ELSEVIER

Biophysical Chemistry 68 (1997) 147–159

Biophysical Chemistry

Biophysical investigations on the *myb*–DNA system

R.V. Hosur *, P.K. Radha, Anup Madan, L.C. Padhy [1]

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

Received 25 November 1996; accepted 13 January 1997

Abstract

The oncogene product c-*myb* is a transcriptional modulator and is known to play important roles in cell growth and differentiation. It binds to DNA in a sequence specific manner and its cognate sequence motifs have been detected in the genes of proteins implying its role in a variety of regulatory functions. The protein has a DNA binding domain consisting of three imperfect repeats with highly conserved tryptophans at regular spacings in each of the repeats. We have carried out a variety of investigations on the structure and interactions of the DNA binding domain of *Drosophila* c-*myb* and its cognate DNA target sequences. The domain has been bacterially over-expressed by subcloning a segment of the gene coding for the domain in a pET 11d vector and transforming it into *E. coli* BL21 (DE3). Circular dichroism of the protein has revealed that the domain is largely helical in nature. Fluorescence investigations indicated that three out of the nine tryptophans are solvent exposed and the others are buried in the interior. The recombinant protein is able to distinguish between specific and non-specific DNA targets in its binding and the interaction is largely electrostatic in nature in both cases. Dynamic fluorescence quenching experiments suggested that the DNA binding sites on the protein for specific and non-specific DNA targets are physically different. Most of the conserved tryptophans are associated with the specific DNA binding site. Simulated annealing and molecular dynamic simulations in a water matrix have been used to predict an energetically favoured conformation for the protein. Calculation of surface accessibilities of the individual residues shows that nearly 60% of the residues are less than 50% accessible to the solvent. Two and three dimensional NMR experiments with isotopically labelled protein have enabled spin system identification for many residue type and the types of residues involved in hydrophobic core formation in the protein. In an attempt to see the DNA surface possibly involved in specific interaction with the protein, a three-dimensional structure of a 12 mer cognate DNA has been determined by NMR in conjunction with restrained energy minimization. The recognition sequence shows interesting structural characteristics that may have important roles in specific interaction. © 1997 Elsevier Science B.V.

Keywords: *myb*–DNA system; R123 domain; Biophysical studies

1. Introduction

The oncogene product c-*myb* is a transcriptional modulator, which seems to play critical roles in cell differentiation and growth [1–6]. It binds to DNA in

* Corresponding author.
[1] Also corresponding author.

PII S0301-4622(97)00026-4

a sequence-specific manner [7–10] and its target sequences detected in many genes in a variety of species indicate that the protein regulates expression of many proteins in a wide range of species. The DNA binding domain of the protein is near to the N-terminus and apparently consists of three imperfect repeats of 51–52 residues. The repeats are generally labelled as R1, R2, R3 and each repeat contains three conserved tryptophans, 18–19 residues apart. There have been several mutational and deletional experiments and it is now known that R2–R3 segment is the minimal DNA binding domain required for specific recognition and R1 seems to play a role in increasing the affinity of the interaction. The conserved tryptophans are believed to form a hydrophobic core where the DNA binds. Mutation of the tryptophans in the R2–R3 segment affects DNA specificity. The protein contains also a conserved cysteine whose mutation affects the specific DNA binding activity. It is speculated that the redox reactions of the cysteine may contribute to the regulatory roles of the protein.

The sequence of the DNA binding domain is conserved to different extents in the three repeats and in different species (Fig. 1). The degree of conservation is highest in R3 (−75%), followed by R2 and in R1 the conservation is as low as 45%. Overall, the extent of conservation is highest among the invertebrates and decreases as one goes to higher organisms. For example the overall homology between R1–R2–R3 (or R123 in brief) of *Drosophila* and mouse is approximately 60%.

