NMR studies on the Myb-DNA system

P. K. Radha, Anup Madan, L. C. Padhy and R. V. Hosur

Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

NMR structural studies on a bacterially produced DNA binding domain of *Drosophila* c-Myb protein as well as its cognate DNA sequence carried out in our laboratory have been surveyed. The structure of a self-complementary dodecamer DNA containing the Myb responsive element (TAACGG) has been determined to atomic resolution by the combined use of two-dimensional NMR, spectral simulations, restrained energy minimization and distance geometry calculations. The structure is seen to possess novel features which may play important roles during its interaction with the Myb protein. The DNA binding domain of c-Myb protein was seen to have a hydrophobic core and we have identified the types of residues contributing to its formation. Residues contributing to the hydrophobic core formation are seen to be well spread out over the whole length of the 160 residues in the protein and include isoleucines, valines, leucines, alanines threonines, aromatic residues, glutamines and possibly aspartates. Our experimental data in combination with those of others indicate that some of the amino acid residues which form the helical motifs that directly interact with DNA may also be a part of the hydrophobic core.

NUCLEAR magnetic resonance (NMR) spectroscopy has grown leaps and bounds during the last decade, establishing itself as the single most reliable technique for precise determination of three-dimensional (3D) structures of biological macromolecules in solution[1]. Starting from the discovery of Fourier transform NMR[2] in the 1960s, a series of developments such as multidimensional NMR[3], rotating frame experiments[4], inverse detection techniques[5] and gradient enhancement techniques[6] have contributed to the possibility of 3D structure determination of proteins and nucleic acids having molecular weights (MWs) as large as 20 kDa. Nearly a hundred protein structures and a similar number of nucleic acid structures in the MW range of 5–15 kDa have been determined to date with the above techniques. The developments in NMR have been facilitated by parallel developments in other areas, namely computer technology and genetic engineering, the former contributing to data-handling, spectrometer control and data-processing abilities and the latter contributing to the production of desired molecules in large quantities. The developments are still occurring at a rapid pace and the objective is to push the molecular size limits to even higher ranges of 30–40 kDa. These developments will take us in the realm of biology, providing insights into structure–function relationships in biomolecules.

The c-Myb protein is a proto-oncogene product which binds to DNA in a sequence-specific manner[7,8]. The protein functions as a transcriptional regulator in vertebrates and activates a large number of genes corresponding to different proteins[9–16]. The DNA binding activity of the protein is mediated by a region close to its amino terminus[17,18]. This domain contains three imperfect, tandem repeats of 51–53 amino acid residues and is represented in general as R1–R2–R3 (or R1R2R3). Each of the repeat units contains three tryptophan residues that are evenly spaced, 18–19 amino acids apart, and are evolutionarily conserved. Besides, the sequences corresponding to the DNA binding domains of the Myb proteins are highly conserved through evolution and have been detected in as widely divergent species as yeast, *Arabidopsis*, *Zea Mays*, *Drosophila*, chicken, mouse and man[19,20]. It is, therefore, envisioned that the tryptophan residues play an important role, directly or indirectly, in sequence-specific DNA binding. This notion is further supported by site-directed mutagenesis experiments, where replacement of tryptophans at many positions was found to affect drastically the DNA binding properties of the R1R2R3 protein[21].

The three-dimensional structure of the R1R2R3 protein is not yet known. To date, there have been three reports of nuclear magnetic resonance (NMR) studies; one on the solution structure of R3 repeat[22] and two on the R2R3 repeats[23,24] from two different species, mouse and chicken. The sequence homology in these proteins is more than 70%, being much better in the R3 domain. In the mouse protein it was observed that both R2 and R3 repeats contain three helices each and the typical helix–turn–helix motif was present in both the domains. In the case of chicken it was observed that the R3 domain contained three helices with a helix–turn–helix motif, while the R2 domain contained only two helices. The R1 domain exhibited largest variations in sequence homology and its exact role in the DNA binding process is not yet fully understood. It is, however, known that the presence of the R1 domain enhances the DNA binding affinity of the protein.

The consensus nucleotide sequence of Myb recognition has been identified to be YAACKG; more recently, this has been extended to 8 bases as YAACKGHH, where Y = C/T, K = G/T and H = A/C/T[25]. However, since the extension has substantial variability, it is unlikely to be

involved in specific contacts with the protein. There are conflicting reports in the literature as to whether the conformation of the protein changes on complex formation with DNA or not. From gel electrophoresis experiments, indications have come that the interaction induces bending in DNA[26]. Recent NMR results on the R23-DNA complex[24] have shown that the core AAC sequence makes the maximum specific contacts with the protein side-chains and the phosphate interactions provide a greater stability to the complex.

In the last few years we have studied the structure and interactions of the *Drosophila melanogaster* (*Dm*) c-Myb protein with its specific cognate DNA sequences. We have bacterially overexpressed the DNA binding domain of *Dm*-Myb, a polypeptide of 160 amino acid residues, and examined its structure and interactions by various spectroscopic techniques[27–31]. This article gives a survey of the significant achievements, with emphasis on the NMR-derived results. First we describe the solution structure of the Myb responsive element (MRE) determined to atomic resolution, and then describe the structural aspects of the hydrophobic core, which probably plays an important role in spatially orienting the helical segments directly involved in protein–DNA interactions.

Solution structure of the cognate DNA sequence

The DNA sequence considered here is d-ACCGTTAACGGT, the underlined stretch being the Myb recognition sequence. The molecule is self-complementary and, hence, exists as a duplex in solution under the usual conditions of room temperature, neutral pH and low salt, 0.1 M NaCl. The structure has been determined by combined use of 2D NMR, restrained energy minimization and distance geometry calculations, whose details have appeared elsewhere[32]. Briefly, the following steps were involved. First, sequential resonance assignments were obtained by standard procedures based on 2D NMR[1]. Then 2D J-correlated spectra were quantitatively interpreted for extracting coupling constant information and NOE correlated spectra were semiquantitatively interpreted to obtain rough estimates of the interproton distances. In the next stage, quantitative interpretation of the NOESY cross-peaks intensities in terms of 3D structure was carried out by explicit relaxation matrix calculations and iterative comparisons of the calculated and the experimental intensities. This step is described in some detail in the following paragraphs.

Experimental peaks in a 300 ms NOESY spectrum were integrated above suitable thresholds by summing up all the points spanning the peak. Calculated NOE intensities for any starting structure were obtained by relaxation matrix calculation[33,34] followed by a scaling procedure inherent in the SIMNOE algorithm[35]. All the integrated intensities thus obtained were referenced to a particular peak in both experimental and calculated spectra for normalization and comparison. In the absence of a good fit, the structure was refined and the procedure repeated until a satisfactory fit resulted. The iterative structure refinement proceeded as follows. A set of distance constraints with upper and lower bounds was derived by looking at the NOE fits at any particular step. Then restrained energy minimization with the NOE potential defined in terms of distance violations was carried out using the software package X-PLOR[36], for about 20–200 cycles. NOE fits were checked and the distance constraint set redefined to drive the structure towards a better fit. The restrained energy minimization included also the 120 dihedral angle constraints to restrict the sugar geometries in both the strands to the

