The EMBO Journal vol.13 no.17 pp.3936–3944, 1994

Solution structure of the LexA repressor DNA binding domain determined by ^{1}H NMR spectroscopy

R.H.Fogh, G.Ottleben[1], H.Rüterjans[1], M.Schnarr[2], R.Boelens and R.Kaptein[3]

Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584-CH Utrecht, The Netherlands, [1]Institute of Biophysical Chemistry, Johann Wolfgang Goethe-Universität, Theodor-Stern Kai 7, Haus 75A, D-6000 Frankfurt 70, Germany and [2]Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 Rue René Descartes, 67084 Strasbourg cédex, France

[3]Corresponding author

Communicated by R.Kaptein

The structure of the 84 residue DNA binding domain of the *Escherichia coli* LexA repressor has been determined from NMR data using distance geometry and restrained molecular dynamics. The assignment of the ^{1}H NMR spectrum of the molecule, derived from 2- and 3-D homonuclear experiments, is also reported. A total of 613 non-redundant distance restraints were used to give a final family of 28 structures. The structured region of the molecule consisted of residues 4–69 and yielded a r.m.s. deviation from an average of 0.9 Å for backbone and 1.6 Å for all heavy atoms. The structure contains three regular α-helices at residues 6–21 (I), 28–35 (II) and 41–52 (III), and an antiparallel β-sheet at residues 56–58 and 66–68. Helices II and III form a variant helix–turn–helix DNA binding motif, with an unusual one residue insert at residue 38. The topology of the LexA DNA binding domain is found to be the same as for the DNA binding domains of the catabolic activator protein, human histone 5, the HNF-3/fork head protein and the *Kluyveromyces lactis* heat shock transcription factor.

Key words: distance geometry/molecular dynamics/NMR spectrum assignment/protein structure

Introduction

The LexA repressor from *Escherichia coli* is a 202 residue protein that regulates transcription of ~20 genes known as SOS genes, which are involved in DNA repair and replication, mutagenesis and cell division [for reviews see Little and Mount (1982), Walker (1984), Little (1991), Schnarr *et al.* (1991) and Schnarr and Granger-Schnarr (1993)]. LexA dimerizes with a rather weak dissociation constant of 50 μM (Schnarr *et al.*, 1985, 1988) to form an active dimer that binds to a 16 bp palindromic sequence (Wertman and Mount, 1985). The molecule is divided into two domains with dimerization mediated by the C-terminal domain (Schnarr *et al.*, 1988). As part of the derepression process *in vivo* (or at a pH ≥9 *in vitro*) the molecule undergoes a (stimulated) autocleavage reaction at residues 84–85 (Little, 1984), in a flexible hinge (Little and Hill, 1985) between the two domains. The N-terminal domain of the molecule, LexA 1–84, binds strongly to at least one of the SOS operators (Bertrand-Burggraf *et al.*, 1987), and forms the same contacts with DNA as the complete protein according to methylation protection and hydroxyl radical footprinting studies (Hurstel *et al.*, 1986, 1988).

The possible presence of the helix–turn–helix (HTH) DNA binding motif in LexA has been the subject of discussion. Notably, Pabo and Sauer (1984) predict that LexA should contain a HTH motif at residues 28–47, while Dodd and Egan (1987, 1990) find a low probability for an HTH motif starting at residue 5, but predict that the motif is most probably absent in LexA. The HTH DNA binding motif consists of an N-terminal recognition helix that lies along the major groove of DNA forming base-specific contacts, preceded by a turn and a short anchoring helix that crosses the major groove contacting the phosphate backbone. The motif is known from a large number of DNA binding proteins [see, for example, Dodd and Egan (1990) for a recent survey], and the structure of several protein–DNA complexes of this type has been determined [for reviews see Steitz (1990), Pabo and Sauer (1992) and Burley (1994)]. LexA contains a concentration of activity-altering mutations in residues 39–46 (Oertel-Bucheit *et al.*, 1990, 1992; Thliveris *et al.*, 1991; Thliveris and Mount, 1992). This strongly suggests that this part of the molecule is in contact with DNA, as would be expected from the proposal of Pabo and Sauer. Especially pertinent results were obtained by Thliveris and Mount (1992) who reported a mutation Glu45→Lys that alters the half-operator recognized by LexA from the wild type CTGTXXXX· to GTGTXXXX· (the raised dot indicates the position of the dyad axis), as well as several altered-specificity mutants with two or three changes among residues 40, 41, 42, 44 and 45. A preliminary report (Lamerichs *et al.*, 1989) has shown that a distorted HTH motif exists at residues 28–47. In this paper the 3-D structure of the DNA binding domain of LexA will be presented.

Results

Assignment

The spectrum of the LexA DNA binding domain (LexA DBD) had been partially assigned previously (Lamerichs *et al.*, 1989). In this work these assignments were confirmed to be valid at pH 5.5 and extended to a virtually complete assignment of the molecule. 2-D double quantum-filtered correlation spectroscopy (DQF-COSY) and 2-D total correlation spectroscopy (TOCSY) spectra in H_2O yielded crosspeaks for essentially all pairs of NH and Hα atoms, as well as crosspeaks between most of the side-chain protons. The remaining side-chain protons were identified later with the help of TOCSY spectra in D_2O.

