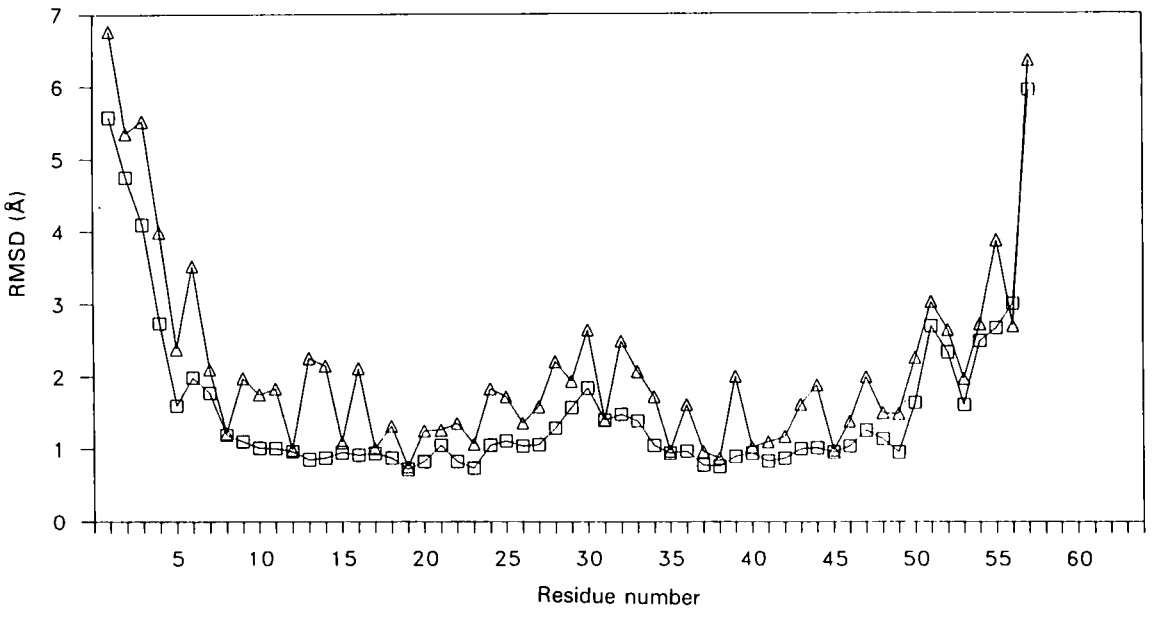


Fig. 5. Average RMSD between NMR structures vs. residue number. Two different symbols are used for the average over backbone atoms (□) and heavy atoms (△). The backbone atoms from residues Gln^4 to Arg^{50} of both monomers were used in the comparison.