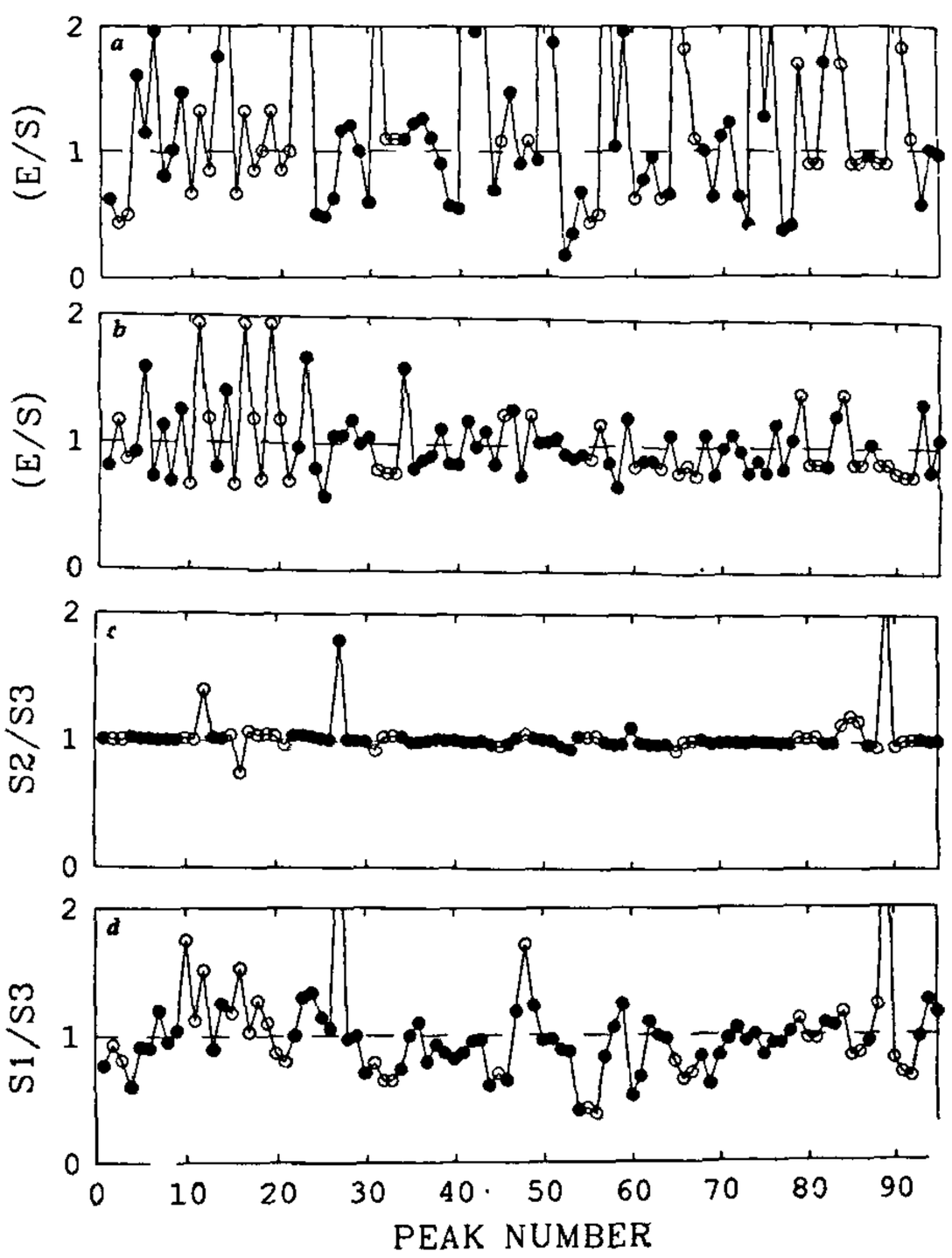


Figure 1. *a*, Comparison of experimental and calculated NOE peak intensities for the initial (B-DNA) model. Open circles represent overlapping peaks and filled circles nonoverlapping peaks. The peaks have been arranged sequencewise from left to right *b*, Peak-to-peak fit of experimental and calculated NOE peak intensities for the final structure. *c*, Peak-to-peak fit for the two structures S2 and S3 differing by only 0.05 Å on the all-atom r.m.s.d. scale *d*, Comparison of simulated NOE intensities for the final structure S1 and another distance geometry generated structure S2, which differs from S1 to an extent of 0.68 Å on the all-atom r.m.s.d. scale.

ranges derived from the coupling-constant information. The whole procedure was repeated with different starting models, and we noted that an energy-minimized B-DNA as the starting structure provided the best NOE fits at the end.

Figures 1 *a* and *b* show the initial and final fits for 95 NOE peaks for each strand of the 12-mer duplex. The NOE set consisted of 61 intranucleotide and 34 internucleotide peaks. To appreciate the adequacy of the NOE set and the quality of the fit in terms of closeness of the structure to the correct structure, we have carried out the following exercise. We selected two intermediate structures labelled S2 and S3, which had an all-atom r.m.s.d. (root mean square deviation) of 0.69 and 0.68 Å, respectively, from the final structure (labelled S1). We then compared the calculated NOEs of all the three structures and these are depicted in Figures 1 *c*, *d* in the same manner as the fits in Figure 1 *b*. These results provide a correlation between the similarity of structures and the similarity of NOE fits for the 95 peaks used. The two structures S2 and S3 which have an r.m.s.d. of 0.05 Å and thus are clearly identical have almost identical NOEs (Figure 1 *c*); the differences seen reflect the small differences in structures. Further, the structures S1 and S2, which are also close to the extent of 0.69 Å on the r.m.s.d. scale, have NOEs differing by 25–50% for a large number of peaks; a few peaks differ to a wider extent. This indicates that even as small a structural difference as 0.69 Å on the r.m.s.d. scale results in