Because of these interesting observations and the biological importance of the protein, there have been

R1:

Mouse c-Myb	LGKTR	W TREEDEKLKKLVEQNGTDD	W	KVIANYLPNRTDVQCQHR	W QKVLNPE
Cicken Myb	LGKTR	W TREEDEKLKKLVEQNGTED	W	KVIASFLPNRTDVQCQHR	W QKVLNPE
Human B-Myb	KCKVK	W THEETEQLR ALVRQFGQQD	W	KFLASHFPNRTDQQCQYR	W LRVLNPD
Human A-Myb	WNRVK	W TRDETDKLKKLVEQHGTDD	W	TLIASHLQNRSDFQCQHR	W QKVLNPE
Drosophila Myb	GFGKR	W SKSEDVLLKQLVETHG EN	W	ELIGPHFKDRLEQQVQQR	W AKVLNPE
Yeast BAS1	LGYEN	G LEDVKTIQQSNHLSKCIA	W	DVLATRFKHTVRTSKDVRKR	W TGSLDPN

R2:

Mouse c-Myb	LIKGP	W TKEEDQRVIKLVQKYGPKR	W	SVIAKHLK GRIGKQCRER	W HNHLNPE
Cicken Myb	LIKGP	W TKEEDQRVIELVQKYGPKR	W	SVIAKHLK GRIGKQCRER	W HNHLNPE
Human A-Myb	LIKGP	W TKEEDQRVIELVQKYGPKR	W	SVIAKHLK GRIGKQCRER	W HNHLNPE
Human B-Myb	LIKGP	W TKEEDQKVIELVKKYGTKQ	W	TVIAKHLK GRLGRQCRER	W HNHLNPE
Drosophila Myb	LIKGP	W TRDEDDMIVIKLVRNFGPKK	W	TVIARYLN GRIGKQCRER	W HNHLNPN
Zea maysC1	VKRGA	W TSKEDDALAAYVKAHGEGK	W	REVPQKAGLRRCGKSCRLR	W LNYLRPN
Yeast BAS1	LKKGK	W TQEEDEQLLKAYEEGHPH_	W	LSISMDIP GRTEDQCAKR	Y IEVLGPGSK

R3:

Mouse c-Myb	VKKTS	W TEEEDRIIYQAHKRLGNR	W	AEIAKLLPGRTDNAIKNH	W NSTMRRKV
Cicken Myb	VKKTS	W TEEEDRIIYQAHKRLGNR	W	AEIAKLLPGRTDNAIKNH	W NSTMRRKV
Human A-Myb	VKKSS	W TEEEDRIIYEAHKRLGNR	W	AEIAKLLPGRTDNSIKNH	W NSTMRRKV
Human B-Myb	VKKSC	W TEEEDRIICEAHKVLGNR	W	AEIAKLLPGRTDNAVKNH	W NSTIKRKV
Drosophila Myb	IKKTA	W TEKEDEIIYQAHLELGNQ	W	AKIAKRLPGRTDNAIKNH	W NSTMRRKY
Zea maysC1	IRRGN	I SYDEEDLIIRLHRLLGNR	W	SLIAGRLPGRTDNELKNY	W NSTLGRRA
Yeast BAS1	GRLRE	W TLEEDLNLISKVKAYGTK	W	RKISSEMEFRPSLTCRNR	W RKITMVV

Fig. 1. Sequence comparison of the R123 protein from different species. The conserved residues are shown in bold.

many reports in the literature on the structural aspects of the individual domains, biophysical investigations on the stabilities of the proteins, and the DNA binding properties of the R123 domains from different species [11–20]. Ogata et al. [20] investigated the structures of the R1, R2, R3 segments separately and also the complex of R23 with a specific DNA segment. They observed that all the three repeat sub-domains in mouse are similar in structure containing three helices under conditions of neutral pH and room temperatures. Jamin et al. [15] however, observed that in R23 of chicken the R3 domain contains three helices, while the R2 repeat contains only two ordered helices and the third is