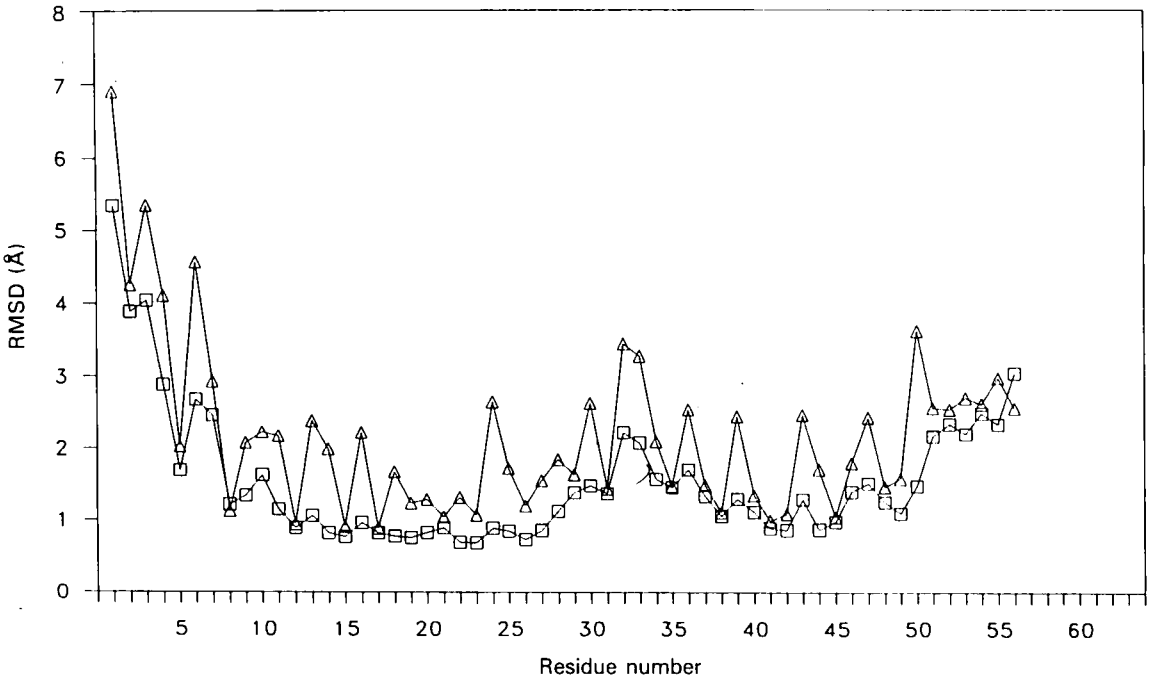


Fig. 6. Average RMSD between X-ray and NMR structures vs. residue number. Two different symbols are used for the average over backbone atoms (□) and heavy atoms (△). The backbone atoms from residues Gln^4 to Arg^{50} of both monomers were used in the fitting.

Visual comparison between the structure in solution and the crystallographic structure (Banner et al., 1987) shows no major differences. The averaged RMSD between NMR and X-ray structures is only slightly higher than the corresponding value for the NMR structures compared to each other (Table 4). Larger differences are found in the loop region, between residues Asp^{32} and Asp^{36}, and around Thr^{7}. Interestingly, Asp^{32}, Asp^{36} and Thr^{7} are among the residues involved in crystal packing as shown by Banner et al., (1987).

A local comparison of the structures can be obtained from Fig. 6, where the average RMSD values between the X-ray structure and the solution structures for each residue are shown. Good agreement along the whole chain is found. The structure of the C-terminal region (residue Gly^{57} to C-terminus) is completely disordered in solution, just as it was found to be in the crystal (Banner et al., 1987).

Crystallographic temperature factors (B-factors) are a measure of the thermal flexibility of molecules in the crystal., We computed the variance of the atom positions averaged over all atoms of each residue under the assumption of isotropic fluctuations from the square roots of the B-factors (Petsko and Ringe, 1984) (Fig. 7). Comparison of the RMSD of the NMR structures and the thermal motions of the protein in the crystal clearly suggests a strong correlation between these two observables (Figs. 5 and 7).

However, in spite of the close similarity of the two conformations, the crystallographic structure does not account for the complete set of NOE restraints observed, as the average NOE violation as calculated from the X-ray structure is 0.29 Å.

In conclusion, we have determined the structure of rop in solution by distance geometry and restrained molecular dynamics. This approach led to structures with low internal energy and almost no distance violations. From the RMSD within the family of solutions equally consistent with the NMR restraints and from a close comparison with the X-ray structure, we can infer a very stable helix-helix packing in the protein interior. The work presented here represents the starting point of a long-term project to study protein stability and mutant predictions.