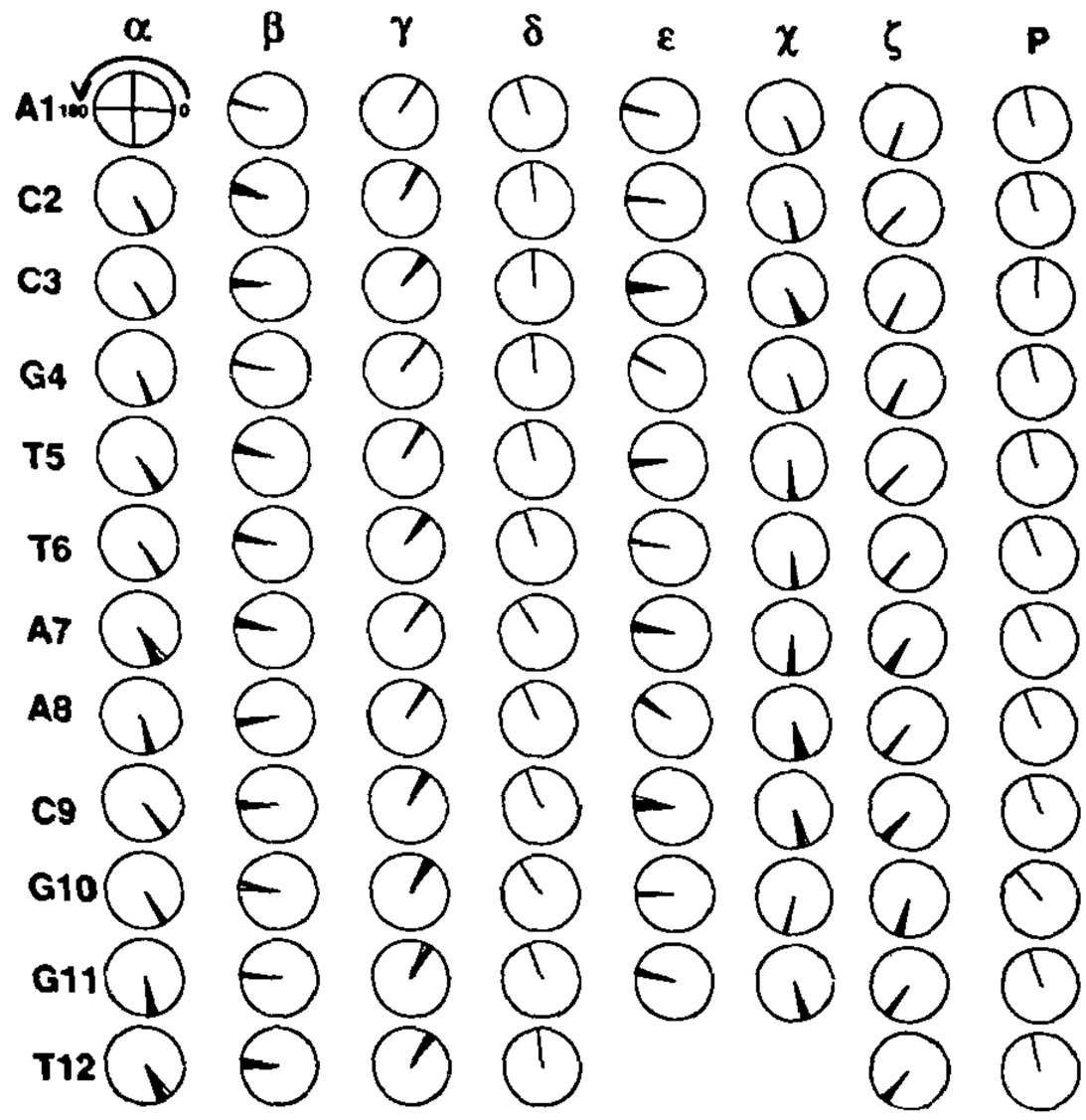


Figure 2. Analysis of the backbone torsion angles of 25 DG-generated structures. The individual torsion angles for all the nucleotides belonging to different structures are indicated by the value in degree in the corresponding dial. The direction of rotation is indicated in the first dial. The torsion angle is indicated in the top row and the nucleotide number in the leftmost column. As can be noted from the analysis, the torsion angles for all the residues are similar within ~ 5–10° for all the structures, indicating a single dominant structure for the DNA segment

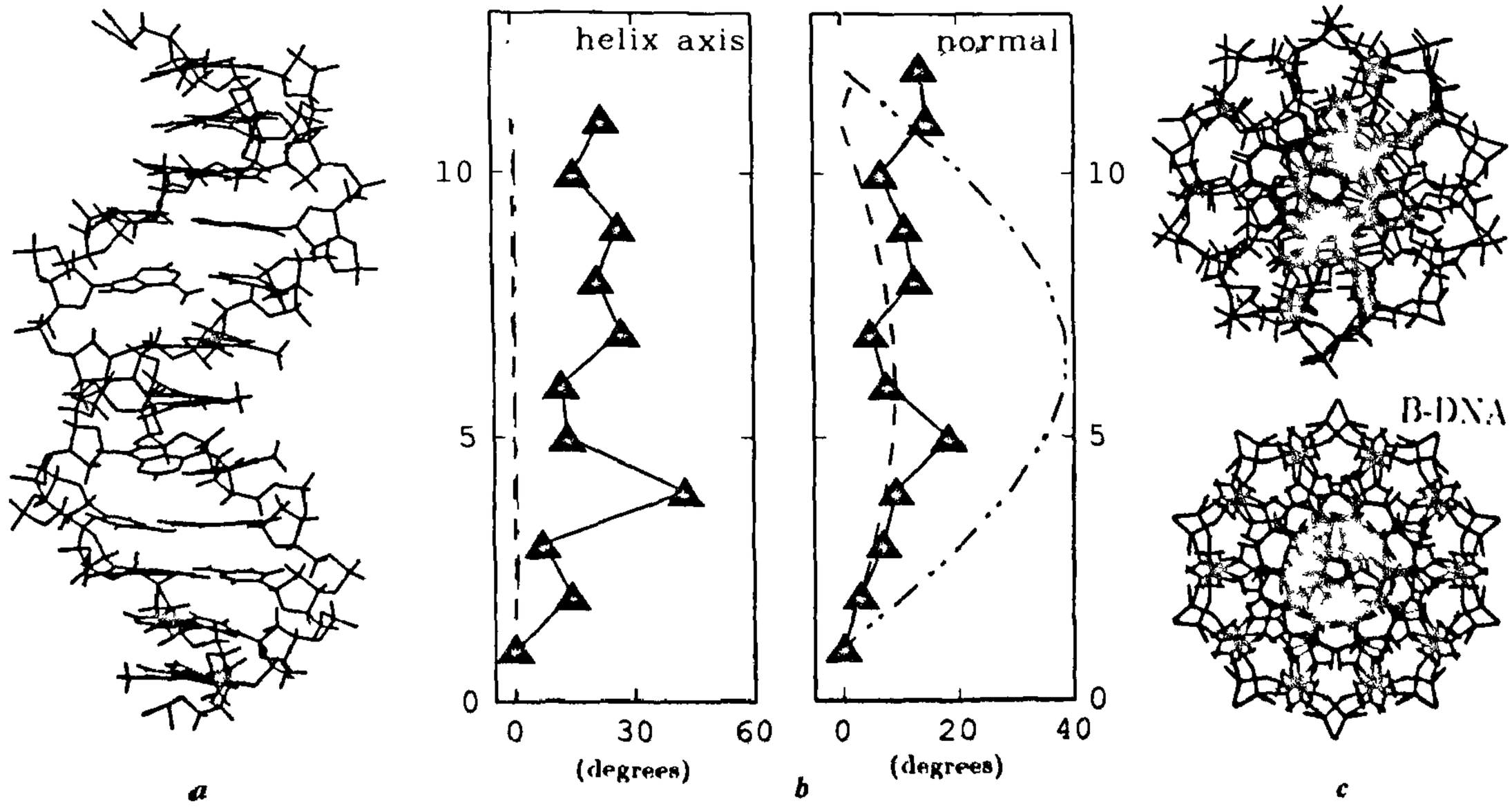


Figure 3. *a*, Stereo view of the final energy-minimized structure *b*, The course of the local helix axis and the base pair normals with respect to the first base pair as the reference The standard values for A-DNA (– .) and B DNA (. .) models have also been indicated *c*, Comparison of the orthogonal views of the final structure and the B-DNA structure.

a

b

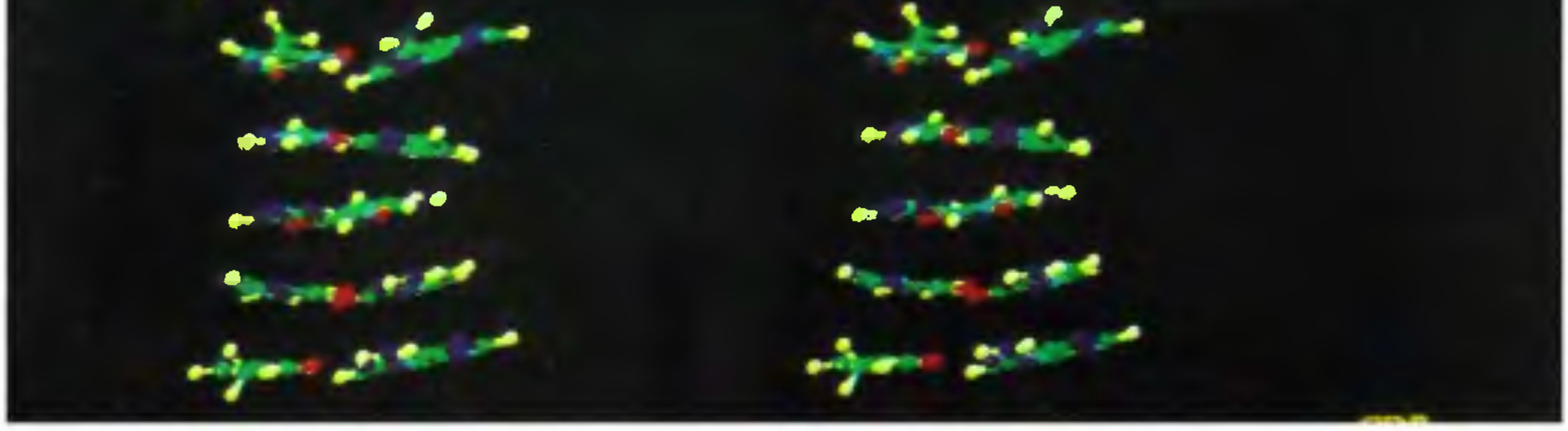


Figure 4. *a*, Stereo view of base stacking for the final energy-minimized structures; it can be noted that some of the base pairs are twisted more compared to others and the overall stacking pattern is symmetrical. *b*, Three-centred H-bond formed at the A7–A8 step. The three-centred bond occurs between the H (white, arrowed) of A7 (N6; blue, left top) and the bonded O4 of T20 (red, right upper) and T19 (red, right lower). Similar three-centred bonds were observed for the A1–C2 and the C2–C3 steps.

