J. Mol. Biol. (1994) 236, 328–341

Nuclear Magnetic Resonance Solution Structure of the Arc Repressor Using Relaxation Matrix Calculations

Alexandre M. J. J. Bonvin, Hans Vis, Jan N. Breg, Maurits J. M. Burgering Rolf Boelens and Robert Kaptein†

Bijvoet Center for Biomolecular Research, Utrecht University Padualaan 8, 3584 CH Utrecht, The Netherlands

The Arc repressor of *Salmonella* bacteriophage P22 is a dimeric sequence-specific DNA-binding protein. The solution structure of Arc has been determined from 2D NMR data using an "ensemble" iterative relaxation matrix approach (IRMA) followed by direct NOE refinement with DINOSAUR. A set of 51 structures was generated with distance geometry and further refined with a combination of restrained energy minimization and restrained molecular dynamics in a parallel refinement protocol. Distance constraints were obtained from an extensive set of NOE build-ups in H_2O and ^2H_2O *via* relaxation matrix calculations from the ensemble of structures. Methyl group rotation, aromatic ring flips and internal mobility effects (*via* order parameters obtained from a free molecular dynamics run in water) were included in these calculations. The best structures were finally refined with direct NOE constraints following a slow-cooling simulated annealing protocol. In this final refinement stage, theoretical NOE intensities were directly compared with the experimental data and forces were derived using a simple two-spin approximation for the gradient of the NOE function. Dynamic assignment was applied to the peaks involving unassigned diastereotopic groups. The structure is determined to a precision (r.m.s.d. from the average excluding the ill defined C and N-terminal region) of 0·55 and 1·10 Å for backbone and all atoms, respectively. The final structures, with R factor values around 0·35, have good stereochemical qualities, contain an extensive network of hydrogen bonds consistent with the secondary structure elements and structural features in concordance with genetic data. The overall folding of the solution and crystal structures is the same.

Keywords: Arc repressor; relaxation matrix refinement; ensemble IRMA; DINOSAUR; internal motions

1. Introduction

Arc repressor is a dimeric sequence-specific, DNA-binding protein that is involved in the switch between lysis and lysogeny of *Salmonella* bacteriophage P22 (Susskind. 1980; Sauer *et al.*, 1983; Vershon *et al.*, 1985; Pabo & Sauer, 1984; Bowie & Sauer, 1989). The protein contains 53 amino acid residues per monomer and shows approximately 40% homology with the Mnt repressor of bacteriophage P22 with which it controls the expression of a third regulatory protein, the antirepressor. The Arc and Mnt repressors belong to a family of β-sheet DNA-binding proteins including the *Escherichia coli* Met J repressor (Rafferty *et al.*, 1989) and the Tra Y proteins of F and related episomes (Bowie & Sauer, 1990).

Previous NMR studies (Breg *et al.*, 1990) resulted in an initial solution structure for the Arc repressor that reveals a strongly intertwined dimer with an antiparallel β-sheet, the DNA-binding motif, formed by residues of both monomers. In this paper we present the refined solution structure of the Arc repressor obtained from 2D‡ NMR data using relaxation matrix calculations. In the first stage, IRMA (iterative relaxation matrix approach) was used to obtain accurate distances. These are used as constraints to generate and refine an ensemble of

† Author to whom all correspondence should be addressed.

‡ Abbreviations used: NMR: nuclear magnetic resonance; 2D, two-dimensional; NOE, nuclear Overhauser effect; IRMA, iterative relaxation matrix approach; DG, distance geometry; DDD, distance bounds driven dynamics; MD, molecular dynamics; EM, energy minimization; SA, simulated anncaling; r.m.s.d., root-mean-squared deviation; p.p.m., parts per million.

0022–2836/94/060328–14 $08.00/0

structures in a parallel protocol (Bonvin *et al.*, 1993a) using distance geometry (DG) and restrained molecular dynamics (MD) and energy minimization (EM) methods. Then, the best structures are directly refined against experimental NOE data (Yip & Case, 1989) following a slow cooling simulated annealing using DINOSAUR (Bonvin *et al.*, 1991, 1993*b*). Information on fast internal motions in the Arc repressor is obtained from a free MD simulation in water and included in the relaxation matrix calculations *via* order parameters, together with the effects of methyl group rotation and aromatic ring flips. The solution structure of Arc is analysed in detail and compared with the results of mutagenesis experiments (Bowie & Sauer, 1989; Vershon *et al.*, 1986) and with the crystal structure (C. R. Kissinger & C. O. Pabo, personal communication).

2. Materials and Methods

(a) *Experimental NMR data*

In all, 824 NOE constraints were obtained from NOESY spectra recorded at 600 MHz on a Bruker AM600 spectrometer in H_2O and ^2H_2O with mixing times of 100, 150, 200 and 50, 80, 120, 160, 200 ms, respectively (for a more complete description see Breg *et al.* (1989)). Additional constraints were obtained from a SCUBA experiment (Brown *et al.*, 1988) revealing NOEs close to the water frequency, and from 3-dimensional (^{15}N, ^1H) heteronuclear experiments (3D NOE-HMQC) (Burgering *et al.*, unpublished results). These were entered in the calculations as qualitative constraints with upper bound values of 2·8, 3·5 and 4·5 Å (before pseudo-atom correction) for strong, medium and weak NOEs, respectively, the peak classification being based on calibration distances. This resulted in a set of 130 additional constraints for the Arc repressor. The total set of 954 constraints consists of 350 intraresidue, 215 sequential, 181 medium-range and 208 long-range constraints; 144 of the total set correspond to intermonomer constraints. Twenty-four hydrogen bonds, each represented by 2 distance constraints of 2·3 and 3·3 Å between hydrogen and acceptor and donor and acceptor, respectively, were also explicitly defined. The acceptors were identified from the secondary structure elements using characteristic NOE patterns (e.g. i-$i+3$, i-$i+4$ NOEs for α-helices). The problem of the interpretation of intra- and intermonomer NOEs, which *a priori* cannot be distinguished for a symmetrical dimer, was initially solved by assuming structural homology with the *Escherichia coli* Met J repressor, for which a crystal structure was available (Rafferty *et al.*, 1989). Successive model building trials based on this hypothesis resulted in an unambiguous interpretation (Breg *et al.*, 1990). The assignments were recently confirmed by double half-filtered 2D NOE experiments on a mixture of unlabelled with (^{15}N, ^{13}C) labelled Arc repressor (Burgering *et al.*, 1993). The constraints were duplicated to account for the dimeric nature of Arc giving a total set of 2004 constraints (NOEs + hydrogen bonds). Of the set of 954 constraints per monomer, 824 NOEs with good quality build-up curves were used in the IRMA and DINOSAUR calculations, the others being introduced as qualitative constraints.

From a combination of J coupling data and NOE intensities 16 β-methylene proton pairs could be stereospecifically assigned, which represents 40% of the possible assignments. In addition, stereospecific assignments were obtained for the methyl groups of 4 valine residues out of 5. From the stereospecific assignment procedure constraints could be imposed on the χ_1 dihedral angles of these residues. In addition, the χ_1 dihedral angle of Val25, for which the methyl groups could not be stereospecifically assigned due to overlap, was constrained to the same value as for the other 4 valine residues, because of the large value of the $^3J_{H^\alpha H^\beta}$-coupling. This resulted in a total set of 21 χ_1 dihedral angle constraints/monomer with values of −60, 60 or 180°.

(b) *Relaxation matrix calculations*

Two types of relaxation matrix calculations were used for the Arc refinement. First, accurate distance constraints were obtained using the iterative relaxation matrix approach (IRMA: Boelens *et al.*, 1988, 1989; Koning 1990*a*). Then, in the final stage, the Arc structures were refined with the NOE intensities as direct structural constraints using DINOSAUR (Bonvin *et al.*, 1991, 1993*b*,*c*).

In IRMA, experimental NOE matrices, in which the missing information is supplemented by theoretical intensities computed from a model (single structure or ensemble of structures), are transformed back into relaxation matrices by standard matrix techniques. The relaxation matrices for the various NOE mixing times are averaged. Lower and upper bound distances are then derived from the average relaxation matrix elements plus or minus their standard deviations. An additional correction of ±10% of the distance was added to the upper and lower bounds. Pseudo-atoms were introduced on prochiral centres when no experimental stereospecific assignment was available and the bounds were relaxed accordingly (Wüthrich *et al.*, 1983). The IRMA calculations were performed for both data sets (H_2O and ^2H_2O) with inclusion of fast methyl rotation, aromatic ring flips and internal motions (*via* order parameters obtained from a free MD simulation in water (see below)) (Koning *et al.*, 1990*b*, 1991). The overall correlation time was first estimated from the size of the molecule and then optimized to 4·25 ns, with the R factors as criteria. An additional leakage rate of 2 s^{-1} corresponding to the inverse of the average proton T_1 of Arc (0·5 s) was added on the relaxation matrix diagonal. Alanine H^α-methyl, tyrosine H^δ-H^ε and methylene $H^{\beta 2}$-$H^{\beta 3}$ cross-peaks of residues in the α-helices were chosen as calibration peaks in IRMA. The exchangeable protons were excluded in the relaxation matrix calculations with ^2H_2O data. Distance constraints obtained from the H_2O and ^2H_2O data sets were combined and symmetrized for the dimer case.