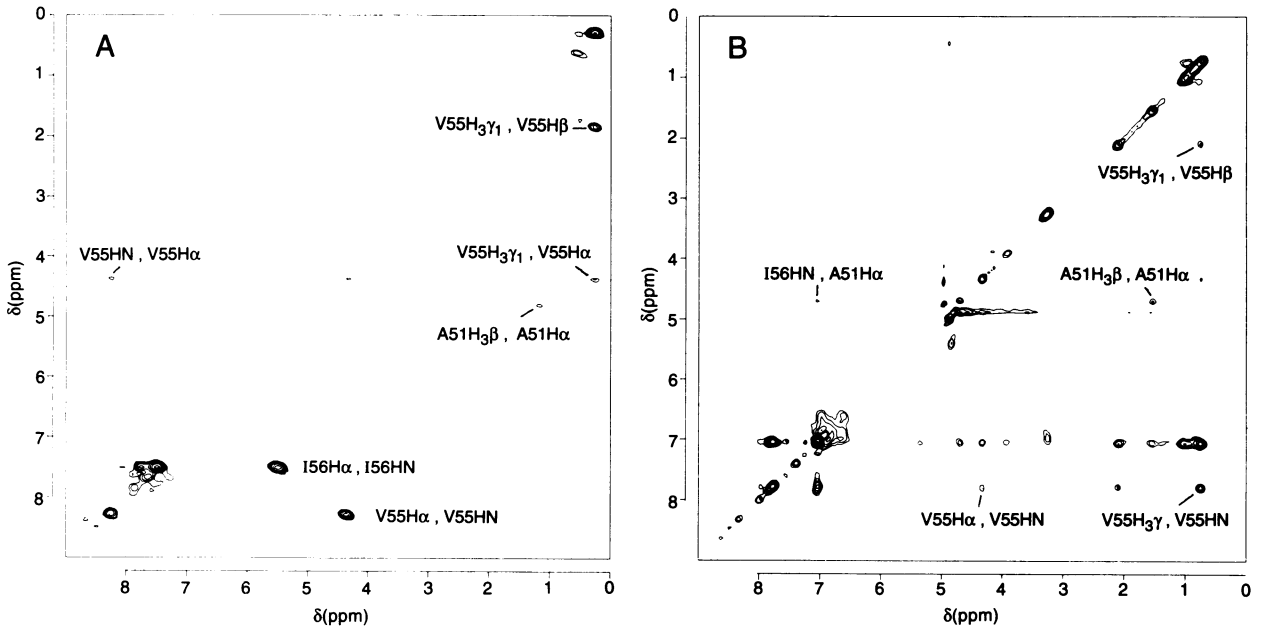

Fig. 1. $\omega_1-\omega_2$ planes of homonuclear 3-D spectra at the ω_3 frequency of I56 HN (7.05 p.p.m.). All peaks have ω_3 coordinates assigned to I56 HN. Assignments of ω_1 and ω_2 for non-diagonal peaks are given in the figure. (**A**) TOCSY–NOESY experiment. Since both the HN and Hα frequencies of L14, K53 and V55 overlap, the V55 γCH3–V55 Hα–I56 HN is the only peak in any spectrum to give an unambiguous sequential connection between V55 and I56. (**B**) NOESY–NOESY experiment.

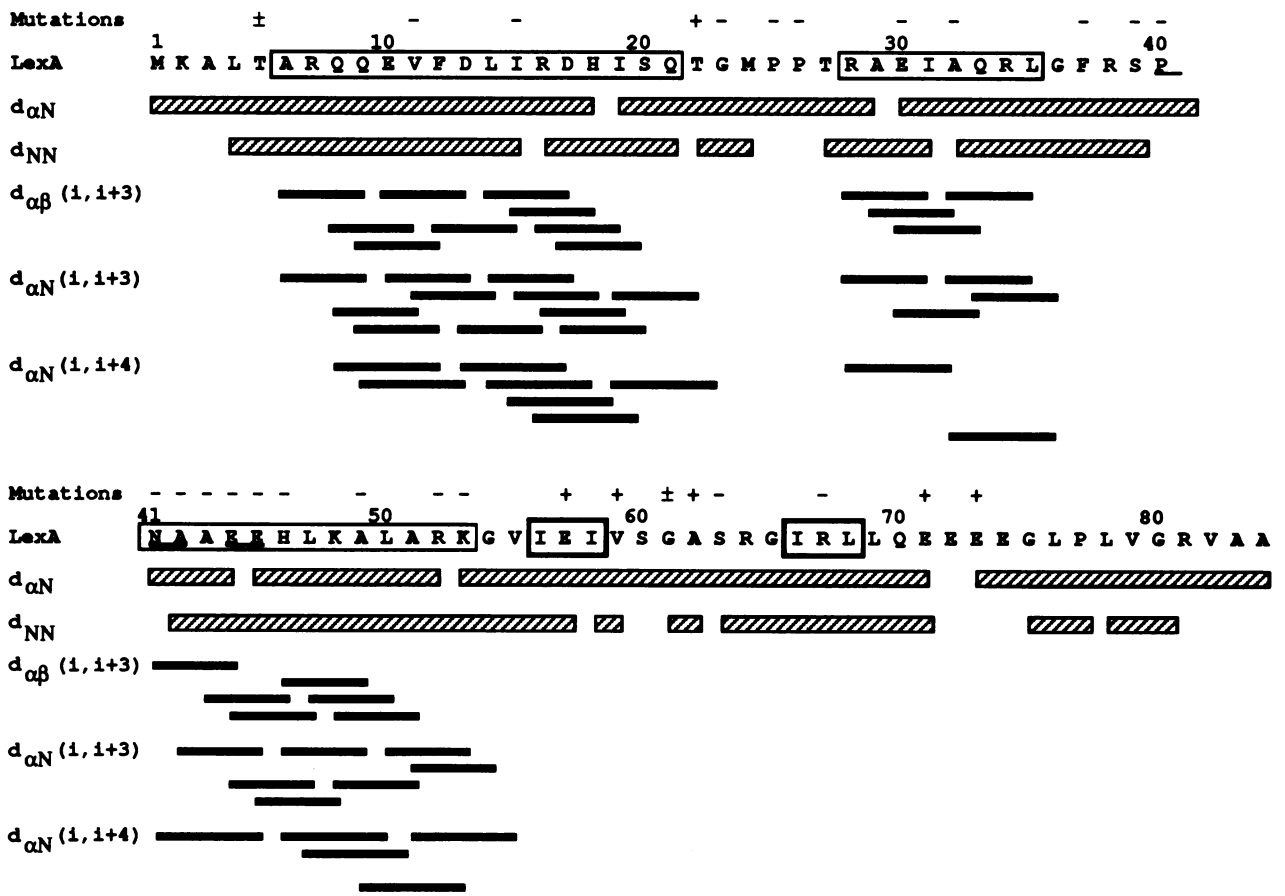

Fig. 2. Sequential assignment, secondary structure and mutations for LexA. Sequential NOEs used to assign LexA are shown as a hatched box and medium-range NOEs as a thick bar connecting the relevant residues. Thin-walled boxes mark residues in α-helices and fat-walled boxes residues in β-sheet strands. A minus sign above a residue indicates that an inactivating mutation has been observed at this residue and a plus sign an activating mutation (Oertel-Bucheit *et al.*, 1990, 1992; Thliveris *et al.*, 1991). Underlined residues show a specificity-changing mutation (Thliveris and Mount, 1992).

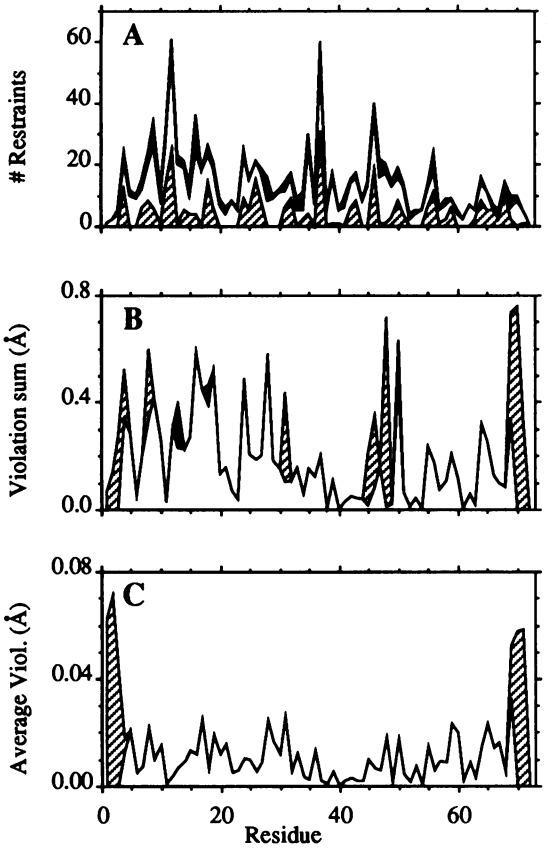

Fig. 3. Restraints and violations per residue for the final set of structures. All restraints are counted twice, once for each of the atoms involved. Additional restraints generated during pseudo-atom correction are included in (**B**) and (**C**), but not in (**A**). (A) Number of long-range (hatched), sequential and medium-range (white), and intraresidue (black) NOE distance restraints per residue. (B) Average sum of restraint violations for each residue. Black areas arise from dihedral restraints, implemented as distance limits. Hatched areas are restraint violations that can be attributed to purely local conformational averaging. These are discussed in the text. (C) Restraint violations averaged over the structures and the number of NOE restraints for each residue. Note the high values found for the poorly defined three residues at either terminus (hatched).