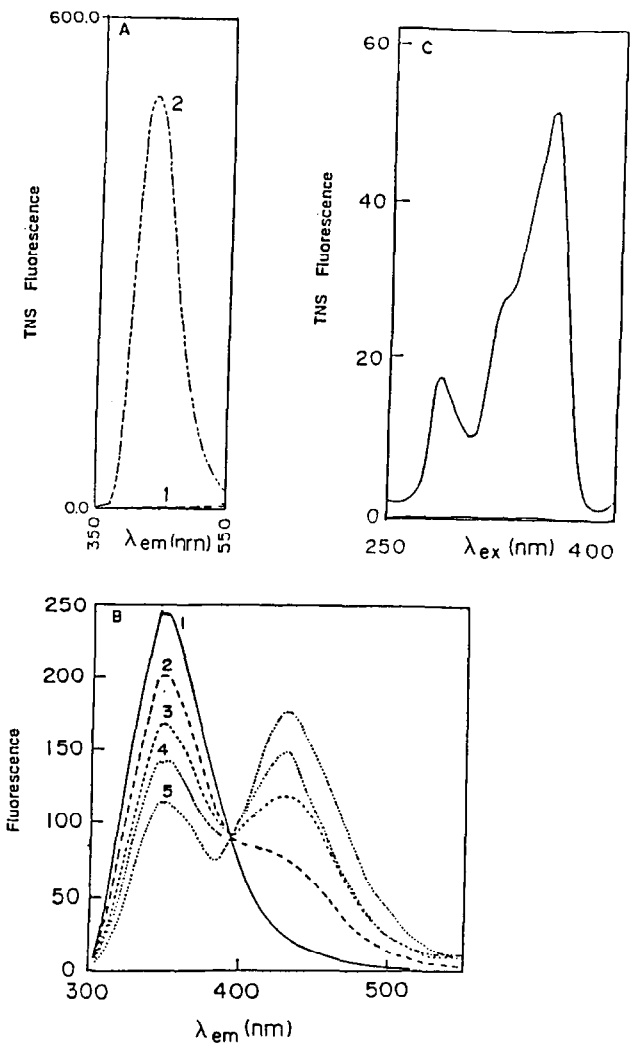


Fig. 2. (a) TNS fluorescence in the free form (curve 1), when bound to the R123 protein (curve 2). (b) Fluorescence spectra of the protein–TNS complex with increasing addition of TNS. The protein fluorescence (349 nm) decreases progressively and the TNS fluorescence (435 nm) increases concomitantly. Excitation is at 279 nm which corresponds to (Tyr + Trp) absorption in the protein. (c) Excitation spectrum of the protein–TNS complex while emission of TNS fluorescence is monitored. A band at 295 nm is observed which is characteristic of tryptophan absorption. This suggests energy transfer between TNS and tryptophans in the protein.

disordered. A similar feature has been noted by Carr et al. [19] in their work on Human B-*myb* R23.

Our work concerns with the c-*myb* related protein in *Drosophila*. In particular, we have studied biophysical characteristics of *Drosophila* R123 and have carried out experiments on the stability, folding and DNA binding properties of the protein. In addition we have determined the structure of a cognate DNA binding motif [21–26]. We present here a survey of the results on the structural aspects and R123–DNA interaction.

2. Expression of the R123 protein

The segment of the gene for the DNA binding domain, R123 of the *Drosophila* c-*myb* protein was generated by PCR (Polymerase Chain Reaction) and subcloned into the pET 11d vector for expression in *E. coli*. The protein was then induced with IPTG at a concentration of 25 uM. The recombinant protein was isolated from the cells and purified by a 4-step protocol involving, PEI precipitation, ammonium sulfate precipitation, DEAE-cellulose and SP-sephadex chromatography. The protein was refolded by dilution and dialysis. The extinction coefficient of the purified protein was determined to be 54300, as per the procedure given by Mach et al. [27]. For the sake of heteronuclear NMR experiments, ^{13}C and ^{15}N isotopically labelled protein was produced by growing the transformed cells in minimal media with the exclusive carbon source of labelled sodium acetate and nitrogen source of labelled ammonium chloride. The protein was isolated and purified in the same manner.

3. Hydrophobic core in the *Drosophila* R123 protein

The presence of the nine conserved tryptophans and their conserved spacing along the sequence has aroused a lot of curiosity about their possible role in the structure and function of the protein. Considering the hypothesis that they may be contributing to the formation of a hydrophobic core in the protein, we have investigated its possibility using TNS as a probe. This molecule has the property that, it has no fluorescence in the free state in aqueous solution, but, when it is bound to a hydrophobic site in a protein it fluoresces strongly. Fig. 2a shows the fluorescence spectra of TNS in free and bound forms. This clearly indicates that there is indeed a hydrophobic core or pocket in the R123 protein. In Fig. 2b the spectra of protein–TNS complex at increasing