In the DINOSAUR refinement, the experimental NOE intensities are directly used as structural constraint in a NOE restraining potential defined as (Bonvin *et al.*, 1991, 1993*c*):

$$V_{\mathrm{NOE}} = \left[\sum_i \delta_i w_i\right]^{-1} \sum_i \delta_i w_i (f A_i^{\mathrm{theo}} - A_i^{\mathrm{exp}})^2 \qquad (1)$$

with

$$w_i = \frac{1}{(N + \varepsilon A_i^{\mathrm{exp}})^2}. \qquad (2)$$

The sum runs over all NOE intensities A_i; N represents the experimental noise level and ε a relative error on the experimental NOE intensities accounting for integration errors; δ_i is a switching factor that is used to switch off the function when the theoretical NOE intensities are within the experimental errors defined as ($N + \varepsilon\ A_i^{\mathrm{exp}}$). The

scaling factor f is chosen such as to minimize the function V_{NOE} for a defined subset of peaks . Forces are calculated from eqn (1) with a simple two-spin approximation. For a more complete description see Bonvin *et al.* (1991, 1993*b*,*c*).

For the DINOSAUR refinement, theoretical NOE intensities were computed with the same parameters as in the IRMA calculations, using a spherical cut-off of 4·5 Å around each proton pair defining a NOE, which allows a considerable speed-up of the computations. With this approximation, the spin diffusion contributions of all neighbours within 2 spheres of 4·5 Å radius around the 2 protons defining a NOE peak are taken into account. For each NOE, a small relaxation matrix is calculated. The use of a spherical cut-off was found to have little effect on the calculated NOE intensities compared to the full relaxation matrix treatment. The scaling factors during the direct NOE refinement were calculated from all peaks, separately for each mixing time. An overall scaling could have been used as well since very little variation, of the order of 4% between the scaling factors for the various mixing times was found. This latter observation gives us confidence in the choice of the overall correlation time. Errors in the correlation time would result in a larger variation in the scaling factors. For future work, however, we recommend the use of an overall scaling, especially when build-up data recorded in one measuring session are available. The weighting factors in eqn (1) were calculated from the experimental noise levels (N) and a relative error ε of 5%. A dynamic assignment procedure (Bonvin *et al.*, 1993*b*,*c*) was applied to all NOEs involving diastereotopic protons that were not stereospecifically assigned and protons on aromatic rings that experience 180° ring flips. In this procedure, NOE intensities are calculated for all possible assignments and the program chooses the assignment giving the best correspondence with the experimental data. Note that this is not used as a stereospecific assignment procedure because the assignment is never fixed during the simulations. This procedure resembles the method proposed by Nilges & Brünger (1991) for the assignment of intra- or intermonomer NOEs in dimeric proteins.

(c) R *factors*

To assess the quality of the structures obtained from NMR data we directly compare calculated and experimental NOEs. The fit can be described by a residual index or R factor. We used 2 definitions; the simplest is similar to the X-ray definition (Gonzales *et al.*, 1991):

$$R = \frac{\sum_i \tau_m |f A_i^{theo} - A_i^{exp}|}{\sum_i \tau_m |A_i^{exp}|}. \quad (3)$$

The second definition was inspired by the $R^{1/6}$ definition proposed both by Gonzales *et al.* (1991) and Thomas *et al.* (1991), and the Q definition of Withka *et al.* (1992):

$$Q^{1/6} = \frac{\sum_i \tau_m |(f A_i^{theo})^{1/6} - (A_i^{exp})^{1/6}|}{\sum_i 0{\cdot}5\tau_m |(f A_i^{theo})^{1/6} + (A_i^{exp})^{1/6}|}. \quad (4)$$

The $Q^{1/6}$ factor is less sensitive to errors on short distances and takes into account the asymmetry of the NOE function. In both definitions the mixing times (τ_m) are introduced as weighting factors to emphasize NOEs measured at longer mixing times, which are expected to have a better signal-to-noise ratio.

(d) *Structure calculations*

An "ensemble" IRMA protocol as proposed by Bonvin *et al.* (1993*a*) was followed. A structure obtained by Breg *et al.* (1990) was chosen as an initial model for IRMA. In the first refinement cycle, 51 structures were generated with the distance geometry algorithm derived from the original EMBED program (Havel *et al.*, 1983) followed by simulated annealing (SA) calculations with distance bounds driven dynamics (DDD) (Kaptein *et al.*, 1988) and further refined in parallel with a combination of restrained EM and MD (10 ps) using the GROMOS force field and programs (van Gunsteren & Berendsen, 1987). The distance and dihedral angle force constants were set to 40 kJ mol^{-1} Å^{-2} and 110 kJ mol^{-1}, respectively. A cosine type potential (de Vlieg *et al.*, 1986; see remark d, Table 1) was used for dihedral angle restraints, the dihedral angle force constant allowing variation of ±20° at 300 K. Restrained MD simulations were done *in vacuo* at constant temperature (300 K) with a time step of 2 fs. The best structure at this point was used as starting conformation in a free MD simulation in water in order to obtain information about the fast internal motions in Arc (see below). Then, an average relaxation matrix, built from the ensemble of structures by averaging the various contributions as $\langle r^{-6} \rangle$ with inclusion of internal mobility, served as starting point for a new IRMA cycle. A new set of constraints was obtained and all 51 structures were further refined with the same combination of restrained EM and MD as in the first cycle. Convergence was obtained after 3 IRMA cycles. The 16 best structures, selected in terms of residual violations, energies, R factors and stereochemical quality (ϕ, ψ and ω), were finally refined with DINOSAUR (Bonvin *et al.*, 1991) using a slow-cooling simulated annealing protocol (Brünger & Krukowski, 1990; Bonvin *et al.*, 1993*c*) that consisted of 50 steps steepest descent restrained EM, 1·5 ps slow-cooling molecular dynamics (time-step 1 fs, from 600 to 1 K, cooling rate of 10 K/25 fs), finally followed by 50 steps restrained EM. Direct NOE refinement of 1 Arc structure required ~25 h CPU on a Silicon Graphics Crimson server, which corresponds to ~58 s per refinement step.

(e) *Generalized order parameter* S^2

Generalized order parameters (Lipari & Szabo, 1982*a*,*b*) were calculated from a free MD simulation in water (Olejniczak *et al.*, 1984; Koning *et al.*, 1991) starting from a structure obtained in the first IRMA cycle. The Arc dimer was put in a box containing 5223 water molecules with periodic boundary conditions. Eight chlorine ions were added at the water positions with the most positive electrostatic potentials to compensate the positive charges on the protein. The system was first subjected to an EM with harmonic position restraining on the Arc co-ordinates (K_{pos} = 80 kJ mol^{-1} Å^{-2}) to remove the unfavourable interactions introduced by adding water. Initial velocities for the MD simulation were taken from a Maxwellian distribution at 300 K. The MD simulation was carried out with a time-step of 1·0 fs and the temperature and pressure in the system were kept constant with relaxation times of 0·1 and 0·5 ps, respectively. After equilibration, the internal correlation functions for all interproton vectors within 5 Å (~9200) were computed from a 40 ps trajectory (800 structures). After visual inspection, generalized order parameters S_{ij}^2 were computed as the averaged value of the internal correlation functions, normalized at the origin, between t = 8 and t = 16 ps.

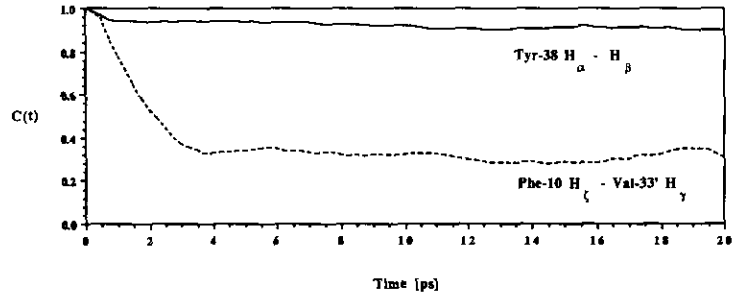

Figure 1. Plot of the first 20 ps of the internal correlation functions (eqn (3) of Bonvin *et al.* (1993*a*), including both angular and radial contributions), normalized at $t = 0$, of the intraresidue H^α-H^β vector of Tyr38 in the second helix and of the intermonomer vector between the aromatic proton H^ζ of Phe10 and a methyl group of Val33′ computed from a 40 ps free MD trajectory in water.

(f) *Crystal structure*

The co-ordinates of a crystal structure of the Arc repressor solved at 1·8 Å resolution were kindly provided by Dr C. R. Kissinger and Professor C. O. Pabo (MIT, Cambridge, U.S.A.). The unit cell contains 2 almost identical Arc dimers. The co-ordinates for the first 6 residues and the C-terminal part (residues 47 to 53) are missing in the crystal structure, crystallographic refinement being still in progress for these regions. We selected the first dimer, which has the best symmetry, for comparison with the solution structures. The missing parts were modelled from the final average solution structure. The crystal structure was solved for the His39Gln mutant of Arc. After changing the side-chain of His into that of Gln, 200 steps of unrestrained energy minimization with the GROMOS force-field were carried out in order to remove unfavourable contacts. The final structure has a total potential energy of -7920 kJ mol^{-1} and is close to the original structure with r.m.s. deviations of 0·13 Å on all atoms and 0·09 Å on backbone atoms. In the following, this structure will be denoted as the X-ray structure.