A series of resonances (later found to arise from the C-terminal end of the molecule) were set apart by their very narrow lines, and showed a doubling for many residues, presumably arising from *cis/trans* isomerism at the Leu76–Pro77. Signals arising from the major conformer could be identified easily by their higher intensity (80–90% of the total).

Sequential assignment was carried out using the basic strategy presented in Wüthrich (1986). The sequential connectivities were obtained from 2-D nuclear Overhauser effect spectroscopy (NOESY) spectra (Wüthrich, 1986), as well as from 3-D homonuclear TOCSY–NOESY spectra (Vuister *et al.*, 1988; Padilla *et al.*, 1990) (see Figure 1A). The 3-D TOCSY–NOESY greatly simplified the assignment of nuclear Overhauser effect connectivities (NOE), since the assignment of at least one of the spins involved is unambiguous in many cases, due to the J magnetization transfer step. Also, at a later stage the 3-D NOESY–NOESY experiment (Figure 1B) was used to choose between alternative assignment possibilities. The NOESY–NOESY experiment is more difficult to analyse than the TOCSY–NOESY because of the large number

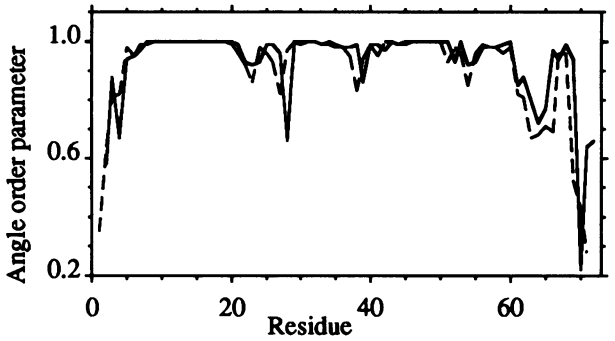

Fig. 4. Angular order parameters for Φ and Ψ angles (full and broken lines respectively) in the final set of structures. Values are calculated according to Hyberts *et al.* (1992).

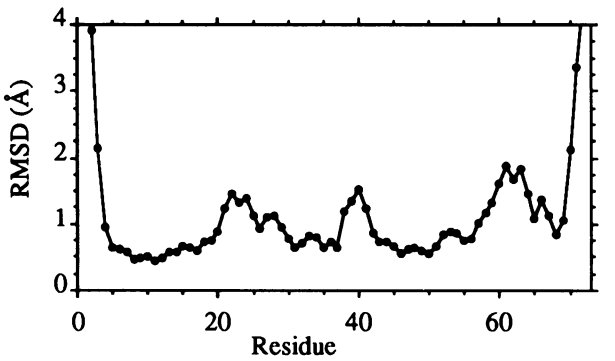

Fig. 5. R.m.s. deviation of Cα atoms in the final set of structures after superposition over Cα of residues 4–69 onto the individual structure closest to structure AV. Note the very high r.m.s. deviation for the terminal residues, and the lower r.m.s. deviation for the α-helices.

of possible alternative assignments for most of the peaks; but the experiment still delivered the assignment of several valuable long-range contacts. In general, the combination of 2-D spectra, with their higher resolution, and 3-D spectra, with the larger information content of 3-D peaks, greatly facilitates the unambiguous assignment of cross-peaks, at least when analysed by an interactive graphics program.

A summary of sequential NOEs is presented in Figure 2. As glutamates 71, 72, 73 and 74 had almost identical chemical shifts, the relative assignment of residues 72 and 73 was performed through an analysis of the largely overlapping $d_{\alpha N}$ (i, i) and $d_{\alpha N}$ (i, j) peaks. Ultimately, 483 out of 498 (97%) of the NH, NH_2 and CH_n resonances were assigned. The list of chemical shifts has been submitted to the BioMagResBank. For Pro40 Hδ only one resonance could be found. Accordingly, the distinction between Hα and Hδ rested solely on the observed chemical shifts (4.31 and 3.99 p.p.m. respectively). The distinction between a *trans* proline (showing $d_{\alpha\delta}$ sequential NOEs) and the alternative *cis* proline (showing $d_{\alpha\alpha}$ sequential NOEs) at residue 40 remains based on the assignments of Pro40 Hα and Hδ.

Restraints

The first set of assigned NOEs gave rise to 552 non-redundant restraints using pseudo-atoms used for all prochiral groups. This set was then refined progressively over several sets of preliminary structures. NOEs involving the His18 ring were persistently violated, and it was

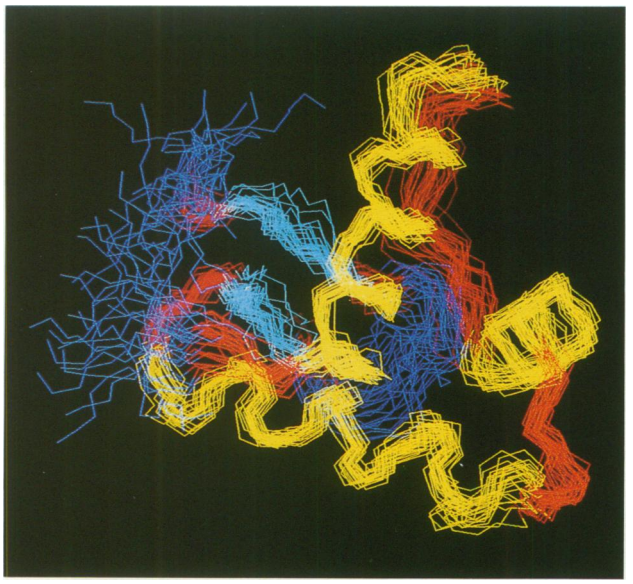

Fig. 6. Final set of structures for the LexA DBD. Helices, yellow; β-sheet, blue-white; connecting loops, red; and poorly defined residues, dark blue. Only residues 1–72 are shown. The anchoring helix of the HTH motif comes out of the paper at the right of the picture, while the recognition helix goes from right to left at the bottom

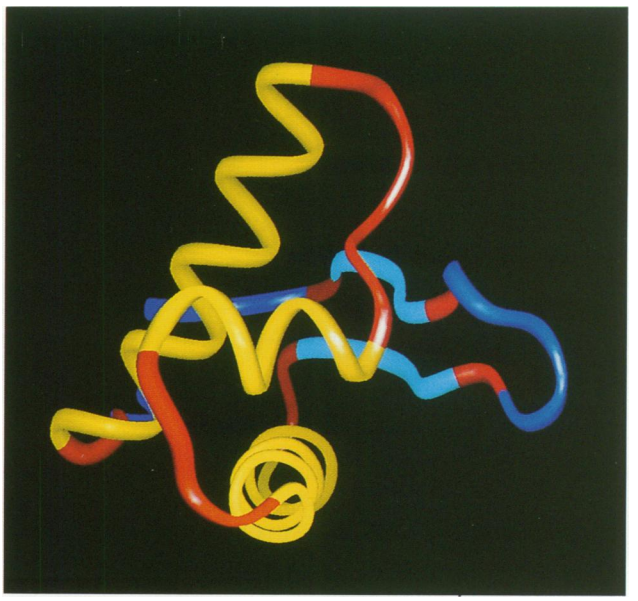

Fig. 7. Structure (AV) of the LexA DBD. Helices, yellow; β-sheet, blue-white; connecting loops, red; and poorly defined residues, dark blue. Only residues 1–72 are shown. The anchoring helix of the HTH motif goes from right to left in the centre of the picture, while the recognition helix is going into the page.

judged that the ring was undergoing rapid rotation ('ring-flip'). Restraints involving it were referred to pseudo-atoms on the axis of rotation accordingly.