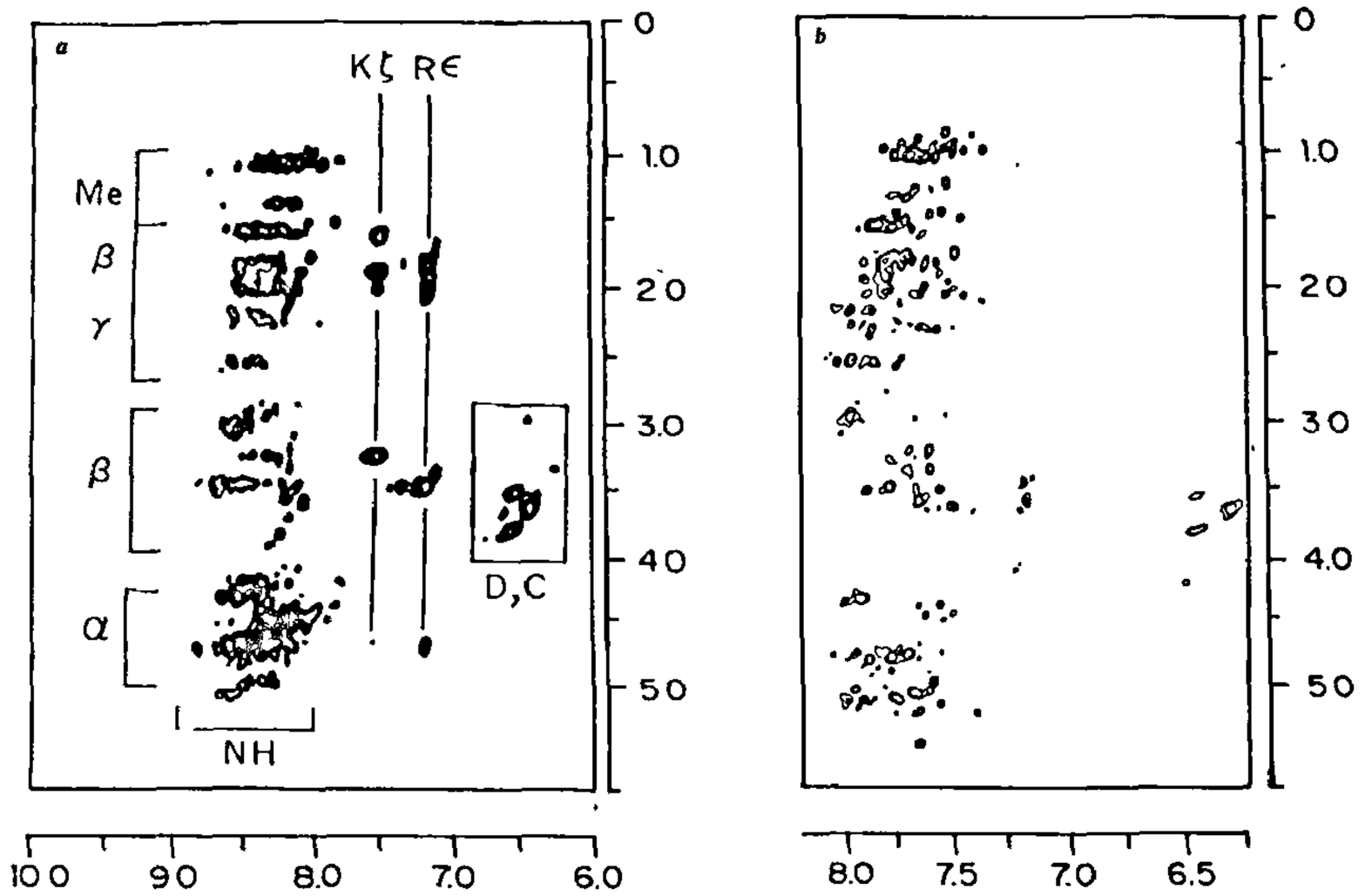


Figure 5. Portions of the clean TOCSY spectra (mixing time 80 ms) of 5 mM R1R2R3 protein in H_2O, pH 4.2 at temperatures 16°C (*a*) and 80°C (*b*). All the cross-peaks originate from the NH protons and gross identifications are given in (*a*). One-letter symbols identify the amino acid types and the Greek symbols α, β, γ, etc., identify particular proton types in the amino acids. It can be noted that many residues are deeply buried and their NH protons do not exchange at higher temperature

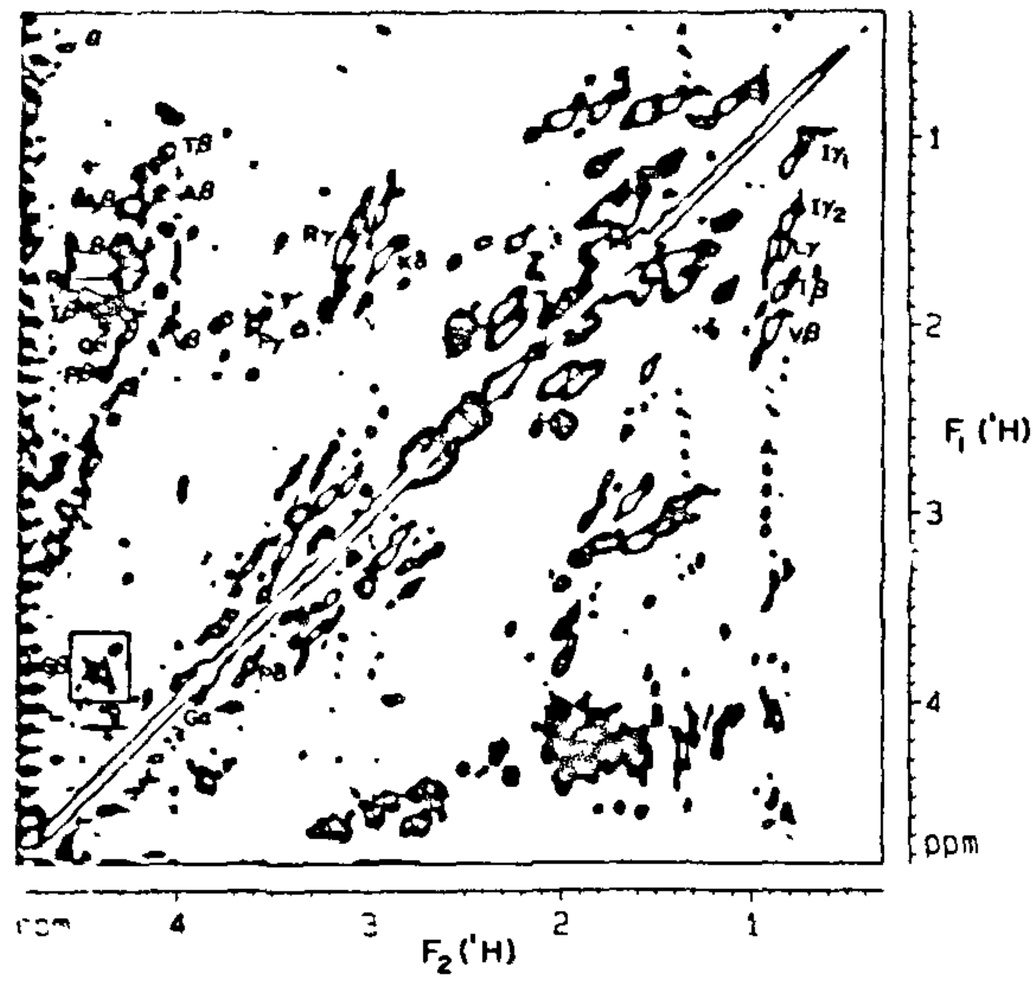
substantial changes in the NOEs for the two structures. This is a consequence of the fact that cross-relaxation rates contributing to NOE intensity have inverse sixth-power dependence on internuclear distances and a 20% deviation in the intensity represents approximately 3% deviation in the distances. Thus, the experimental to calculated NOE fit in Figure 1 *b* where the deviations are less than 25% for most peaks, represents a satisfactory fit and the structure S1 corresponding to the above fit represents a very dominant structure of the oligonucleotide in aqueous solutions.