3. Results and Discussion

(a) *Generalized order parameters* S^2

The internal correlation functions of all inter-proton vectors within 5 Å were computed from the 40 ps free MD trajectory. The first 20 ps of two of these functions are plotted in Figure 1 for the intra-residue H^α-H^β vector of Tyr38 situated in the second helix and for an intermonomer contact between the aromatic proton H^ζ of Phe10 and a methyl group of Val33′. Due to fast internal motions, a decay is observed within the first picoseconds, after which a plateau value is reached. It is clear from Figure 1 that, as expected, more mobility is predicted for the long-range vector than for the intraresidue vector within the helix. As already observed in previous studies (Bonvin *et al.*, 1993*a*), a correlation can be found between computed order parameters and secondary structure elements. This is illustrated by the plot of the intraresidue H^α-H^N order parameters as function of the residue sequence in Figure 2: the two α-helices and the β-sheet clearly display the highest values for S^2. More mobility is found for the N and C-terminal regions and for the loop between the two helices. The calculated order parameters were symmetrized for the dimer case and subsequently used in all IRMA and DINOSAUR calculations. Although these results were obtained from a rather short MD trajectory (40 ps after equilibration), the correlation functions reach a plateau value within the first picoseconds and our experience showed that order parameters extracted from such short simulations already allow a better fit of the experimental NOE data (Koning *et al.*, 1991; Bonvin *et al.*, 1993*a*).

(b) *Structure calculations*

A view of the backbone of the 16 best Arc structures at various stages during the refinement (DG/SA, IRMA 1st, 2nd and 3rd cycles and DINOSAUR) is presented in Figure 3, which clearly shows the improvement of the structures. The precision of the well-defined region of Arc including the

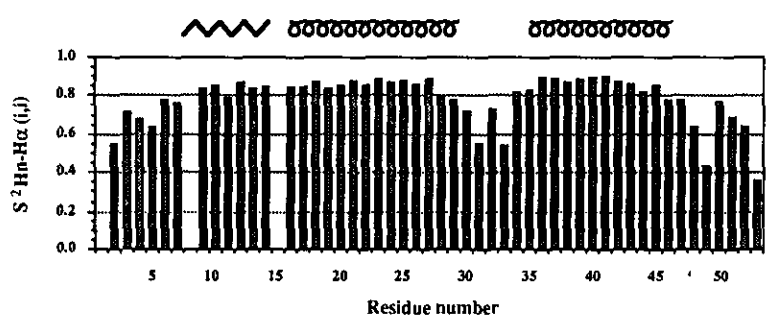

Figure 2. Plot of the intraresidue H^N-H^α order parameters calculated from the free MD simulation in water against the residue sequence and comparison with the secondary structure elements in the Arc repressor.

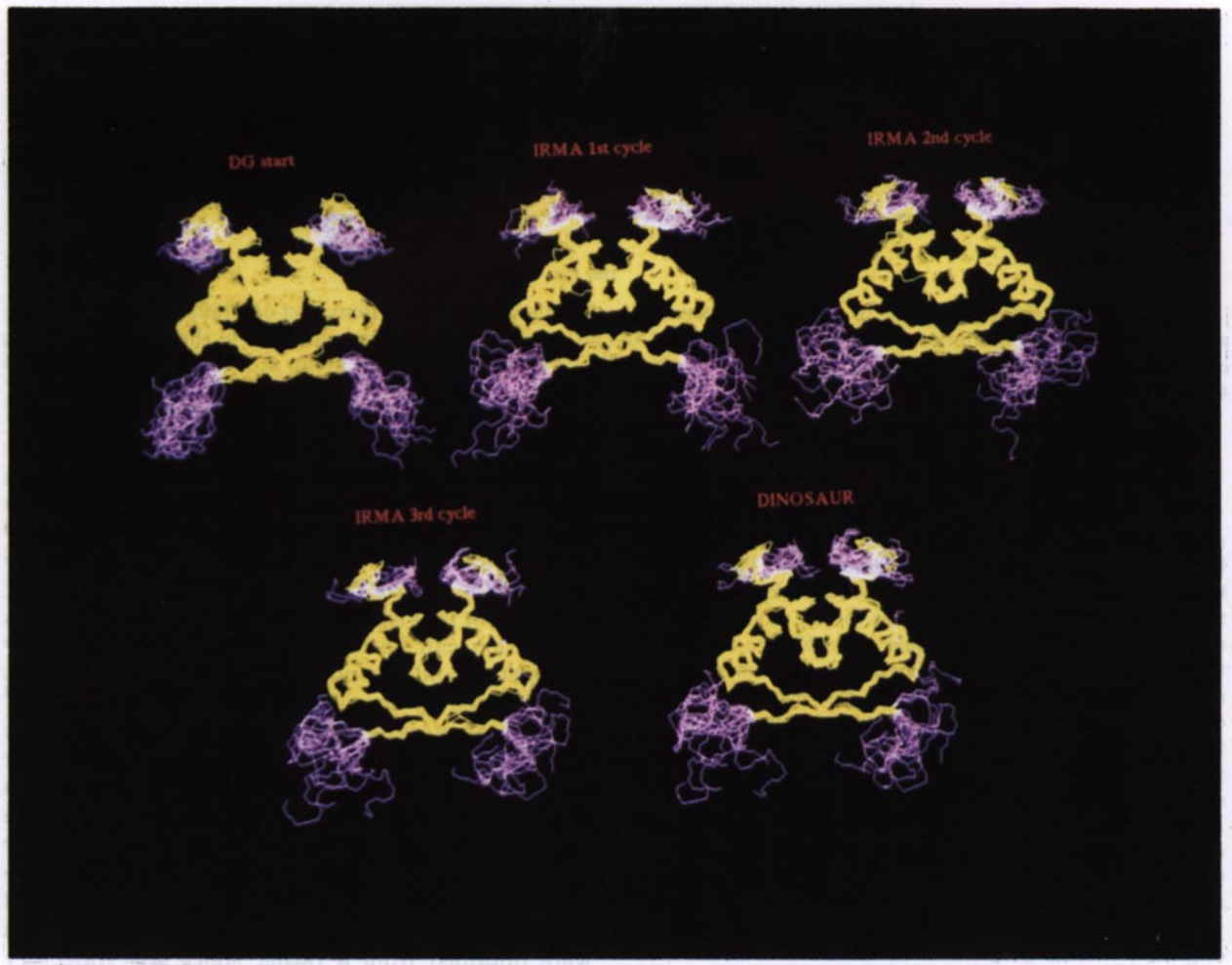

Figure 3. View of the Arc repressor structures during the refinement with NMR constraints. The structures corresponding to the 16 best final conformations are shown directly after the DG/SA calculations in the first IRMA cycle, after each IRMA cycle and after direct NOE refinement with DINOSAUR. The structures were superimposed on C^{α}, C and N atoms of the well-defined region of Arc (residues 8 to 48 of each monomer (shown in yellow)).

β-sheet and the two α-helices (residues 8 to 48) increases during the IRMA refinement as confirmed by the r.m.s.d. (from the average) given in Table 1. The r.m.s.d. for the backbone atoms decrease from 1·0 Å for the DG/SA structures to 0·50 Å for the structures obtained after the third IRMA cycle. This value increases slightly to 0·55 Å after the direct NOE refinement. The R factors and NOE restrained energies show a substantial decrease during the DINOSAUR refinement step as can be seen from Table 1. The distance restraint energies calculated with the set of constraints obtained in the last IRMA cycle increase after direct NOE refinement. This indicates that somewhat different minima of the energy function are obtained with NOE or distance constraints. It is noteworthy that more than half of the increase in distance restraint energy can be accounted for by a set of approximately 40 constraints corresponding to weak NOEs with upper distance restraints almost all larger than 4·5 Å, which were already violated but to a lesser extent in the IRMA structures. Nevertheless, the $Q^{1/6}$ factor, which is sensitive to larger distances, shows a considerable decrease during the DINOSAUR refinement.