The final set of restraints used floating chirality rather than pseudo-atoms where possible, and consisted of 613 independent restraints divided into 151 intraresidue, 138 sequential, 171 medium range (two to four residues apart) and 153 long range. It should be emphasized that restraints that were redundant with respect to the covalent constraints were removed from the restraint set and are not included in any of the numbers quoted. An additional 127 restraints were added as part of the smoothing associated with pseudo-atom corrections. The list of restraints used has been submitted together with the structure to the Brookhaven Protein Data Bank.

Structures

Of 80 embedded structures, the 33 with the highest values of the DGII error function were discarded. The 47 remaining structures were submitted to molecular dynamics.

In the final refined structures, residues 1–3 and 70–72 were characterized by a high local r.m.s. deviation, completely undefined dihedral angles, higher average restraint violations and a high proportion of residues in disallowed regions of the $\Phi-\Psi$ plot (see Figures 3–5). The conformation of these residues was clearly undefined, and violations of restraints involving them were subsequently disregarded. At this point the 14 structures with the highest total energy, including those with the highest sum of violations, were rejected. A further five structures were rejected because they contained a region with a significant deviation in the backbone topology coupled with a local concentration of restraint violations.

The 28 accepted structures, the final set, were super-

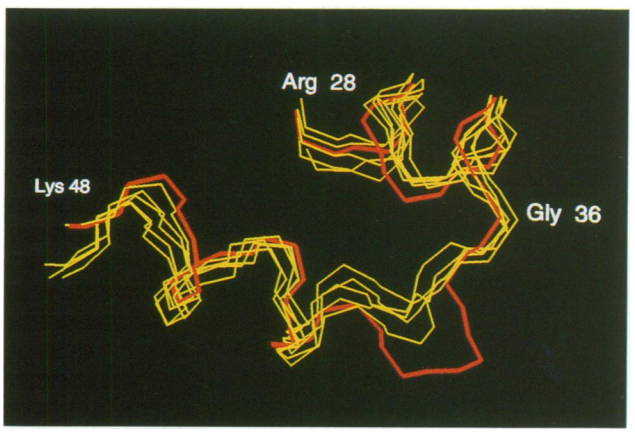

Fig. 9. HTH motif of LexA and selected other proteins superimposed on C, N and Cα of LexA residues 28–37 and 40–48, corresponding to residues 1–10 and 12–20 of the HTH motif in the nomenclature of Pabo and Sauer (1984). LexA is in red, while the other proteins are in yellow. Note the insert in LexA at residue 38/39. Other proteins are (protein data bank names in parentheses): catabolite activator protein (3GAP, B chain; Weber and Steitz, 1987), phage 434 Cro repressor (2CRO; Mondragon *et al.*, 1989), lac repressor headpiece (de Vlieg *et al.*, 1988), λ repressor (1LRD; Jordan and Pabo, 1988) and Trp repressor (2WRP; Otwinowski *et al.*, 1988).

Table I. R.m.s. deviation from average of final structures[a]

R.m.s. deviation over	CN, Cα (Å)	Heavy atoms (Å)
Final set residues 4–55	0.72 (0.15)	1.42 (0.16)
Final set residues 56–69	0.80 (0.26)	1.60 (0.33)
Final set residues 4–69	0.94 (0.19)	1.57 (0.18)
EMAV residues 4–69	0.65	1.19

[a]R.m.s deviations are calculated relative to the average structure AV. Numbers in parentheses are standard deviations.

Table II. Key data for final structures[a]

	Final set	EMAV
Number of structures	28	1
Total energy (kcal/mol)	−48 ± 14	−104
Distance violation sum, residues 4–69 (Å)	6.7 ± 1.4	2.7
Dihedral violation sum (Å)[b]	0.37 ± 0.43	0.0
Ramachandran plot, residues 4–69		
Percentage of residues in favoured region (%)[c]	90.6 ± 3.6	96.4
Residues in forbidden regions[c,d]	57 ± 0.63	0

[a]Numbers in parentheses are standard deviations.
[b]Dihedral restraints were entered as restraints on the distances across the dihedral (see text).
[c]As defined in Morris *et al.* (1992).
[d]Includes 'generously allowed' regions.

imposed and averaged to create the AV (average) structure, which was further energy minimized to yield the energy-minimized average (EMAV) structure.

Discussion

Structure of the LexA DBD

The final set of 28 LexA structures is shown in Figure 6, while the average backbone structure is given in Figure 7. The LexA DBD contains three α-helices extending over residues 6–21 (helix I), 28–35 (helix II) and 40–53 (helix III), according to the Kabsch−Sander secondary structure definition rules, as incorporated in the program PROCHECK (Morris *et al.*, 1992). The $i - (i + 4)$ hydrogen bonds characteristic of an α-helix span the same residues, except that the regular part of helix III extends only over residues 41–52 according to this criterion, residue 39 C=O and 54 NH being involved in $i - (i + 3)$ hydrogen bonds more characteristic of a 3_{10} helix. An antiparallel β-sheet is found at residues 56–58 and 66–68 according to the Kabsch−Sander rules, and the presence of hydrogen bonds involving the NH groups of residues 57, 59, 67 and 69 confirms this finding. LexA DBD contains three tight turns according to the definition of Wilmot and Thornton (1990). All of them involve loop residues with less precisely defined main-chain dihedral angles. Two, at residues 37–40 and 62–65, fit the description of a type VIII turn which is not normally associated with a hydrogen bond. The last turn, at residues 59–62, could be described as a distorted type II turn based on the average values for the (poorly defined) dihedral angles. The hydrogen bond typical of such turns is not observed, but a hydrogen bond from Gly61 NH to Val59 C=O is observed in >80% of the structures.