To check for the uniqueness of the structure derived, extensive distance geometry calculations were carried out in the torsion angle space[37] with several different initial structures using the final distance constraint set obtained in the iterative procedure described above. NOE fits were calculated for each of the convergent structures to screen for the structures with similar NOE fits. Figure 2 presents an 'analysis of the torsion angles in the various structures so generated. It is clear that all the convergent structures are very similar.

Figure 3 shows the final energy-minimized structure (*a*), its helical characteristics (*b*) and an orthogonal view along with a similar view of the standard B-DNA (*c*). The structure shows significant departure from both B- and A-DNA families and there are substantial variations in each of the parameters along the sequence of the molecule. In particular, the local helix axis takes a zig-zag path, resulting in a variation in groove widths along the length of the molecule. The helical axis rise is irregular and varies from 2.79 to 3.73 Å and the base pairs are seen to exhibit high negative propeller twists at the recognition site. Three-centre H-bonds are seen in the major groove at AC, CC and AA steps. Figure 4 *a* shows the stacking of the bases in stereo view and Figure 4 *b* shows a particular three-centre H-bond as an illustration. Finally, the duplex exhibits a widened major groove at the AAC core. These features are of potential significance for the recognition of the cognate site by the Myb protein.

Hydrophobic core in the R1R2R3 protein

Several experiments employing protein fluorescence, fluorescence quenching by a neutral quencher acrylamide have suggested that the specific DNA interacts with R1R2R3 in 1 : 1 stoichiometry and binding occurs at a single site located in the hydrophobic pocket in the protein[29]. The binding site has been found to involve four to five conserved tryptophan residues, three of which are buried in the interior and not available to the solvent water molecules. The site involves also at least one tyrosine and one cysteine residue of the protein. Further, in protein denaturation studies by urea, simultaneous monitoring of the tryptophan unfolding to the



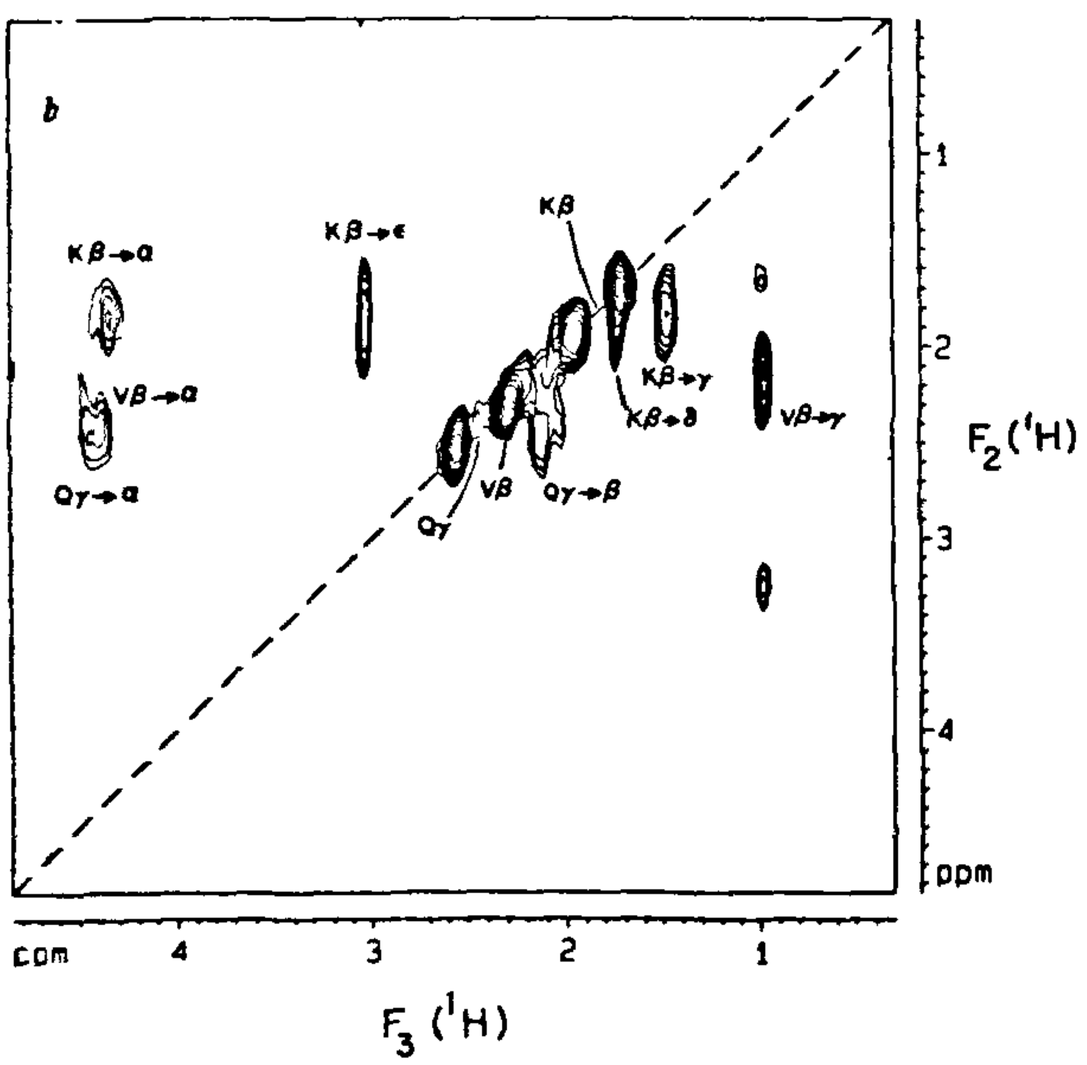


Figure 6. Experimental illustration of the spectral simplifications achieved in the 3D spectrum as compared to the 2D spectrum. *a* Portion of 2D ^1H–^1H clean TOCSY spectrum. *b*, A ^{13}C plane through the full 3D spectrum, at a ^{13}C shift of 33.4 ppm. A number of correlations which are hard to find out clearly in the top spectrum are readily identified in the lower spectrum and these have been indicated. The spin systems of K, V and Q residues can be clearly seen.

solvent, and loss of secondary structure elements indicated that the hydrophobic core may be the most stable element since it is the last one to get disrupted by urea. The fluorescence results also indicated that all the three repeats, R1, R2 and R3, in the protein contribute to the formation of the hydrophobic core. Because of these significant observations, we have tried to characterize the nature of the hydrophobic core in terms of identifying the participating amino acids by NMR spectroscopy.