In the first cycle of the "ensemble" IRMA refinement, restrained EM and MD calculations result in a considerable improvement of the energies and R factors of the starting structures (DG/SA). Only small differences are found between the subsequent IRMA cycles indicating a fast convergence of the procedure. The refinement with NOEs as direct structural constraints allows, in a short protocol, a substantial further decrease in R factors, which goes along with an increase of the total potential energy due to somewhat larger deviations from ideal bond lengths and angles. The final Arc structures still have good stereochemical qualities and show an extensive pattern of hydrogen bonds, consistent with the secondary structure elements (see below). Some differences are found between the IRMA and DINOSAUR refined structures. The pairwise r.m.s.d. between the average IRMA and DINOSAUR structures is 0·6 Å for backbone atom

Table 1

Comparison of the energies (in kJ mol^{-1}), R factors and r.m.s. deviations of the Arc structures during the refinement

	Start (DG/SA)	IRMA (1st cycle)	IRMA (2nd cycle)	IRMA (3rd cycle)	DINOSAUR
E_{pot}†	23,100±18,500	−5980±207	−7637±152	−7858±111	−7059±133
E_{disre}‡	13,700±3600	3251±249	1767±144	1355±28	4728±269
E_{dihed}§	4970±460	587±201	264±35	176±15	104±13
E_{NOE}‖	10,400±680	8295±445	7796±335	7571±262	3325±186
R	0·72±0·04	0·60±0·01	0·57±0·01	0·57±0·01	0·35±0·01
$Q^{1/6}$	0·186±0·024	0·135±0·004	0·126±0·003	0·123±0·002	0·085±0·002
R_{ens}	0·73	0·61	0·58	0·58	0·34
$Q^{1/6}_{ens}$	0·171	0·135	0·125	0·124	0·078
r.m.s.d. (Å)¶					
All atoms	1·65±0·15	1·25±0·15	1·15±0·15	1·05±0·15	1·10±0·15
Backbone	1·00±0·10	0·70±0·15	0·60±0·20	0·50±0·10	0·55±0·10

Average and ensemble R factors (eqns (3) and (4)) calculated from the ensemble of 16 structures using all NOE peaks. The ensemble R factors were calculated from an average relaxation matrix. The theoretical NOE intensities were computed with inclusion of methyl rotation and internal motions *via* order parameters obtained from a free MD run in water. The overall correlation time was optimized to 4·25 ns and the methyl group rotation correlation time was taken to be 0·1 ns. An additional leakage rate of 2 s^{-1} was added on the diagonal of the relaxation matrix to account for external relaxation contributions.

† GROMOS force field (van Gunsteren & Berendsen, 1987), excluding restraint energies.

‡ Harmonic potential (Kaptein *et al.*, 1985), force constant = 40 kJ mol^{-1} Å^{-2}, energy calculated for the set of 2×1002 distance constraints obtained in the third IRMA cycle. These latter were replaced by direct NOE constraints in the DINOSAUR refinement stage.

§ Dihedral angle restraining potential $V_{dihed} = K_{dihed}(1-\cos(\phi-\phi_0))$ (de Vlieg *et al.*, 1986), force constant = 110 kJ mol^{-1}, energies calculated for a set of 2×21 χ_1 dihedral angle constraints.

‖ NOE potential of eqn (1) force constant = 400 kJ mol^{-1}. The weighting factors w_{ij} were calculated with the experimental noise levels (N) and a relative error $\varepsilon = 5\%$.

¶ The r.m.s. deviations from the average for the well-defined region of Arc (residues 8 to 48 of each monomer).

and 0·8 Å for all heavy atoms for the well defined core of the protein (residues 8 to 48). Although the average IRMA structure is within the spread of DINOSAUR structures, significant differences are found, from an analysis of residue r.m.s.d., around Arg23, Val33 and Glu43. In these three regions, the DINOSAUR refined structures are found to be closer to the X-ray structures than the IRMA structures. The relatively large drift of 0·6 Å and the much lower R factors justify the use of direct NOE refinement for protein structure determination.

The final refinement stage with direct NOE constraints, although resulting in a much better fit of the experimental NMR data with lower NOE and dihedral angle restraints energies, causes a slight increase in r.m.s.d. (+0·05 Å). There could be different reasons for this. The use of direct NOE constraints might introduce more variability in the structures: due to spin diffusion effects there are more ways to fulfil a NOE constraint than a distance constraint. On the other hand, the slight increase in r.m.s.d. could just be the effect of the slow-cooling simulated annealing protocol used, the high temperatures allowing a better sampling of the conformational space. To check this we submitted the structures obtained in the third IRMA cycle to the same slow-cooling annealing protocol, but with standard distance restraints. This does not affect the precision of the structures, no increase in r.m.s.d. being observed, and indicates that the slight increase in r.m.s.d. is the result of the use of direct NOE constraints. The precision of the structures obtained after the third IRMA cycle might in fact exceed their actual accuracy, possibly due to too tight constraints. Such effects have been reported by Liu *et al.* (1992) in an analysis of various structure calculation methods with synthetic NMR data.

(c) *Analysis of the solution structure of the Arc repressor*

The solution structure of the Arc repressor is well determined for the core of the dimer, residues 8 to 48 (and 8′ to 48′ of the second monomer), including the antiparallel β-sheet and the two α-helices with final r.m.s.d (from the average) of 0·55(±0·10) Å on backbone atoms and 1·10(±0·15) Å on all heavy atoms (the average pairwise r.m.s.d. are 0·80(±0·15) Å and 1.65(±0·25) Å, respectively). The structures have a good symmetry with pairwise r.m.s.d. between the two monomers (residues 8 to 48) of 0·80±(0·20) Å and 1·70(±0·20) Å for backbone and all heavy atoms, respectively. A large number of constraints, corresponding to an average of 19 constraints per residue, were used in the structure calculations and their distribution as function of the residue sequence is plotted in Figure 4, together with the precision (r.m.s.d. on C^α and all heavy atoms), the angular order parameters (Hyberts *et al.*, 1992) and the $Q^{1/6}$ factors. A correlation can be noticed between the number of constraints and the precision per residue, but no such relation can be found for the $Q^{1/6}$ factors,

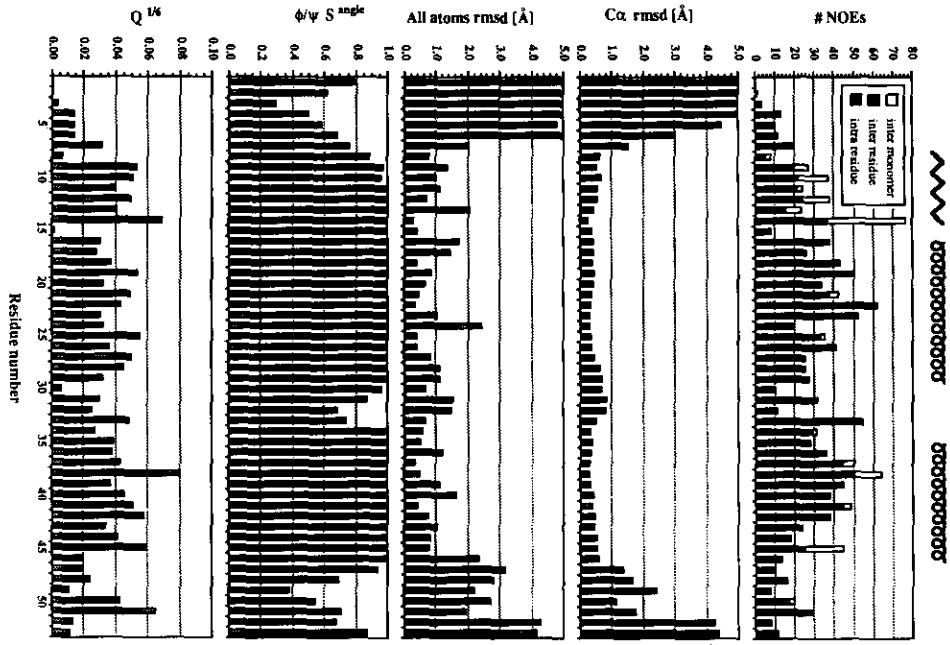

Figure 4. Number of constraints, r.m.s.d. (from the average) on C^{α} atoms and on all heavy atoms (the structures were superimposed on C^{α}, C and N atoms of the well-defined region of Arc (residues 8 to 48 of each monomer)), backbone angular order parameters (averages from ϕ and ψ angular order parameters: Hyberts *et al.*, 1992) and $Q^{1/6}$ factors (eqn (4)) as function of the residue sequence for the final Arc structures. Secondary structure elements in Arc are indicated. (The r.m.s.d. of the first 4 residues are off-scale and have values up to 10 Å.)

which all represent low values (~0·05: for comparison, the $Q^{1/6}$ factors for a linear chain of Arc have values around 0·55). Regions of high variability are found in the N and C-terminal parts for which few constraints are available. Internal mobility was predicted from the free MD simulations in water for these regions (see Fig. 2), but not enough to account for the high level of variability, especially in the N-terminal part; the lack of constraints and, consequently, the high r.m.s.d. might be the result of motions taking place at longer time-scales, which were not monitored in the free MD simulations. Two residues in the C-terminal part, Arg50 and Ile51, display somewhat less variability; this can be explained by the presence of intermonomer NOEs with residues of the first helix. Interestingly, this observation correlates with the order parameters obtained from the unrestrained MD in water in Figure 2. The structures show typical characteristics of solution structures with more disorder in the side-chains on the surface and a well-defined hydrophobic core formed by residues of both monomers. Details of the secondary structure elements in Arc (the DNA-binding β-sheet and the 2 α-helices) are presented in stereoviews in Figure 5. One can clearly recognize the residues pointing toward the interior of the protein (right side of the stereoviews) with very well defined side-chains and those on the surface of protein showing somewhat more variability.