The degree of definition of the various parts of the structure can be judged from the data in Figures 4–6. The α-helices represent the best defined parts of the structure, with average pairwise r.m.s. deviations <0.7 Å for the Cα atoms, and angular order parameters close to 1.0. At the other extreme residues 1–3 and 70–72 are completely disordered with respect to both angles and position. The loops between the secondary structure elements are less well defined than the helices. In particular the loop between the two β-strands is more poorly defined than the rest, with order parameters for the backbone dihedrals <0.8 and a higher coordinate r.m.s. deviation. The β-sheet has a special position, in that the angular order parameters are

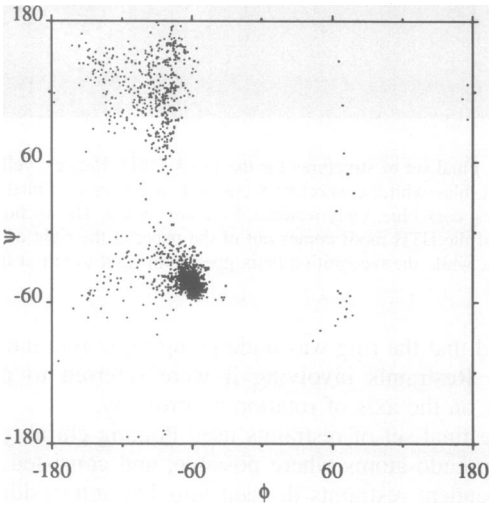

Fig. 8. Ramachandran plot for residues 4–69 from the final set of LexA. Only non-glycine residues are shown. Residues with 0° <Φ <90° are Arg28 (six cases, Ψ ~−70°), Ser63 (three cases, Ψ ~−70°), Leu4 (four cases), Arg64 (four cases), Ala6, Arg52 and Ala62.

≥0.95 for both strands, whereas the strand at residues 66–68 has Cα r.m.s. deviations as high as 1 Å. This is due chiefly to uncertainty in the positioning with respect to the rest of the molecule, rather than to local disorder in the sheet. The r.m.s. deviation calculated for the molecule as a whole reflects this uncertainty as well as the local uncertainty in the structures, as may be clearly seen by comparing the different backbone r.m.s. deviation values in Table I.

The overall quality of the structures can be judged from the data in Table II. It is worth noting that the EMAV structure is better than the other structures according to all the criteria used, while being as close to the AV structure as any other single structure. This is the behaviour expected for a set of structures clustered around a single average structure. The distribution of the backbone dihedral angles is shown in the Ramachandran plot in Figure 8. The quality of the structures was evaluated with the program PROCHECK (Morris *et al.*, 1992). The quality of the stereochemistry in the final set of structures is excellent, as would be expected from a structure refined by molecular dynamics simulation.

Hydrogen bonds were judged to be present where the H−acceptor distance was <2.5 Å and the

donor−H−acceptor angle was >90° (Baker and Hubbard, 1984) in a final energy minimized refined structure. Only hydrogen bonds observed in 66% or more of the structures in the final set were accepted as reliable. The use of molecular dynamics *in vacuo* during refinement will distort the hydrogen bonding pattern, especially for mobile side chains, introducing a bias towards intramolecular hydrogen bond formation. Nevertheless, hydrogen bonds that show up in most of the structures are likely to be representative of the solution structure.

Outside the secondary structure (see above) the most reliable hydrogen bonds were to the NH and C=O groups exposed at the ends of α-helices. Although all three helices are sandwiched between an N-terminal serine or threonine and a C-terminal glycine, compatible with the formation of classic helix caps (Richardson and Richardson, 1988), the corresponding hydrogen bonds are formed only for the C-terminus of helix III. This does not exclude that the residues in question may be important for folding (Presta and Rose, 1988). The N-terminus of helix I is uncapped, but Thr5 NH and Thr5 side-chain OH form hydrogen bonds to Gln8 Oε. It has been found (Oertel-Bucheit *et al.*, 1992) that the mutation Thr5 → Arg, but not Thr5 → Ser, destroys LexA activity *in vivo*, but that mutations in unrelated parts of the LexA DBD can restore activity. It is possible that the Thr5−Gln8 interaction plays a role in stabilizing the structure around helix I. The N-termini of helices II and III both contain $i - (i + 3)$ hydrogen bonds (starting at residues 27 and 39), but lack side-chain−backbone hydrogen bonds. The C-terminus of helix I was stabilized by side-chain OH groups acting as hydrogen bond donors, while the C-terminus of helix II had an irregular bonding pattern, with hydrogen bonds 35-NH−31-C=O, 36-NH−32-C=O and 37-NH−32-C=O formed in 57, 39 and 96% of cases, respectively. An interesting hydrogen bond was formed between Arg7 Hη and Leu35 C=O. This hydrogen bond is likely to be genuine, considering that the side chain of Arg7 is fixed in place by NOEs observed to Phe37 and His46, and serves to position the Arg7 guanidinium group close to the DNA binding HTH part of the molecule.

Restraints and violations

The total number of NOE restraint violations for the 28 final set structures is 6.7 ± 1.4 Å on average, when restraints involving the ill-defined residues 1–3 and 70–72 are excluded. This corresponds to $9 \pm 2 \times 10^{-3}$ Å for each of the 719 restraints involved. The range 0.7–1.1 Å contains the 10 largest restraint violations, with a further 33 in the range 0.5–0.7 Å. This is <0.4 respectively 1.2 per structure. Only two restraints have average distances longer then the restraint distance, in both cases by ≤0.2 Å. The distribution of restraint violations (see Figure 3B and C) is clearly not uniform, and the different concentrations of violations have different implications for the overall quality of the structures.

The few short-range restraints involving the disordered residues 1–3 and 70–72 show very high average violations (Figure 3C). This surprising result is not caused by incompatible restraints. The most likely explanation is that the structure generation algorithm is unable to get close to the global energy minimum, given the many local minima of similar energy available to unconstrained residues. Since the distance restraint force constant has been chosen deliberately to give comparable magnitudes to restraint and energy terms, favourable non-bonded interactions in an arbitrary local minimum could easily compensate for restraint violations. It is worth noting that the structures contain on average one presumably artefactual NH(i)−C=O(i − 2) hydrogen bond at both the C- and N-terminus. Similar effects have also been observed in other cases where poorly defined residues were simulated *in vacuo*, e.g. Fogh *et al.* (1990).

The excess violations at residues 8, 31, 46 and 48 could all be the result of conformational averaging involving only a single side chain. The concentration of violations at residues 16–25 could, if significant, have a similar origin, but the averaging involved would have to include the conformation of the whole loop at residues 23–27.

The floating chirality used for prochiral groups yielded seven groups with identical prochirality in all 47 fully refined structures, and thus with implicit stereospecific assignments. All other prochiral groups were found in either conformation in at least 25% of even the 28 final set structures. The groups with well-determined prochirality encompassed five of the six βCH_2 groups with the χ_1 angle restrained, as well as the isopropyl group of Val11. These stereospecific assignments are likely to be correct, since they coincided with the assignments already deduced (with varying degrees of certainty) from short-range restraints and the topology found in preliminary structures. It should be emphasized that these stereospecific assignments were not used as input in the structure generation process, a practice explicitly discouraged by Havel (1991) because of the risk of introducing incorrect assignments. The usefulness of floating chirality assignments has been questioned by Havel (1991) for cases where the β-methylene protons could be stereospecifically assigned. However, where the data do not allow such assignments to be performed reliably, as in the present case, floating chirality allows what may well be a significant tightening of the restraints without requiring the introduction of potentially erroneous assumptions.