Specific monitoring and distinction of the interior residues from the exterior ones can be made in NMR by making use of the fact that the amide and other side-chain NH_2 groups lying on the surface of the protein exchange protons more rapidly with solvent compared to those in the interior of the protein. Deuterium exchange studies provide the most direct method to identify the slowly exchanging amides (exchange rates $< 10^3\ s^{-1}$) in the protein. Amides exchanging faster than this rate will not be visible in the spectra in ^2H_2O solutions. In the case of R1R2R3 protein, we observed that only four or five out of the 162 NH–$C^{\alpha}H$ cross-peaks could be counted in the clean TOCSY spectrum in ^2H_2O solution at 16°C and these were present even after two months of keeping the protein in ^2H_2O. This implied that a small fraction of the protein is deeply buried in the hydrophobic environment and the rest of the protein is easily accessible to the solvent. We resolved to attempt a finer gradation of the exchange rates of amides by performing temperature variation experiments in H_2O solutions, that is, if the amide proton resonances are monitored as a function of temperature, the fast exchanging amides disappear from the spectra faster and this provides a useful filter to study the slowly exchanging amide protons. We may mention here that the ^1H spectra in ^2H_2O solutions did not show perceptible major changes as the temperature was raised slowly from 16 to 80°C. This indicated rather qualitatively that the structural core of the protein did not undergo major changes, and also, more relevantly in the present context, that the chemical shift assignments we would obtain at a higher temperature can be extrapolated to lower temperatures.

Figure 5 shows the same regions of clean TOCSY spectra of R1R2R3 protein at 16 and 80°C recorded with the same mixing time of 80 ms. Several points are noteworthy from these two spectra: (i) All the side-chain NH_2's of lysines and arginines, which can be easily identified from their characteristic relay peaks to other nonexchangeable protons along their chains, have exchanged out with water at 80°C. This implies that these two residue types lie entirely on the surface of the protein. (ii) Quite substantial simplification of the spectrum has occurred at 80°C and it is possible to count peaks in different regions; nearly 75 NH–$C^{\alpha}H$ peaks could be counted in the fingerprint region of the spectrum at 80°C. Here it may be noted that several $C^{\alpha}H$ protons occurring at 4.20–4.25 ppm are killed by water presaturation at 80°C. (iii) Several NH–$C^{\alpha}H$–$C^{\beta}H$ relay peaks arising from aromatic residues and possibly aspartates were detectable. This identification is provisional and is based on the typical chemical shifts of $C^{\beta}H$ protons. Actually, $C^{\beta}H$ protons of aspargines,

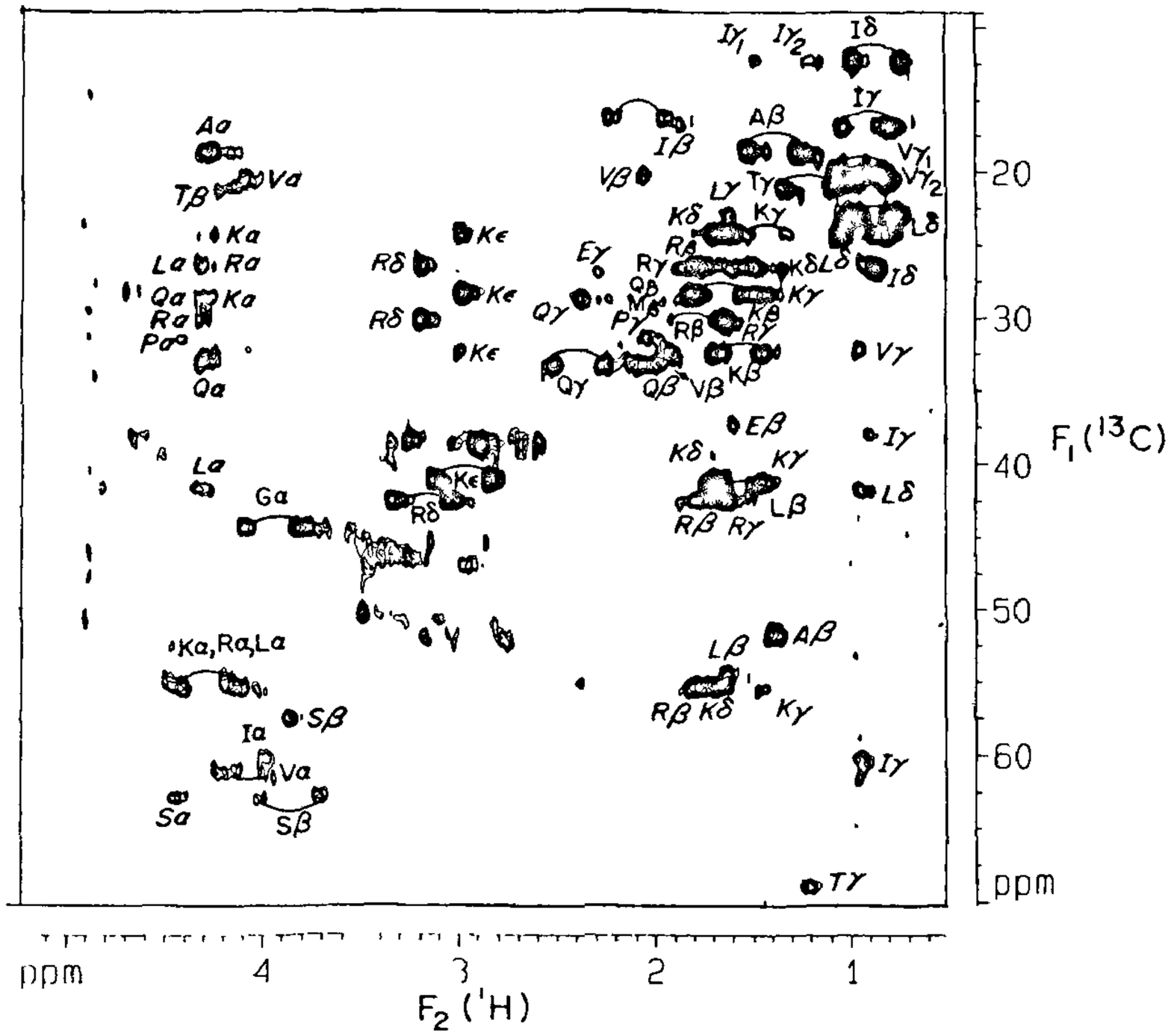


Figure 7. ^{1}H–^{13}C HMQC-TOCSY spectrum at natural abundance of the R1R2R3 protein recorded on BRUKER AMX 500 NMR spectrometer. Sample conditions. 5 mM protein, pH 5 6, temperature 16°C. 2048 t_2 and 512 t_1 points were used with carbon offset placed in the middle of the C^{β} region. The carbon spectral width was 90 ppm Quadrature direction along F_1 was achieved by the standard TPPI procedure. The MLEV mixing time was 51 ms. No ^{13}C decoupling was employed during detection and thus the direct peaks in the spectrum retain the one-bond $^{1}J_{CH}$ coupling (they have been joined). The relay peaks are, however, single peaks. The various peak assignments indicated have been obtained by analysis of different 2D spectra and the 3D spectrum

glutamines and cysteines are also expected in this region, but the presence of aspargines and glutamines can be easily checked from the double quantum (DQ) spectra. Indeed, the DQ spectra in H_2O showed that aspargines were absent but some glutamines were present whose $C^{\beta}H$ positions have consequently been identified. The presence or absence of cysteine was not possible to ascertain from these experiments. The number of remaining peaks suggested that both aromatic and aspartate residues were retained. (iv) There are clearly several methyl-bearing residues such as isoleucines, leucines, alanines, threonines and valines, and the NH protons of these show a reasonable spread.

Additional residue type identification of peaks at 80°C was obtained from the heteronuclear ^{1}H-^{13}C correlated 2D and 3D spectra. The spectra shown in Figure 6 illustrate how 3D spectra can be used for spin system identifications. The spectrum in (*a*) is the usual clean TOCSY spectrum in $^{2}H_2O$, which contains all the observable ^{1}H-^{1}H correlations and relay peaks for the protein. Clearly, the spectrum is highly crowded and difficult to analyse for unambiguous connectivities. The spectrum in (*b*) is a slice through the three-dimensional ^{13}C separated ^{1}H–^{1}H clean TOCSY spectrum taken at a particular carbon chemical shift. The spectrum is now simpler and the peaks can be readily identified as originating from particular spin systems. Such an analysis of all possible ^{13}C planes allows the identification of all the amino acid residues in the protein. Figure 7 summarizes all the identified spin systems in a 2D HMQC-TOCSY spectrum, which is basically a ^{1}H–^{13}C correlation spectrum including the ^{1}H–^{1}H relays at specific ^{13}C chemical shifts. The peaks in this spectrum provide a cross-check for the validity of the assignments made from 3D spectral analysis. An important observation made from this spectrum is that the methyl-bearing residues which are difficult to distinguish in ^{1}H-^{1}H correlated spectra are easily distinguished on the basis of the ^{13}C chemical shifts.