The side-chain χ_1 dihedral angles satisfy the experimental values obtained from J coupling data as can be seen from Table 2. No significant deviation from the experimental values is found. The solution structures were searched for hydrogen bonds using as criteria a maximum distance of 2·5 Å and a minimum angle between donor and acceptor of 135°. Results of this search and comparison with

Table 2

Side chain χ_1 dihedral angles of those residues for which experimental data are available for the Arc structures

	Exp. (deg.)	NMR (deg.)		Exp. (deg.)	NMR (deg.)
Pro8	0	−2±9	Pro8′	0	−1±11
Gln9	60	65±5	Gln9′	60	64±12
Asn11	180	−176±6	Asn11′	180	−176±5
Arg13	60	55±6	Arg13′	60	54±6
Trp14	−60	−77±3	Trp14′	−60	−65±3
Pro15	0	−7±9	Pro15′	0	−5±10
Arg16	−180	−180±5	Arg16′	−180	−175±6
Val18	−180	−187±4	Val18′	−180	183±3
Asp20	−60	−62±2	Asp20′	−60	−63±3
Leu21	−180	−184±4	Leu21′	−180	−177±5
Val22	−180	−185±5	Val22′	−180	−189±4
Val25	−180	−185±6	Val25′	−180	−177±9
Glu27	−60	−72±4	Glu27′	−60	−75±3
Val33	−180	−198±11	Val33′	−180	−182±14
Asn34	−180	−185±7	Asn34′	−180	−177±4
Ser35	−180	−179±3	Ser35′	−180	−179±2
Tyr38	−180	177±4	Tyr38′	−180	−176±4
Arg40	−60	−64±7	Arg40′	−60	−67±6
Val41	−180	−182±9	Val41′	−180	−187±6
Glu43	−180	−203±4	Glu43′	−180	−191±3
Phe45	−60	−66±5	Phe45′	−60	−59±3

Exp. denotes the experimental values based on J coupling data and NMR the average angles from the final 16 DINOSAUR structures.

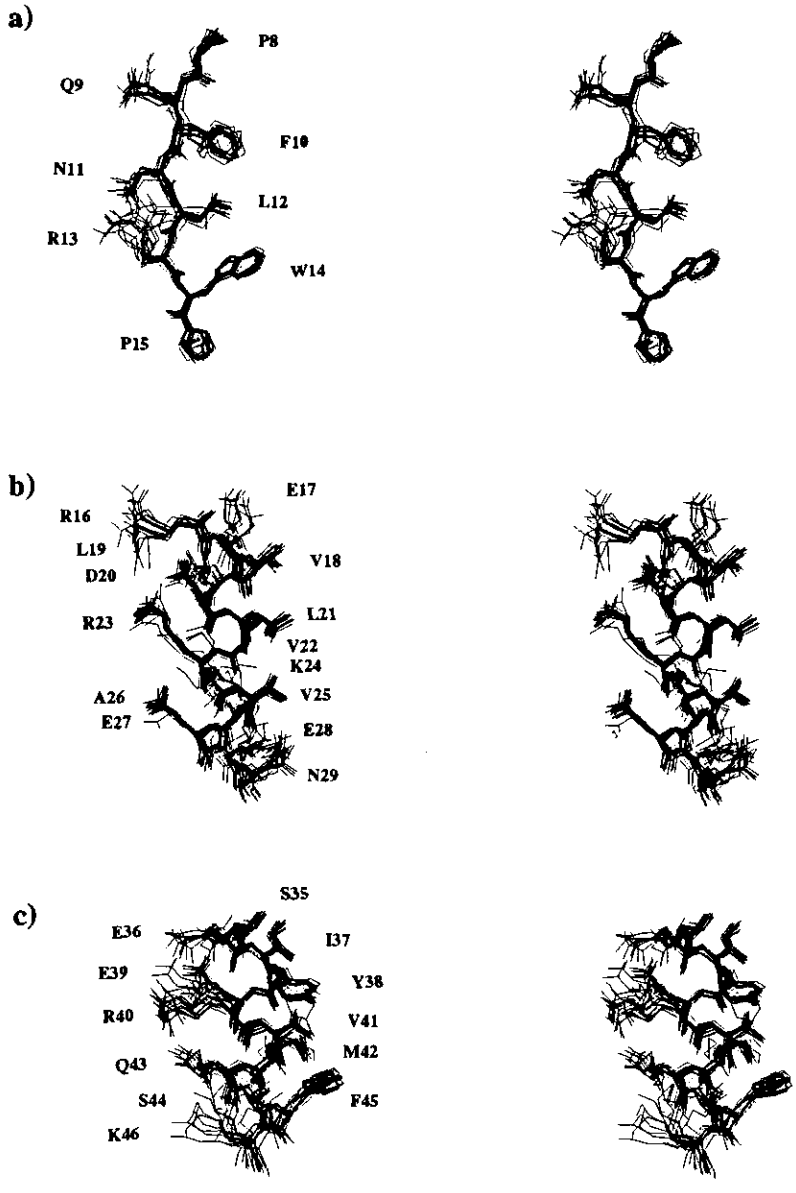

Figure 5. Stereoviews of the secondary structure elements in Arc. The 16 final Arc structures were superimposed on backbone atoms (C^{α}, C and N) of the well-defined core (residues 8 to 48 of each monomer). Details of the structure including all heavy atoms are shown for: a) the DNA-binding β-sheet (residues 8 to 15), b) the first α-helix (residues 16 to 29) and c) the second α-helix (residues 35 to 46).

the experimental exchange data (Burgering *et al.*, unpublished results) are presented in Table 3. Hydrogen bonds are indicated if found in 75, 50 and 25% of the structures, respectively. The hydrogen-bond pattern is consistent with the secondary structure elements in Arc: bonds are found for the amide protons in the β-sheet (Pro8 to Trp14) and in the two α-helices (Arg16 to Asn29 and Ser35 to Lys46). The slow exchange of the amide protons of Gly30 and Arg31 following the first helix and the intermediate exchange of Asn34 in the loop between the two helices is well accounted for by strong (for Gly30 and Arg31) and weak (for Asn34) hydrogen bonds in the solution structure, respectively. A few structures also show hydrogen bonds in the N-terminal region where fast exchange was monitored. A weak hydrogen bond is found in the C-terminal part between the amide proton of Ile51 and the carbonyl group of Gly49. An ill-defined turn seems to be present in solution at this position.

We checked the stereochemical quality of the ensemble of 16 structures with the program PROCHECK (Morris *et al.*, 1992). The results are summarized in Table 4. The ensemble of 16 Arc structures show good stereochemical qualities, close to the ideal values reported by Morris *et al.* (1992) and corresponds to a classification of 1, 2, 2. Taken individually, all structures have good stereo-

Table 3

Comparison between the hydrogen bonds found in the final NMR (N), X-ray (X) structures and the amide exchange (E) NMR data