The HTH motif

The structure of the LexA DBD has been the focus of some interest given the different opinions about the nature and location of the DNA binding motif (Pabo and Sauer, 1984; Dodd and Egan, 1987, 1990). A partial model based on preliminary NMR results from our laboratory (Lamerichs *et al.*, 1989) suggested that a HTH motif was present at residues 28–47, as predicted by Pabo and Sauer (1984), but that the details of the structure might be atypical. With the complete and detailed LexA structure presented for the first time in this work, it is clear that LexA contains a HTH motif at residues 28–48. The observed structure falls completely within the parameters of a classic HTH motif (with the important difference that the LexA HTH motif contains a one residue insert in the turn between the helices, at residue 38). The remarkable similarity between the HTH motif of LexA and some acknowledged HTH DNA binding proteins can be seen in Figure 9. The r.m.s. deviation for C, N and Cα of LexA residues 28–37 and 40–48 [residues 1–10 and 12–20 of the motif in the nomenclature of Pabo and Sauer (1984)]

is 0.96 ± 0.16 Å comparing LexA with the proteins depicted in Figure 9, as compared with 0.87 ± 0.16 Å between the proteins themselves. The side-chain packing around the HTH motif in LexA corresponds to what is found in other HTH proteins. Notably, the 'helix-positioning' contacts between Ile31, Ala32 and Ala43 in LexA are very similar to the equivalent contacts in, for example, lac repressor or catabolic activator protein (CAP; see legend to Figure 9 for protein names and references). The side chain of Phe37 lies on top of these residues like a lid, contacting residues on both helices, but does not penetrate between the helices. Another contact between the helices is formed at Ala32–Pro40.

Homology to other proteins

Recently, a structure of HNF-3, the HNF-3/fork head DNA binding domain, complexed with DNA (Clark *et al.*, 1993), and new structures of GH5 (the human histone 5 DNA binding domain; Ramakrishnan *et al.*, 1993) and HSF (the *Kluyveromyces lactis* heat shock transcription factor; Harrison *et al.*, 1994) have appeared in the literature. All three proteins show the same topology as LexA, as is also found for CAP (Weber and Steitz, 1987) and BirA (the DNA binding domain of the *E.coli* biotin repressor; Wilson *et al.*, 1993), and the proteins superimpose remarkably well over all three helices as well as the β-sheet (results to be published elsewhere). Of the six proteins, CAP and BirA contain a classic HTH motif, while GH5, HNF-3 and HSF, like LexA, have an insert in the turn between the helices. GH5, HNF-3 and HSF also have differences in the position and orientation of the first helix in the HTH motif relative to CAP, BirA and LexA. The other main difference between the proteins is in the vicinity of the β-sheet. GH5 and HNF-3 have a three-stranded antiparallel β-sheet with the third strand formed by the loop between helices I and II, whereas in LexA and BirA this loop runs approximately parallel to the β-sheet strands without forming part of the sheet. In CAP and HSF the loop forms two strands of the complete four-stranded sheet. The loop between the ends of the β-sheet strands in LexA is situated on the same face of the molecule as the N-terminal end of helix II. It has its counterpart in GH5, HNF-3 and HSF, but is reduced to a turn in CAP and BirA. As is the case for LexA, the loop is flexible in GH5, where it is found in two different conformations in the two molecules in the unit cell, and in HSF, where the central part of the loop shows no discernible electron density. The similarities between LexA DBD and other HTH DNA binding proteins, as well as the impact of the LexA structure on the sequence homology recognition pattern for HTH proteins, will be discussed in more detail elsewhere.

Mutations and DNA binding

Positions in LexA 1–72 where mutations are known to alter the activity or specificity of the molecule are shown in Figure 2. Of the residues involved, V11, I15, P25, P26, F37 and A43 are buried in the core of the molecule. Mutations at these positions are likely to affect the overall structure rather than the DNA binding site, as is the case for G23 which forms part of a turn. T5 may be involved in stabilizing the N-terminus of helix I, but could also take part in a possible interaction between the N-terminal end of helix I and the DNA backbone. Such an interaction is found for CAP and HNF-3, the two proteins with a similar topology where the structure of the complex with DNA is known (Schultz *et al.*, 1991; Clark *et al.*, 1993). Residues S39, N41, E44, E45, R52 and K53 are completely solvent-exposed on the surface of helix III. Since mutations in residues 40–45 can alter the binding specificity of LexA (Thliveris and Mount, 1992), these residues must be involved in base-specific contacts with DNA. Residue 52 is also known from the photo-crosslinking experiment of Dumoulin *et al.* (1993) to face the DNA upon binding. These data lead to the conclusion (Dumoulin *et al.*, 1993) that on binding, helix III of LexA is inserted in the major groove of DNA in such a way that the N-terminus is directed towards the palindromic centre of the dimeric operator.

The mutations in residues 57–67, both activity-suppressing and activity-enhancing, involve surface-exposed residues at a distance from the recognition helix (helix III). In both CAP and HNF-3 a residue in this region makes a contact with the DNA backbone. A similar interaction might well be present in LexA where, given the known orientation of the LexA DBD relative to the dyad axis of the operator, it could account in part for the protected phosphates found at the dyad axis (Hurstel *et al.*, 1988). It should be noted that such an interaction would place residues 59–63 in close proximity to the dyad axis, so that some of the activity-modifying mutations found here (Figure 2) could also be explained as the result of interactions with the same region of the other LexA monomer rather than protein–DNA interactions.

The DNA complexes of the topologically similar CAP and HNF-3 are similar in their contacts to the DNA backbone, but differ markedly in the degree of bending induced in the DNA and in which part of the recognition helix contacts the DNA base-pairs. Considering that HNF-3 does not bind as a dimer and that the monomers of CAP are oriented opposite to those of LexA, the LexA–DNA complex may well turn out to be significantly different from DNA complexes of HNF-3 and CAP, in spite of the close similarities in topology and structure between the three DNA binding domains.

Materials and methods

NMR experiments

The LexA DBD was isolated as described previously (Hurstel *et al.*, 1986) and transferred by repeated ultrafiltration into a 10 mM NaP_i buffer (95% H_2O/5% D_2O), pH 5.5, 300 mM NaCl, with a few grains of NaN_3. The final sample contained ~5 mM LexA DBD. For experiments in D_2O, the sample was lyophilized and redissolved in pure D_2O.

^1H NMR spectra were recorded at 600 MHz on Bruker AM and AMX spectrometers. All spectra were acquired with solvent suppression by presaturation during the relaxation delay and NOESY mixing times, and with sweep widths of 10–11 p.p.m. Experiments were carried out at either 293 or 300 K. 2-D spectra were acquired with at least 400 (t_1)×2048 (t_2) real points, 32–56 scans per t_1-value and a recycle delay of 1.1–1.4 s. The DQF-COSY (Ernst *et al.*, 1987) was recorded at t_1 and t_2 acquisition times of 0.041 and 0.655 s, respectively. TOCSY spectra were recorded using the 'clean' MLEV or DIPSI-2 sequence (Griesinger *et al.*, 1988; Shaka *et al.*, 1988) with mixing times of 22–73 ms. NOESY spectra (Ernst *et al.*, 1987) were acquired with mixing times from 50 to 200 ms, quantitative distances being taken from a build-up series with mixing times of 50, 100, 150 and 200 ms in 95% H_2O/5% D_2O at 293 K. A 3-D TOCSY–NOESY spectrum using the 'clean' MLEV mixing sequence was recorded using the pulse sequence described in Padilla *et al.* (1990) with the TOCSY trim pulses omitted.