The above assignments provide a basis for the identification of the peaks in Figure 5. Figure 8 shows the specific side-chain assignments to residue types obser-

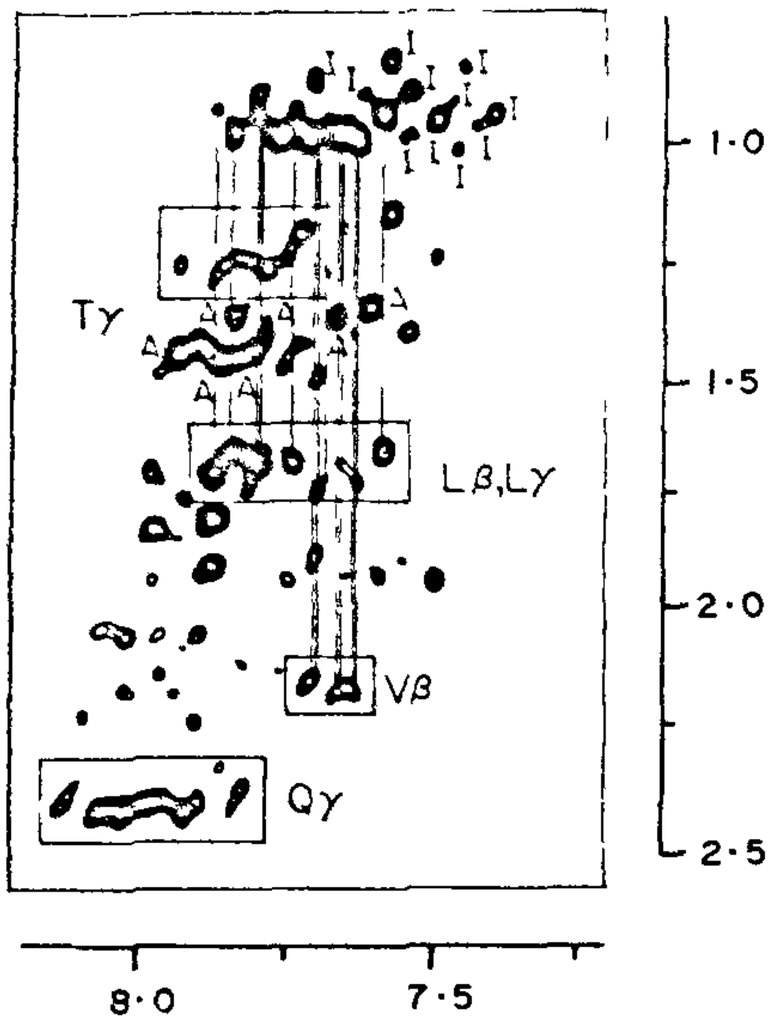


Figure 8 Specific side-chain assignments to residue types observed in the clean TOCSY spectrum (80 ms spin lock time) of R1R2R3 protein in H_2O recorded at 80°C. Cross-peaks belonging to valines and leucines are indicated by vertical lines. Peaks belonging to glutamine and threonine γ have been boxed, and those belonging to alanines have been labelled A. The 11 isoleucine γ's have been labelled as I.

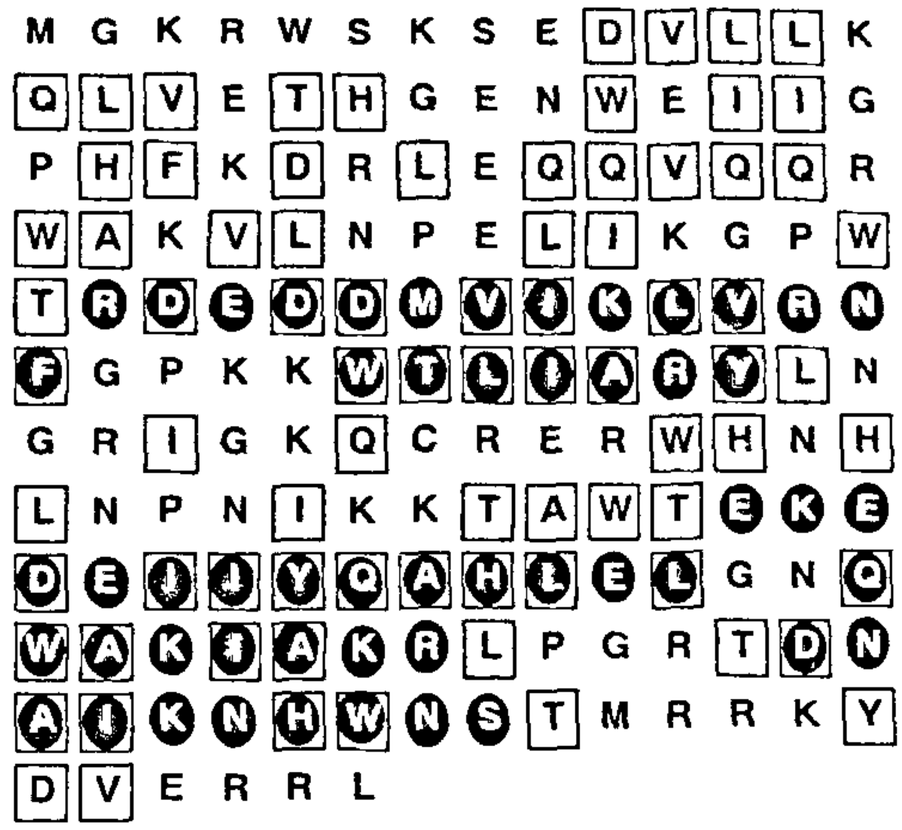


Figure 9. Amino acid sequence of R1R2R3 protein indicating the residues (boxed) for which the amides are observed at 80°C. All the residues (excepting the terminal ones) of any particular type that has been seen in the high-temperature spectrum have been marked, although we were able to count the numbers only in few cases. For example, separate identification of all peaks belonging to individual aromatic residues was not done. But the total count in the figure tallies closely with the total number of residues seen in the spectrum. We notice that the slowly exchanging amides are spread out all along the sequence, implying participation of all the portions of the protein in the formation of the hydrophobic domain. The circled residues are expected to be forming helices, from sequence comparison with R2–R3 domains of mouse and chicken c-Myb, whose NMR secondary structures have been reported recently[23, 24].

vable at 80°C that probably constitute the hydrophobic core in the R1R2R3 protein. We noticed here that backbone amide protons had reasonable spread and we could count 5–6 valines, 7 alanines, 6 threonines, 5–6 leucines, 11 isoleucines and 7–8 glutamines. There were several other peaks which originated, very likely, from aspartates and aromatic residues; the latter have hydrophobic character and hence are expected to remain shielded from the water.

Figure 9 shows the sequence-wise location of the various residues identified in Figure 8. We notice that the residues are distributed along the whole length of the protein. This indicates that all the three repeats contribute to the formation of the hydrophobic core in the protein and about 70–80 residues contribute to its formation. It is also worthwhile noting that glutamines and aspartates, which are hydrophilic residues, are also included in the above list. A possible explanation could be that they are involved in the formation of hydrogen bonds through their polar side-chains with other residues in the protein. Examination of the relative positions of the hydrophobic residues in the sequence, *vis-à-vis* the known secondary structure elements of vertebrate R1R2R3 from NMR studies in the conserved R3 and R2–R3 domains and their DNA complexes[22, 24], reveals that some of the hydrophobic core residues may form parts of helices involved in specific interactions with the DNA. The helices may be lost at elevated temperatures but the hydrophobic core appears to be the last one to get disrupted[31].