Amide	Acceptor		*N*†	*X*	*E*		Amide	Acceptor		*N*†	*X*	*E*
Met1	—				—		Met1′	—				—
Lys2	—				f		Lys2′	—				f
Gly3	Met1	C=O	±		f		Gly3′	Met1′	C=O	±		f
Met4	Lys2	C=O	±		f		Met4′	Lys2′	C=O	+		f
Ser5	Gly3	C=O	±		f		Ser5′	—				f
Lys6	Met4	C=O	±		f		Lys6′	Gly3′	C=O	±		f
Met7	Ser5	C=O	±		f		Met7′	Ser5	C=O	±		f
Pro8	—				—	β	Pro8′	—				—
Gln9	—				nm	β	Gln9′	Met7′	C=O	±		nm
	—				nm	β		Gln9′	O^{ε}	±		nm
Phe10	Leu12′	C=O	+ +	+	nm	β	Phe10′	Leu12	C=O	+ +	+	nm
Asn11	—				nm	β	Asn11′	—				nm
Leu12	*Phe10′*	C=O	+ +	+	*i*	β	*Leu12′*	*Phe10*	C=O	+ +	+	*i*
Arg13	Asn34′	O^{δ}	+	+	i	β	Arg13′	Asn34	O^{δ}	+	+	i
Trp14	Pro8′	C=O	+ +	+	s	β	Trp14′	Pro8	C=O	+ +	+	s
Pro15	—				—		Pro15′	—				—
Arg16	Lys6′	C=O	±		f	α	Arg16′	Lys6	C=O	±		f
Glu17	Glu17	O^{ε}	+	+	nm	α	Glu17′	Glu17′	O^{ε}	±	+	nm
Val18	Pro15	C=O	+ +		s	α	Val18′	Pro15′	C=O	+		s
Leu19	Pro15	C=O	+ +	+	s	α	Leu19′	Pro15′	C=O	+ +	+	s
Asp20	Arg16	C=O	+ +	+	s	α	Asp20′	Arg16′	C=O	+ +	+	s
Leu21	Glu17	C=O	+ +	+	s	α	Leu21′	Glu17′	C=O	+ +	+	s
Val22	Val18	C=O	+ +	+	s	α	Val22′	Val18′	C=O	+ +	+	s
Arg23	Leu19	C=O	+ +	+	s	α	Arg23′	Leu19′	C=O	+ +	+	s
Lys24	Asp20	C=O	+ +	+	s	α	Lys24′	Asp20′	C=O	+ +	+	s
Val25	Leu21	C=O	+ +	+	s	α	Val25′	Leu21′	C=O	+ +	+	s
Ala26	Val22	C=O	+ +	+	s	α	Ala26′	Val22′	C=O	+ +	+	s
Glu27	Arg23	C=O	+	+	s	α	Glu27′	Arg23′	C=O	±	+	s
	Lys24	C=O			s	α		Lys24′	C=O	+		s
Glu28	Lys24	C=O	+ +	+	s	α	Glu28′	Lys24′	C=O	+ +	+	s
Asn29	Val25	C=O	+ +	+	s	α	Asn29′	Val25′	C=O	+ +	+	s
Gly30	Glu27	C=O	+ +	+	s		Gly30′	Glu27′	C=O	+ +	+	s
Arg31	Ala26	C=O	+ +	+	s		Arg31′	Ala26′	C=O	+ +	+	s
Ser32	—				i		Ser32′	—				i
Val33	Arg31	C=O	+		f		Val33′	Arg31′	C=O	±		f
Asn34	Ser32	C=O	+		i		Asn34′	Ser32′	C=O	±		i
Ser35	Ser32	C=O	+		s	α	Ser35′	Ser32′	C=O	+		s
	Ser32	O^{γ}		+	s	α		Ser32′	O^{γ}		+	s
Glu36	Ser32	C=O	+	+	s	α	Glu36′	Ser32′	C=O	+	+	s
Ile37	Val33	C=O	+ +	+	s	α	Ile37′	Val33′	C=O	+ +	+	s
Tyr38	Asn34	C=O	+ +	+	s	α	Tyr38′	Asn34′	C=O	+ +	+	s
Gln39	Ser35	C=O	+ +	+	s	α	Gln39′	Ser35′	C=O	+ +	+	s
Arg40	Glu36	C=O	+ +	+	s	α	Arg40′	Glu36′	C=O	+ +	+	s
Val41	*Ile37*	C=O	+ +	+	s	α	*Val41′*	*Ile37′*	C=O	+ +	+	s
Met42	Tyr38	C=O	+ +	+	s	α	Met42′	Tyr38′	C=O	+ +	+	s
Glu43	Gln39	C=O	+	+	s	α	Glu43′	Gln39′	C=O	+ +	+	s
Ser44	Arg40	C=O	+ +	+	s	α	Ser44′	Arg40′	C=O	+ +	+	s
Phe45	Val41	C=O	±	+	s	α	Phe45′	Val41′	C=O	+ +	+	s
	Met42	C=O	+		s	α		Met42′	C=O			s
Lys46	*Met42*	C=O	±	+	s	α	*Lys46′*	*Met42′*	C=O	±		s
	Glu43	C=O	+		s	α		Glu43′	C=O	±	+	s
Lys47	Ser44	C=O	+ +		nm		Lys47′	Ser44′	C=O	+ +		nm
Glu48	Ser44	C=O			nm		Glu48′	Ser44′	C=O			nm
	Lys46	C=O	+		nm			Lys46′	C=O	±		nm
Gly49	Lys46	C=O			nm		Gly49′	Lys46′	C=O			nm
	Lys47	C=O	±		nm			Lys47′	C=O	±		nm
Arg50	Glu48	C=O			nm		Arg50′	Glu48′	C=O	±		nm
	Glu28′	O^{ε}	±		nm			Glu28	O^{ε}			nm
Ile51	Phe45	C=O			nm		Ile51′	Phe45′	C=O			nm
	Glu48	C=O	±		nm			Glu48′	C=O	±		nm
	Gly49	C=O	±		nm			Gly49′	C=O	±		nm
Gly52	Lys47	C=O	±		f		Gly52′	—				f
Ala53	Arg50	C=O	±		f		Ala53′	Arg50′	C=O	+		f
	Ile51	C=O	±		f			Ile51′	C=O	±		f

Table 3 *(continued)*

Other hydrogen bonds found in the NMR structures

Donor		Acceptor			Donor		Acceptor		
Gln9	H^{ε}	Asn11′	O^{δ}	±	Gln9′	H^{ε}	Asn11	O^{δ}	±
Asn11	H^{δ}	Gln9′	C^{ε}	±	Asn11′	H^{δ}	Gln9	O^{ε}	±
Trp14	$H^{\varepsilon 1}$	Asn34′	C=O	+	Trp14′	$H^{\varepsilon 1}$	Asn34	C=O	±
	$H^{\varepsilon 1}$	Asn34′	O^{δ}	+		$H^{\varepsilon 1}$	Asn34	O^{δ}	+ +
Asn29	$H^{\delta 2}$	Arg50′	C=O	±	Asn29′	$H^{\delta 2}$	Arg50	C=O	±
Arg31	H^{ε}	Gln39	O^{ε}	±	Arg31′	H^{ε}	Gln39′	O^{ε}	±
	$H^{\eta 2}$	Gln39	O^{ε}	+		$H^{\eta 2}$	Gln39′	O^{ε}	+
Asn34	$H^{\delta 2}$	Arg13′	C=O	+ +	Asn34′	$H^{\delta 2}$	Arg13	C=O	+ +
Ser35	H^{γ}	Ser35	C=O	±	Ser35′	H^{γ}	Ser35′	C=O	±
	H^{γ}	Gln39	O^{ε}	+		H^{γ}	Gln39′	O^{ε}	±
Arg40	H^{ε}	Glu36	O^{ε}	±	Arg40′	H^{ε}	Glu36′	O^{ε}	±
	$H^{\eta 1}$	Glu43	O^{ε}	+		$H^{\eta 1}$	Glu43′	O^{ε}	+
	$H^{\eta 2}$	Arg50′	C=O			$H^{\eta 2}$	Arg50	C=O	±
	$H^{\eta 2}$	Ile51′	C=O	±		$H^{\eta 2}$	Ile51	C=O	
Ser44	H^{γ}	Glu43′	O^{ε}	+	Ser44′	H^{γ}	Glu43	O^{ε}	+
Arg50	$H^{\eta 2}$	Glu28′	C=O	±	Arg50′	$H^{\eta 2}$	Glu28	C=O	

† The hydrogen bonds were calculated from the final set of 16 structures obtained after DINOSAUR refinement. The β-sheet and the 2 α-helices found in the Arc solution structures are indicated by β and α, respectively. A maximum distance of 2·5 Å and a minimum angle of 135° were used as criteria to indicate a hydrogen bond. + +, + and ± indicate hydrogen bonds found in at least 75, 50 and 25% of the structures, respectively; f, i and s indicate fast, intermediate and slow exchange, respectively; nm indicates that the exchange could not be monitored experimentally and can be regarded as intermediate exchanging.

chemical qualities, six having a classification of 1, 1, 2, and none exceeding a classification of 2 for any of the criteria.

(d) *Comparison with the results of mutagenesis studies*

A considerable amount of genetic data is available for the Arc repressor (for a review, see Knight *et al.*, 1989). From an analysis of functionally and/or structurally allowed amino acid substitutions, residues that are likely to be involved in protein function or to play important structural roles in Arc could be identified. This also allowed a prediction of secondary structure elements (Bowie & Sauer, 1989; Vershon *et al.*, 1986). About one-third of the residues in Arc repressor were found to be functionally important and about one-half structurally important. A short region of β-strand structure for residues 9 to 14 and two helical regions at positions 16 to 28 and 35 to 47, with a clustering of hydrophobic residues on one side of the helices that should be favourable to the formation of a hydrophobic core, were predicted. These predictions are in remarkable agreement with the NMR structures in which a β-sheet from Pro8 to Trp14 and two α-helices from Arg16 to Asn29 and Ser35 to Lys46 are found.

Among the residues essential for the stability of the dimer, 15 were postulated to be important for the formation of the hydrophobic core; these are Phe10, Leu12, Trp14, Val18, Leu21, Val22, Val25, Ala26, Val33, Ile37, Tyr38, Val41, Met42, Phe45 and Lys46. The side-chains of those residues are all very well defined in the solution structures, except for Lys46, as can be seen from the plot of the all atoms r.m.s.d. per residue in Figure 4. The other structurally important residues were postulated to stabilize the structure by tertiary hydrogen bonds or ion pairs (Arg23, Glu28, Arg31, Glu36 and Arg40) or to be involved in special backbone conformations (Pro15 and Gly30). Pro15 and Gly30 are, indeed, found in loop regions, between the β-sheet and the first helix and between the two helices, respectively. No interactions are found for the amino protons of Arg23, the side-chain of which points in fact toward the DNA-binding side. This residue is thus likely to play a role in DNA binding. Some intermonomer hydrogen bonds are found in the structures between Arg50 and Glu28′, both between the amide proton H^{N} of Arg50 and the carboxyl group of Glu28′ and between the H^{η} protons of Arg50 and the carbonyl

Table 4

Stereochemical quality of the ensemble of 16 final Arc repressor structures

Parameter	Arc values	Ideal values†
χ_1 dihedral $g(-)$ (deg.)	60·5 ± 10·5	64·1 ± 15·7
χ_1 dihedral *trans* (deg.)	179·8 ± 12·7	183·6 ± 16·8
χ_1 dihedral $g(+)$ (deg.)	−67·5 ± 12·5	−66·7 ± 15·0
χ^2 dihedral (deg.)	178·6 ± 18·2	177·4 ± 18·5
Proline ϕ (deg.)	−63·7 ± 17·1	−65·4 ± 11·2
Helix ϕ (deg.)	−61·7 ± 12·3	−65·3 ± 11·9
Helix ψ (deg.)	−41·5 ± 13·9	−39·4 ± 11·3
ω dihedral (deg.)	179·0 ± 5·2	180·0 ± 5·8
H-bond energy (kJ)	−2·1 ± 0·7	−2·0 ± 0·8
Chirality C^{α} (deg.)	31·0 ± 4·3	33·9 ± 3·5

Parameter	Arc value	Class
Residues with ϕ/ψ in most favoured regions (%)	81·4	1
χ_1 standard deviation	12·5	2
H-bond energy standard deviation	0·74	2

† According to Morris *et al.* (1992) (PROCHECK).