Mixing times of 43 (TOCSY) and 150 ms (NOESY) were used, with a size of 128 (t_1)×215 (t_2)×512* (t_3) points, where * denotes complex points, eight transients per free induction decay (FID) and a relaxation delay of 0.6 s. The 3-D NOESY–NOESY spectrum (Boelens *et al.*, 1989) was recorded in 95% H_2O/5% D_2O at 293 K, with both mixing times of 120 ms, a size of 160 (t_1)×153 (t_2)×512* (t_3) points, eight transients per FID and a relaxation delay of 0.8 s.

2-D spectra were processed using the program TRITON to a size of 1024 (ω_1)×2048 (ω_2) real points, while both 3-D spectra were processed to a size of 256 (ω_1)×256 (ω_2)×512 (ω_3) real points. For the 3-D NOESY–NOESY spectrum, t_1 and t_2 were extended to 210 points by linear prediction (Press *et al.*, 1992) immediately before Fourier transformation.

Generation of restraints

Quantitative distance restraints were calculated from NOE build-up rates at $t = 0$, obtained by fitting the crosspeak volumes from a NOESY build-up series in H_2O to a function of the form $I = I_0[1 - \exp(-kt)]$ using least-squares minimization. $d_{NN}(i, i + 1)$ distances from the helical parts of LexA DBD were used for calibration (theoretical value 2.8 Å; Wüthrich, 1986). As a check for this approach we determined the following averaged distances for the helical residues (theoretical values in parentheses): $d_{N\alpha}(i, i)$ 2.87 (2.7) Å, $d_{\alpha N}(i, i + 1)$ 3.60 (3.5) Å, $d_{\alpha N}(i, i + 3)$ 3.44 (3.4) Å. Upper distance restraint limits were set by lengthening all determined distances by 20% to allow for spin diffusion, molecular motion, etc., taking the maximum rather than the average distance from symmetry-equivalent peaks. Distance restraints involving methyl groups were increased by 7% as a partial compensation for the higher intensity expected for these groups. NOEs that were observed at a mixing time of ≤150 ms in any spectrum but could not be quantified were given an upper distance limit of 5.5 Å. All lower distance limits were set to the sum of the van der Waals radii (0.95 Å for H, 1.45 Å for C). The distance restraints used for structure generation are summarized in Figure 3. Pseudo-atom corrections (Wüthrich *et al.*, 1983) were not used for prochiral groups in general, but only where resonances were degenerate.

For six residues where the relative intensity of DQF-COSY and NOESY crosspeaks to Hβ protons showed that one proton (e.g. Hβ1) was *trans* to Hα, the distance restraints $r_{H\beta1-H\alpha} > 3.0$ Å, $r_{H\beta2-H\alpha} < 2.85$ Å and $r_{C\gamma-H\alpha} < 3.3$ Å were introduced.

During distance geometry calculations only, extra restraints (1.8 Å $< r_{H-O} < 2.2$ Å, 2.7 Å $< r_{N-O} < 3.2$ Å) were added for seven α-helical hydrogen bonds where both a $d_{\alpha N}(i, i + 4)$ and either a $d_{\alpha N}(i, i + 3)$ or a $d_{\alpha\beta}(i, i + 3)$ connectivity was observed. The hydrogen bonds found involved residues 9, 13; 13, 17; 14, 18; 15, 19; 16, 20; 46, 50; and 47, 51. To improve convergence of the distance geometry-embedding algorithm (Havel and BIOSYM Technologies, 1992), a series of redundant lower distance limits were likewise added for the α-helices found in the preliminary structures. Distances for N (respectively C and Cα) in helical residues were set to, for example, $r_{Ni-Nj} > 1.4 \times (j - i) + 1.0$ Å, $j - i > 5$.

Structure generation

For residues 73–84, no long-range and only a few medium-range NOEs were observed. Furthermore, these residues gave rise to resonance lines much sharper than the rest of the molecule. Therefore it is supposed that these residues undergo fast random motion, and that the C-terminus does not have a single well-defined structure. Accordingly, structures of the LexA DBD were generated only for residues 1–72.

For the generation and analysis of structures, version 2.2.0β of the InsightII package (BIOSYM Technologies Inc., San Diego, CA) was used. The metric matrix distance geometry program DGII (Havel and BIOSYM Technologies Inc., 1992) from the package was used in a modified form allowing 'floating chirality' for prochiral groups. Restraints to individual prochiral atoms were entered as such, and the chirality of the group in question was determined by the program during the optimization phase. The chirality of prochiral groups, like the coordinates of the structures, were thus chosen as a random sampling of the conformational space allowed by the restraints. Tetrangle smoothing for sequential pairs of residues was used in addition to triangle smoothing in generating the distance matrix. Structures were embedded by prospective metrization in four dimensions using a uniform probability distribution for selecting trial distances. The fit of the embedded structures was improved using 10 Guttman transformations, with distances weighted proportionally to the inverse of the squared range plus the squared average distance. For optimization, the structures were submitted to 15 000 steps of simulated annealing in four dimensions using an initial energy of 2500 kcal/mol, a maximum temperature of 200 K, a time step of 0.2 ps and atomic masses set to 1000 Da. Finally, the structures were submitted to 250 steps of conjugate gradient minimization.

The structures were refined further by restrained energy minimization and molecular dynamics using version 2.8 of Discover (BIOSYM Technologies Inc.). The protocol consisted of a minimization phase (I), followed by dynamics at 600 K (II), cooling to 300 K (III), dynamics at 300 K (IV) and a final minimization phase (V). The CVFF force field was used, with cross-correlation terms, Morse potentials and electrostatic charges. To avoid the very large forces associated with net charges on the amino acids, all charged residues were converted to a neutral form and a fixed value was used for the dielectric constant. Throughout the calculations all physical energy terms were given a weighting factor of 1, while the distance restraint force constant was set to k_{dis} = 5 kcal.mol^{-1}.Å^{-2} ($E = k_{dis} \cdot \Delta r^2$). Peptide bonds were forced to be *trans* with force constants of 1.0–2.0 kcal.mol^{-1}.rad^{-2}. For non-bonded interactions a double distance cut-off was used. An inner cut-off of 8.5 Å determined which atom–atom interactions were calculated at every step, while forces between atoms within the outer cut-off of 13.5 Å were calculated every 20 steps when the neighbour list was updated, and assumed to be constant between updates. The weighting factor for non-bonded interactions was changed smoothly from 1.0 to 0.0 starting at the cut-off distance less 1.5 Å.

The phase I energy minimization consisted of (i) 20 steps of steepest descent minimization with quartic non-bonded repulsion terms, followed by (ii) 180 steps of steepest descent and (iii) 1200 steps of conjugate gradient minimization. In phase II, the molecules were simulated for 1000 steps of 0.2 fs each, strongly coupled to a temperature bath at 600 K. Cooling to 300 K (III) took place over 3000 steps of 0.5 fs using weak coupling to a reference temperature, with temperature reductions of 13% followed by 600 steps of equilibration. After cooling the structures were simulated through 3000 steps of 1 fs with weak coupling to a 300 K bath (IV). The final structure was energy minimized for 1800 steps (V), alternating between 200 steps of steepest descent and 400 steps of conjugate gradient minimization.

The final set of structures used to represent the conformation of the LexA DBD were now selected, superimposed over the C, N and Cα atoms of residues 4–69 and averaged to give the average structure, AV. This structure was first energy minimized for 400 steps using a quartic repulsion potential without electrostatic charges, then minimized for 3600 steps as in phase V above to give the EMAV structure.