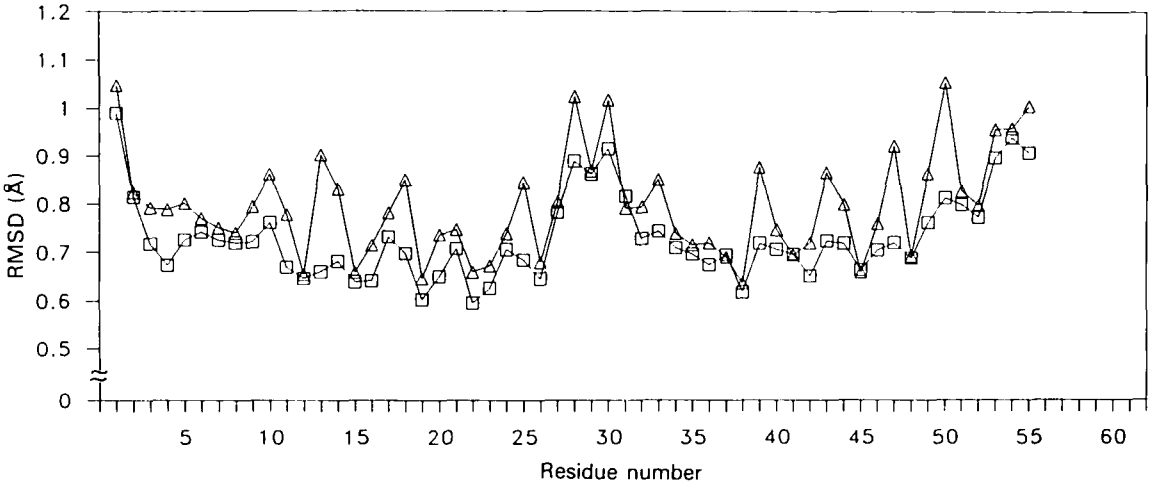


Fig. 7. Standard deviation of atom positions computed from the crystallographic B-factors, averaged over the backbone atoms (▽) and all of the heavy atoms of each residue (□).

REFERENCES

Banner, D.W., Cesareni, G. and Tsernoglou, D. (1983) *J. Mol. Biol.*, **170**, 1059-1060.
Banner, D.W., Kokkinidis, M. and Tsernoglou, D. (1987) *J. Mol. Biol.*, **196**, 657-675.
Braun, W. and Go, N. (1985) *J. Mol. Biol.*, **186**, 611-626.
Brünger, A.T. (1987) In *'Methods and Applications in Crystallographic Computing'* (Ed, Isaacs, N.), Oxford Press, Oxford.
Castagnoli, L., Scarpa, M., Kokkinidis, M., Banner, D.W., Tsernoglou, D. and Cesareni, G. (1989) *EMBO J.*, **8**, 621-629.
Cesareni, G. and Banner, D.W. (1985) *Trends Biochem. Sci.*, **10**, 303-306.
Chotia, C. and Lesk, A. (1986) *EMBO J.*, **5**, 823-826.
Cohen, C. and Parry, D.A.D. (1986) *TIBS*, **11**, 245–248.
Eberle, W., Klaus, W., Cesareni, G., Sander, C. and Rösch, P. (1990) *Biochemistry*, **29**, 7402–7407.
Emery, S.C., Blöcker, H., Cesareni, G., Kingswell, A. and Sander, C. (1990) *Protein Eng.*, in press.
Eijsink, V.G.H., Vriend, G., van den Burg, B., Venema, G. and Stulp, B.K. (1990) *Protein Eng.*, in press.
Ho, S.P. and DeGrado, W.F. (1987) *J. Am. Chem. Soc.*, **109**, 6751–6758.
Kabsch, W. (1976) *Acta Cryst.*, **A32**, 922.
Kabsch, W. and Sander, C. (1983) *Biopolymers*, **22**, 2577–2637.
Lesk, A.M. and Chothia, C. (1980) *J. Mol. Biol.*, **136**, 225–270.
Markley, J.L., Croll, D.H., Krishnamoorthi, R., Ortiz-Polo, G., Westler, W.M., Bogard, W.C., Jr. and Laskowski, M., Jr. (1986) *J. Cell. Biochem.*, **30**, 291–309.
Nilges, M., Gronenborn, A.M., Brünger, A.G. and Clore, M.G. (1988) *Protein Eng.*, **2**, 27–38.
Petsko, G. and Ringe, D. (1984) *Ann. Rev. Bioeng.*, **13**, 331–371.
Presnell, S.R. and Cohen, F.E. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6592–6596.
Sander, C. (Ed) (1987) *EMBL BIOcomputing Technical Document 1.*
Schlichting, I., John, J., Frech, M., Chardin, P., Wittinghofer, P., Zimmermann, H. and Roesch, P. (1990) *Biochemistry*, **29**, 504–511.
Twigg, A.J. and Sherratt, D.J. (1980) *Nature*, **283**, 216–218.
Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Go, N. and Wüthrich, K. (1987) *J. Mol. Biol.*, **196**, 611–639.
Weber, P.C. and Salemme, F.R. (1980) *Nature*, **287**, 82–84.
Wüthrich, K., Billeter, M. and Braun, W. (1983) *J. Mol. Biol.*, **169**, 946–961.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.