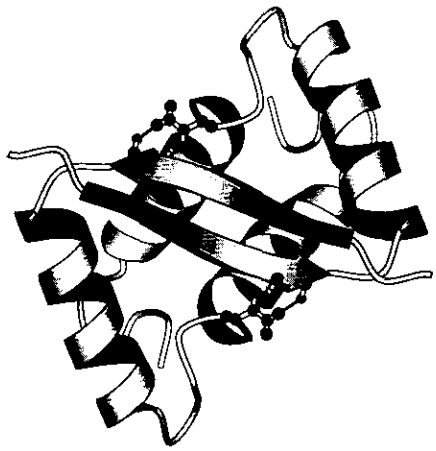

Figure 6. Ribbon view of Arc along the symmetry axis. The side-chain atoms of the 2 Asn34 and Asn34′ stabilizing the DNA-binding β-sheet by hydrogen bonds (indicated in black) with the carbonyl groups and amide protons of Asp13′ and Asp13, respectively, are shown explicitly. (This picture was created with the MOLSCRIPT program (Kraulis, 1991).)

group of Glu28′ that could explain the functional role of these residues, DNA binding being improved by a stabilization of the dimer. The side-chain of Arg50 is, however, not very well determined in solution and, in addition, both residues are at positions that could be involved in dimer–dimer contacts. The other residues mentioned, Arg31, Glu36 and Arg40, together with Gln39 and Glu43 form a hydrophilic pocket that contributes to the stabilization of the structure by a network of hydrogen bonds (see Table 3) and a salt bridge between Arg40 and Glu43. Hydrogen bonds are found between the H^{ε} protons of Arg31 and Arg40 and the carboxyl groups of Gln39 and Glu36, respectively, which are consistent with the slow exchange monitored for these protons (Burgering *et al.*, unpublished results). The presence of a salt bridge between Arg40 and Glu43, and the hydrogen bond to Glu36 could explain the downfield shift (9·36 p.p.m.) of the H^{ε} proton of Arg40 in the NMR spectrum.

Several steps are important for the specific binding of Arc repressor to its operator (Brown & Sauer, 1993). In the first step, denatured monomers form dimers and then these dimers will bind sequentially to the operator half-sites forming the Arc tetramer–operator complex. All steps contribute to the DNA binding, but the last two steps (DNA binding and tetramer formation) will do most for specificity. Many of the functionally important residues of Arc repressor are found in the N terminus (from residue 2 to 14), where they either contribute to a stabilisation of the β-sheet or are directly involved in DNA recognition (Gln9, Asn11 and Arg13). However, some functional residues were found at other places in the Arc repressor (Asp20, Asn34, Glu48 and Arg50). The side-chain of Asp20 points toward the β-sheet and may be directly involved in its stabilization. The role of Asn34 might be dual. It forms intermonomer hydrogen bonds with the carbonyl and backbone amide of Arg13′ and with the indole hydrogen of Trp14′ (see Table 3), which stabilizes the DNA-binding region, but is also part of the loop between both helices that is supposed to be involved in dimer–dimer contacts (Breg *et al.*, 1990). Figure 6 shows a ribbon view of Arc with the two explicit Asn34 side-chains. It could be speculated that Glu48 and Arg50 have a dual role as well. Model building of the Arc tetramer–operator complex (Breg *et al.*, 1990) shows that in one monomer of each dimer those residues are in a position to make dimer–dimer interactions. However, Arg50 is also involved in a hydrogen-bond network involving Ala53, Glu48 and Glu28′, which favours dimer formation and thus could enhance the DNA-binding affinity. The formation of hydrogen bonds in this region is made possible by the fact that the C-terminal part of one monomer is pointing toward the first helix of the other monomer in solution. A few intermonomer NOEs are observed in solution between these two regions (from Arg50 to Glu28′ and Asn29′ and from Ile51 to Leu19′ and Ala26′). These have been confirmed by the double half-filtered 2D NOESY experiments (Burgering *et al.*, 1993).

(e) *Comparison with the X-ray structure*

The r.m.s.d. between the ensemble of 16 solution structures at various stages during the refinement and the X-ray structure are given in Table 5. We notice the continuous decrease in r.m.s.d. as a function of the refinement, even in the final stage with direct NOE constraints, during which a slight increase in r.m.s.d. is observed within the set of solution structures. All values are, however, somewhat larger than the r.m.s.d. within the sets of solution structures given in Table 1, indicating that some significant differences are present. Figure 7 shows a stereoview of the superimposed minimized average solution structure and the crystal structure of the Arc repressor. The structures have pairwise r.m.s.d. of 0·90 Å for backbone atoms and 1·70 Å for all heavy atoms for the part of the protein being well defined in the crystal (residues 8 to 46). Small differences in the backbone conformation are found in the loop region between the two helices and at the

Table 5

The r.m.s.d. (Å) between the crystal and the ensemble of 16 solution structures at various stages during the refinement

	All atoms	Backbone
IRMA 1st cycle	2·20±0·10	1·40±0·15
IRMA 2nd cycle	2·10±0·15	1·25±0·15
IRMA 3rd cycle	2·00±0·05	1·10±0·10
DINOSAUR	1·90±0·10	1·00±0·10

Average r.m.s. deviations calculated for the well-defined region of Arc in the crystal (residues 8 to 46 of each monomer).

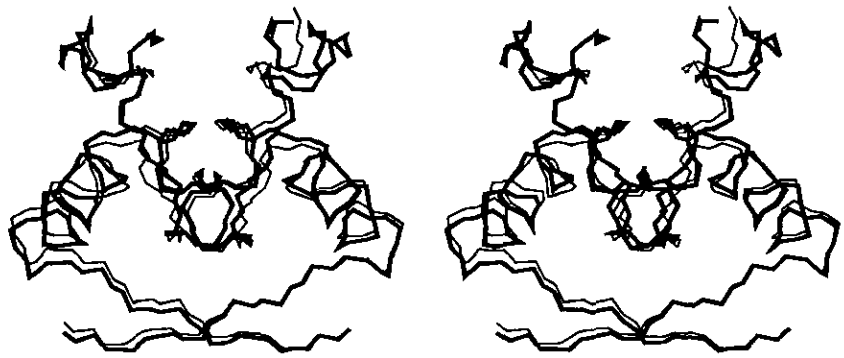

Figure 7. Stereoview of the backbone of the minimized average solution (residues 8 to 53) (broad lines) and X-ray (residues 8 to 46) structures of the Arc repressor. The structures were superimposed on C^{α}, C and N atoms of the well-defined region of Arc (residues 8 to 46 of each monomer).

end of the second helix. The hydrogen-bond patterns for amide protons in the two structures are in good agreement (see Table 3). Finally, the various energies and R factors for the X-ray structure are presented in Table 6. The crystal structure has a good GROMOS potential energy but very high distance and dihedral restraint energies. The NOE restraint energy and R factors are close to the values obtained for the IRMA structures, which is not surprising considering the good agreement between solution and crystal structures for the core of the protein and the fact that the missing C and N-terminal parts were modelled from the NMR structure.

4. Conclusion

The use of relaxation matrix calculations for the refinement of the Arc repressor from 2D NMR data resulted in an ensemble of structures that gives an accurate representation of the solution structure of Arc. The inclusion of internal mobility in the relaxation matrix calculations, *via* order parameters, fast methyl group rotation and aromatic ring flips, allows a dynamic and more realistic description of the structure in solution. In the first stage of the refinement, the "ensemble" IRMA showed fast convergence toward a very well defined set of structures. Then, direct refinement against experimental NOE data with DINOSAUR allowed, in an efficient protocol, a considerable decrease in R factors of about 30 to 40%. A new R factor ($Q^{1/6}$) that combines the advantages of previous definitions, was introduced.

The final structures satisfy the NMR constraints with no significant deviation from dihedral angle restraints obtained from J coupling data and very low R factors, indicating a very good agreement with the experimental NOE data. The structures, which have been deposited into the Brookhaven Protein Data Bank (entry 1ARR), have good stereochemical qualities and present an extensive pattern of hydrogen bonds and a good packing with a very well defined core. They are consistent with the secondary structure predictions and explain the structural and/or functional role of many residues previously found to be important from mutagenesis studies (Bowie & Sauer, 1989). Particularly interesting is the role of the asparagine residues at positions 34/34′, which appear to stabilize the DNA-binding β-sheet by intermonomer hydrogen bonds to the carbonyl groups and amide protons of Arg13′/13 and to the indole hydrogen atoms of Trp14′/14. Mutants at position 34 do not affect the folding of the dimer, but the stabilization of the β-sheet that is required for DNA binding. Finally, The role of Arg50 in the C-terminal region, which is determined by intermonomers NOEs, could be dual; favouring dimerization, but also contributing to dimer–dimer contacts in the complex with DNA.