Conclusions and future directions

We have described in this article structural information derived from NMR studies in our laboratory, on a specific target DNA sequence of the DNA binding domain of *Drosophila melanogaster* c-Myb protein and also on the hydrophobic core present in the protein where the DNA binds. The structure determination has unravelled important features which could constitute hot points of recognition. Extensive use of homonuclear and heteronuclear two- and three-dimensional NMR experiments has enabled gross identification of residue types involved in the formation of the hydrophobic core in the protein, which has an important role to play in DNA recognition. The hydrophobic core seems to be constituted by all the three repeats in the protein. Sequence comparison with R2–R3 domain of mouse and chicken c-Myb, for which NMR structures were reported recently, provided clues to identification of the secondary structure and of the so-called helix–turn–helix motif of DNA recognition. This provides a supplementary character to our previous work on R1R2R3–DNA interaction by fluorescence quenching studies, in that the cognate

DNA binding site of this protein is associated with a hydrophobic core element being surrounded by 4–5 tryptophans, at least one tyrosine and a cysteine[27–29].

The work reported in this paper represents an important step forward in our efforts to understand DNA–Myb interaction and recognition at the atomic level. A lot more remains to be done from the viewpoint of 3D structure determination of the protein and of the complex. Sequence-specific assignment of resonances is the most crucial step requiring the employment of a battery of multidimensional NMR experiments on isotopically (^{15}N and ^{13}C) labelled protein. Indeed, we have just successfully achieved the production of large amounts of labelled R1R2R3 protein.

Atomic-level structure determinations of the R1R2R3 and also of their segments R1R2, R2R3 and R3, which are also in the offing in our laboratory are expected to throw light on several important questions. For example: (1) What is the role of R1 repeat in specific recognition? (2) Do the individual repeats retain their structural motifs when any of the other repeats is knocked off from the protein? (3) Which are the residues making contact with the DNA and how is the specificity of interaction brought about? (4) Is it possible to alter the DNA binding specificity by structural-knowledge-based mutations in the protein?

1 Wuthrich, K., *NMR of proteins and Nucleic Acids*, Wiley, New York, 1986.
2. Ernst, R R and Anderson, W. A., *Rev. Sci. Instrum.*, 1966, **37**, 93–102.
3. Ernst, R. R, Bodenhausen, G., and Wokaun, A, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1987.
4 Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D. and Jeantoz, R W., *J Am Chem. Soc.*, 1984, **106**, 811–813.
5 Muller, L, *J. Am. Chem Soc.*, 1979, **101**, 4481–4484.
6. Davis, A. L., Keeler, J., Laue, E. D. and Moskau, D., *J. Magn Reson*, 1992, **98**, 207–216.
7. Biekenkapp, H., Borgmeyer, U, Sippel, A. E. and Kempnauer, K. H, *Nature*, 1988, **335**, 835–837.
8. Nakagoshi, H., Nagase, T., Kanei-Ishii, C., Ueno, Y., and Ishii, S., *J. Bio Chem*, 1990, **265**, 3479–3483.
9 Ness, S. A., Maarknell, A and Graf, T., *Cell*, 1989, **59**, 1115–1125.
10 Dasgupta, P., Saikumar, P, Reddy, C. D. and Reddy, E. P., *Proc. Natl Acad. Sci USA*, 1990, **87**, 8090–8094.
11. Dasgupta P., Reddy, C. D, Saikumar, P. and Reddy, E P, 1980, *J. Virol.*, **66**, 270–276.
12. Evans, J. L., Moore, T L., Kuehl, W. M, Bender, T and Ling, J., *Mol. Cell Biol.*, 1980, **10**, 5747–5752
13. Nicolaides, N C., Gualdi, R., Casadevall, C., Manzella, L. and Calabretta, B., *Mol. Cell. Biol.*, **11**, 6166–6176
14 Zobel, A., Kalkbrenner, F., Guemann, S, Nawarath, M., Vorbruggen, G. and Moelling, K., *Oncogene*, 1991, **6**, 1397–1407.
15 Siu, G., Wurster, A L., Lipsick, J. S. and Hedrick S M, *Mol. Cell. Biol.*, 1992, **12**, 1592–1604.
16 Sureau, A, Soret, J., Vellard, M, Crochet, J. and Perbal, B, *Proc. Natl Acad. Sci. USA*, 1992, **89**, 11683–11687.
17. Katzen, A. L., Kornberg, T. B. and Bishop, J. M. *Cell*, **41**, 449–456.
18 Peters, C. W. B, Sippel, A E, Vingron, M and Klempnauer, K-H, *EMBO J.*, 1987, **6**, 3085–3090.
19 Luscher, B and Eiseman, R, *Genes Develop.*, 1990, **4**, 2235–2240
20. Shen-Ong, G. L. C., *Biochem Biophys Acta*, 1990, **1032**, 39–52.
21. Saikumar, P., Murali, R. and Reddy, E. P., *Proc. Natl. Acad Sci. USA*, 1990, **87**, 8452–8456
22. Ogata, K., Hojo, H., Aimoto, S, Nakai, T., Nakamura, A. S, Ishii, S and Nishimura, Y., *Proc. Natl. Acad Sci. USA*, 1992, **89**, 6428–6432.
23. Jamin, N., Gabnelsen, O. S, Gilles, N., Lirsac, P.-N. and Toma, F, *Eur. J. Biochem*, 1993, **216**, 147–154.
24 Ogata, K., Morikawa, S, Nakamura, H, Sakikawa, A, Inoue, T., Kanai, H, Sarai, A, Ishii, S and Nishimura, Y., *Cell*, 1994, **79**, 639–659.
25. Weston, K. *Nucleic Acids Res.*, 1992, **20**, 3043–3049
26. Saikumar, P, Gabreil, J L. and Reddy, P., *Oncogene*, 1994, **9**, 1279–1287.
27. Madan, A, Hosur, R V and Padhy, L. C., *Biochemistry*, 1994, **33**, 7120–7126.
28. Radha, P. K., Madan, A, Padhy, L. C. and Hosur, R V., *Curr. Sci*, 1994, **66**, 287–294
29. Madan, A., Radha, P. K., Hosur, R. V. and Padhy, L. C., *Eur. J. Biochem.*, 1995, in press
30 Radha, P. K., Madan, A, Padhy, L. C. and Hosur, R. V., *Proc Indian Acad Sci. Chem. Sci.*, 1994, **106**, 7, 1537–1549.
31. Madan, A., Radha, P. K, Srivastava, A., Hosur, R. V and Padhy, L. C., *Eur J. Biochem.*, 1995, in press.
32. Radha, P K., Madan, A, Nibedita, R and Hosur, R. V., *Biochemistry*, 1995, **34**, 5193–5922.
33. Macura, S. and Ernst, R. R, *Mol Phys.*, 1980, **41**, 95–117.
34. Keepers, J W., and James, T., *J. Mag Reson*, 1984, **57**, 404–426
35. Nibedita, R, Ajay Kumar, R, Majumdar, A. and Hosur, R V, *J. Biomol. NMR*, 1992, **2**, 467–476
36 Brunger, A. T., X-PLOR (Version 2 1) Manual, Fellows of Harvard University, New York, 1990.
37 Radha, P. K., Nibedita, R, Ajay Kumar, R. and Hosur, R. V., *Methods Enzymol*, **261**, in press.

ACKNOWLEDGEMENTS. The facilities provided by the National Facility for High-Field NMR supported by the Department of Science and Technology are gratefully acknowledged. We thank Bhabha Atomic Research Centre for the use of X-PLOR package and Dr Manju Bansal for the NUPARM package for DNA structure analysis.