Acknowledgements

This work was supported by an EC grant ST2J-0291-C. H.R. acknowledges a grant from the Deutsche Forschungsgemeinschaft (Ru 145/11-1).

References

Baker,E.N. and Hubbard,R.E. (1984) *Prog. Biophys. Mol. Biol.*, **44**, 97–179.

Bertrand-Burggraf,E., Hurstel,S., Daune,M. and Schnarr,M. (1987) *J. Mol. Biol.*, **105**, 293–302.

Boelens,R., Vuister,G.W., Koning,T.M. and Kaptein,R. (1989) *J. Am. Chem. Soc.*, **111**, 8525–8526.

Burley,S.K. (1994) *Curr. Opin. Struct. Biol.*, **4**, 3–11.

Clark,K.L., Halay,E.D., Lal,E. and Burley,S.K. (1993) *Nature*, **364**, 412–420.

de Vlieg,J., Scheek,R.M., van Gunsteren,W.F., Berendsen,H.J.C., Kaptein,R. and Thomason,J. (1988) *Proteins*, **3**, 209–218.

Dodd,I.B. and Egan,J.B. (1987) *J. Mol. Biol.*, **194**, 557–564.

Dodd,I.B. and Egan,J.B. (1990) *Nucleic Acids Res.*, **18**, 5019–5026.

Dumoulin,P., Oertel-Buchheit,P., Granger-Schnarr,M. and Schnarr,M. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 2030–2034.

Ernst,R.R., Bodenhausen,G. and Wokaun,A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Clarendon Press, Oxford, UK.

Fogh,R.H., Kem,W.R. and Norton,R.S. (1990) *J. Biol. Chem.*, **265**, 13016–13028.

Griesinger,C., Otting,G., Wüthrich,K. and Ernst,R.R. (1988) *J. Am. Chem. Soc.*, **110**, 7870–7872.

Harrison,C.J., Bohm,A.A. and Nelson,H.C.M. (1994) *Science*, **263**, 224–227.

Havel,T.F. (1991) *Prog. Biophys. Mol. Biol.*, **56**, 43–78.

Havel,T.F. and BIOSYM Technologies (1992) *NMRchitect Users Guide.*

BIOSYM Technologies Inc., San Diego, CA, USA.
Hurstel,S., Granger-Schnarr,M., Daune,M. and Schnarr,M. (1986) *EMBO J.*, **5**, 793–798.
Hurstel,S., Granger-Schnarr,M. and Schnarr,M. (1988) *EMBO J.*, **7**, 269–275.
Hyberts,S.G., Goldberg,M.S., Havel,T.F. and Wagner,G. (1992) *Protein Sci.*, **1**, 736–751.
Jordan,S.R. and Pabo,C.O. (1988) *Science*, **242**, 893–899.
Lamerichs,R.M.J.N., Padilla,A., Boelens,R., Kaptein,R., Ottleben,G., Rüterjans,H., Granger-Schnarr,M., Oertel,P. and Schnarr,M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 6863–6867.
Little,J.W. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 1375–1379.
Little,J.W. (1991) *Biochimie*, **73**, 411–422.
Little,J.W. and Hill,S.A. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 2301–2305.
Little,J.W. and Mount,D.W. (1982) *Cell*, **29**, 11–22.
Mondragon,A., Wolberger,C. and Harrison,S.C. (1989) *J. Mol. Biol.*, **205**, 179–188.
Morris,A.L., MacArthur,M.W., Hutchinson,E.G. and Thornton,J.M. (1992) *Proteins*, **12**, 345–364.
Oertel-Bucheit,P., Lamerichs,R.M.J.N., Schnarr,M. and Granger-Schnarr,M. (1990) *Mol. Gen. Genet.*, **223**, 40–48.
Oertel-Bucheit,P., Porte,D., Schnarr,M. and Granger-Schnarr,M. (1992) *J. Mol. Biol.*, **225**, 609–620.
Otwinowski,Z., Schevits,R.W., Zhang,R.-G., Lawson,C.L., Joachimiak,A., Marmorstein,R.Q., Luisi,B.F. and Sigler,P.B. (1988) *Nature*, **335**, 321–329.
Pabo,C.O. and Sauer,R.T. (1984) *Annu. Rev. Biochem.*, **53**, 293–321.
Pabo,C.O. and Sauer,R.T. (1992) *Annu. Rev. Biochem.*, **61**, 1053–1095.
Padilla,A., Vuister,G.V., Boelens,R., Kleywegt,G.J., Cave,A., Parello,J. and Kaptein,R. (1990) *J. Am. Chem. Soc.*, **112**, 5024–5030.
Press,W.H., Flannery,B.P., Teukolsky,S.A. and Vetterling,W.T. (1992) *Numerical Recipes (FORTRAN Version)*. Cambridge University Press, Cambridge, UK, pp. 444–447.
Presta,L.G. and Rose,G.D. (1988) *Science*, **240**, 1632–1641.
Ramakrishnan,V., Finch,J.T., Graziano,V., Lee,P.L. and Sweet,R.M. (1993) *Nature*, **362**, 219–223.
Richardson,J.S. and Richardson,D.C. (1988) *Science*, **240**, 1648–1652.
Schnarr,M. and Granger-Schnarr,M. (1993) In Eckstein,C.J.F and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Vol. 7. Springers Verlag, Berlin, Germany.
Schnarr,M., Pouyet,J., Granger-Schnarr,M. and Daune,M. (1985) *Biochemistry*, **24**, 2812–2818.
Schnarr,M., Granger-Schnarr,M., Hurstel,S. and Pouyet,J. (1988) *FEBS Lett.*, **234**, 56–60.
Schnarr,M., Oertel-Buchheit,P., Kazmaier,M. and Granger-Schnarr,M. (1991) *Biochimie*, **73**, 423–431.
Schultz,S.C., Shields,G.C. and Steitz,T.C. (1991) *Science*, **253**, 1001–1007.
Shaka,A.J., Lee,C.J. and Pines,A. (1988) *J. Magn. Reson.*, **77**, 274–293.
Steitz,T.A. (1990) *Q. Rev. Biophys.*, **23**, 205–280.
Thliveris,A.T. and Mount,D.W. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 4500–4504.
Thliveris,A.T., Little,J.W. and Mount,D.W. (1991) *Biochimie*, **73**, 449–455.
Vuister,G.W., Boelens,R. and Kaptein,R. (1988) *J. Magn. Reson.*, **80**, 176–185.
Walker,G.C. (1984) *Microbiol. Rev.*, **48**, 60–93.
Weber,I.T. and Steitz,T.A. (1987) *J. Mol. Biol.*, **198**, 311–326.
Wertman,K.F. and Mount,D.W. (1985) *J. Bacteriol.*, **163**, 376–384.
Wilmot,C.M. and Thornton,J.M. (1990) *Protein Engng*, **3**, 479–493.
Wilson,K.P., Shewchuk,L.M., Brennan,R.G., Otsuka,A.J. and Matthews,B.M. (1993) *Proc. Natl Acad. Sci. USA*, **89**, 9257–9261.
Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
Wüthrich,K., Billeter,M. and Braun,W. (1983) *J. Mol. Biol.*, **169**, 949–961.

Received on May 25, 1994; revised on June 22, 1994