Table 6
Energies (in kJ mol^{-1}) and R *factors of the crystal structure*

E_{pot}	−7920
E_{disre}	8955
E_{dihed}	3323
E_{NOE}	7037
R	0·56
$Q^{1/6}$	0·128

Energies and R factors calculated as for the solution structures (see remarks in Table 2).

The authors thank Dr C. R. Kissinger and Dr C. O. Pabo (MIT, Cambridge, U.S.A.) for providing us with the crystal co-ordinates of the Arc repressor. This work was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for Scientific Research (NWO).

References

Boelens, R., Koning, T. M. G. & Kaptein, R. (1988). Determination of biomolecular structures from proton–proton NOEs using a relaxation matrix approach. *J. Mol. Struct.* **173**, 299–311.

Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H. & Kaptein, R. (1989). Iterative procedure for structure determination from proton–proton NOEs using a full relaxation matrix

approach. Application to a DNA octamer. *J. Magn. Reson.* **82**, 290–308.

Bonvin, A. M. J. J., Boelens, R. & Kaptein R. (1991). Direct NOE refinement of biomolecular structures using 2D NMR data. *J. Biomol. NMR*, **1**, 305–309.

Bonvin, A. M. J. J., Rullmann, J. A. C., Boelens, R. & Kaptein, R. (1993*a*). "Ensemble" iterative relaxation matrix approach: a new NMR refinement protocol applied to the solution structure of crambin. *Proteins: Struct. Funct. Genet.* **15**, 385–400.

Bonvin, A. M. J. J., Boelens, R. & Kaptein, R. (1993*b*). Determination of biomolecular structures by NMR. Use of relaxation matrix calculations. In *Computer Simulations of Biomolecular Systems: Theoretical and Experimental Applications* (van Gunsteren, W. F., Weiner, P. K. & Wilkinson, A. J., eds), vol. 2, ESCOM Science Publishers, Leiden, in the press.

Bonvin, A. M. J. J., Boelens, R. & Kaptein, R. (1993*c*). Direct NOE refinement of crambin using a slow-cooling simulated annealing protocol. *Biopolymers*, in the press.

Bowie, J. U. & Sauer, R. T. (1989). Identifying determinants of folding and activity for a protein of unknown structure. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 2152–2156.

Bowie, J. U. & Sauer, R. T. (1990). Tra Y proteins of F and related episomes are members of the Arc and Mnt repressor family. *J. Mol. Biol.* **211**, 5–6.

Breg, J. N., Boelens, R., George, A. V. E. & Kaptein, R. (1989). Sequence-specific ^{1}H NMR assignment and secondary structure of the Arc repressor of bacteriophage P22, as determined by two-dimensional ^{1}H NMR spectroscopy. *Biochemistry*, **28**, 9826–9833.

Breg, J. N., van Opheusden, J. H. J., Burgering, M. J. M., Boelens, R. & Kaptein, R. (1990). Structure of Arc repressor in solution: evidence for a family of β-sheet DNA-binding proteins. *Nature (London)*, **346**, 586–589.

Brown, B. M. & Sauer, R. T. (1993). Assembly of the Arc repressor-operator complex: co-operative interactions between DNA-bound dimers. *Biochemistry*, **32**, 1354–1363.

Brown, S. C., Weber, P. L. & Mueller, L. (1988). Toward complete ^{1}H NMR spectra in proteins. *J. Magn. Reson.* **77**, 166–169.

Brünger, A. T. & Krukowski, A. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. sect. A*, **46**, 585–593.

Burgering, M. J. M., Boelens, R., Caffrey, M., Breg, J. & Kaptein, R. (1993). Observation of inter-subunit Nuclear Overhauser Effects in a dimeric protein: application to the Arc repressor. *FEBS Letters*, **330**, 105–109.

de Vlieg, J., Boelens, R., Scheek, R. M., Kaptein, R. & van Gunsteren, W. F. (1986). Restrained molecular dynamics procedure for protein tertiary structure determination from NMR data: a Lac repressor headpiece structure based on information on J-couplings and from presence and absence of NOE's. *Isr. J. Chem.* **27**, 181–188.

Gonzales, C., Rullmann, J. A. C., Bonvin, A. M. J. J., Boelens, R. & Kaptein, R. (1991). Toward an NMR R factor. *J. Magn. Reson.* **91**, 659–664.

Havel, T. F., Kuntz, I. D. & Crippen, G. M. (1983). The theory and practice of distance geometry. *Bull. Math. Biol.* **45**, 665–720.

Hyberts, S. G., Goldberg, M. S., Havel, T. F. & Wagner, G. (1992). The solution structure of Eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736–751.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985). A protein structure from NMR data: Lac repressor headpiece. *J. Mol. Biol.* **182**, 179–182.

Kaptein, R., Boelens, R., Scheek, R. M. & van Gunsteren, W. F. (1988). Protein structures from NMR. *Biochemistry*, **27**, 5389–5395.

Knight, K. L., Bowie, J. U., Vershon, A. K., Kelley, R. D. & Sauer, R. T. (1989). The Arc and Mnt repressors. A new class of sequence-specific DNA-binding protein. *J. Biol. Chem.* **264**, 3639–3642.

Koning, T. M. G. (1990*a*). IRMA: iterative relaxation matrix approach for NMR structure determination. Application to DNA fragments. Ph.D. thesis, University of Utrecht, The Netherlands.

Koning, T. M. G., Boelens, R. & Kaptein, R. (1990*b*). Calculation of the nuclear Overhauser effect and the determination of proton–proton distances in the presence of internal motions. *J. Magn. Reson.* **90**, 111–123.

Koning, T. M. G., Boelens, R., van der Marel, G. A., van Boom, J. H. & Kaptein, R. (1991). Structure determination of a DNA octamer in solution by NMR spectroscopy: effect of fast local motions. *Biochemistry*, **30**, 3787–3797.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 945–949.

Lipari, G. & Szabo, A. (1982*a*). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Amer. Chem. Soc.* **104**, 4546–4559.

Lipari, G. & Szabo, A. (1982*b*). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Amer. Chem. Soc.* **104**, 4559–4570.

Liu, Y., Zhao, D., Altman, R. & Jardetzky, O. (1992). A systematic comparison of three structure determination methods for NMR data: dependence upon quality and quantity of data. *J. Biomol. NMR*, **2**, 373–388.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). Stereochemical quality of protein structure co-ordinates. *Proteins: Struct. Funct. Genet.* **12**, 345–364.

Nilges, M. & Brünger, A. T. (1991). Automated assignment of ambiguous cross peaks in symmetric dimers. *J. Cell. Biochem.* **15G**, abstract CG 422.

Olejniczak, E. T., Dobson, C. M., Karplus, M. & Levy, R. M. (1984). Motional averaging of proton nuclear Overhauser effects in proteins. Predictions from a molecular dynamics simulation of lysozyme. *J. Amer. Chem. Soc.* **106**, 1923–1930.

Pabo, C. O. & Sauer, R. T. (1984). Protein-DNA recognition. *Annu. Rev. Biochem.* **53**, 293–321.

Rafferty, J. B., Somers, W. S., Saint-Girons, I. & Phillips, S. E. V. (1989). Three-dimensional crystal structures of *Escherichia coli met* repressor with and without corepressor. *Nature (London)*, **341**, 705–710.

Sauer, R. T., Krovatin, W., DeAnda, J., Youderian, P. & Susskind, M. M. (1983). Primary structure of the *imm*I immunity region of bacteriophage P22. *J. Mol. Biol.* **168**, 699–713.

Susskind, M. M. (1980). A new gene of bacteriophage P22 which regulates synthesis of antirepressor. *J. Mol. Biol.* **138**, 685–713.

Thomas, P. D., Basus, V. J. & James, T. L. (1991).

Protein solution structure determination from 2D NOE experiments: effects of approximations on the accuracy of derived structures. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 1237–1241.

van Gunsteren, W. F. & Berendsen, H. J. C. (1987). Groningen Molecular Simulation (GROMOS) Library Manual. Biomos BV, Nijenborgh 16, 9747 AG Groningen, The Netherlands.

Vershon, A. K., Youderian, P., Susskind, M. M. & Sauer, R. T. (1985). The bacteriophage P22 Arc and Mnt repressors. Overproduction, purification and properties. *J. Biol. Chem.* **260**, 12124–12129.

Vershon, A. K., Bowie, J. U., Karplus, T. M. & Sauer, R. T. (1986). Isolation and analysis of Arc repressor mutants: evidence for an unusual mechanism of DNA binding. *Proteins: Struct. Funct. Genet.* **1**, 302–311.

Withka, J. M., Srinivasan, J. & Bolton, P. H. (1992). Problems with, and alternatives to, the NMR *R* factor. *J. Magn. Reson.* **98**, 611–617.

Wüthrich, K., Billeter, M. & Braun, W. (1983). Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton–proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949–961.

Yip, P. & Case D. A. (1989). A new method for refinement of macromolecular structures based on nuclear Overhauser effect spectra. *J. Magn. Reson.* **83**, 643–648.

Edited by P. E. Wright

(Received 3 August 1993; accepted 22 October 1993)