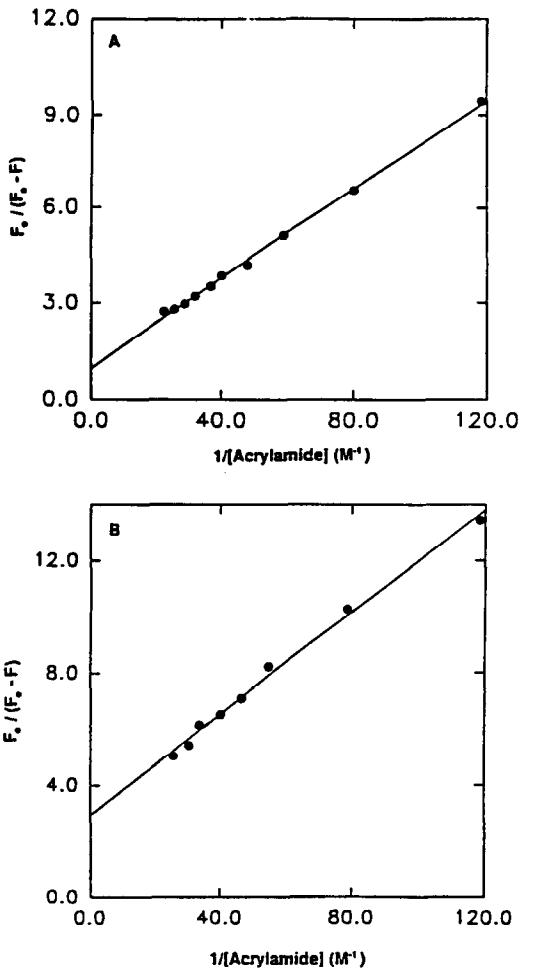


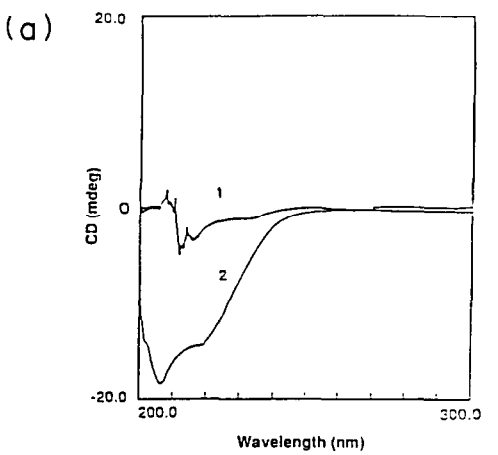
Fig. 3. Lehrer plots of dynamic quenching of protein fluorescence by neutral quencher acrylamide. plots (A) and (B) are for the native and denatured R123 (by 8 M urea) respectively. The intercepts on the *y*-axis allow calculation of the number of exposed tryptophans in the protein as per Eq. (1). It is observed that in denatured protein all the tryptophans are accessible, while in native protein only a third of the tryptophans are accessible to the solvent.

concentrations of TNS are shown. It is observed that the TNS fluorescence grows as the TNS concentration increases and there is a concomitant decrease in the intrinsic protein fluorescence. This seems to suggest an interaction between the tryptophans which are largely responsible for the protein fluorescence and TNS in the protein–TNS complex. Fig. 2c confirms this observation through the excitation spectrum of TNS fluorescence in the complex. There is a peak at 295 nm which is characteristic of tryptophan absorption and free TNS does not absorb at this wave length. This indicates that there is resonance energy transfer between tryptophan and TNS occurring in the protein–TNS complex. In other words the tryptophans are located near or in the hydrophobic core in the protein. Quantitative estimation of TNS binding indicated that there is one binding site and the binding constant is approximately $1 \times 10^7\ M^{-1}$.

To further investigate the extent of accessibility of the tryptophans to the solvent, we have used the technique of dynamic quenching of fluorescence by a collisional quencher, acrylamide. Collisions between this external molecule and the tryptophans in the protein quenches the fluorescence of the latter. Therefore, the extent of fluorescence quenching depends upon the concentration of the quencher and follows a relation [28]:

$$\frac{F_o}{F_o - F} = \frac{1}{k[Q]\tau f_a} + \frac{1}{f_a} \tag{1}$$

where, F_o is the protein fluorescence in the absence



(b)

R1:

MGKR W KS KEDVLLKQLVETHG_EN W EIIGPHFKDRLEQQVQQR– W AKVLNPE

R2:

LIKGP W TRDEDDMVIKLVRNFGPKK W TV IARYLNGRIGKQCRER W HNHLNPN

R3:

IKKTA W TEKEDEI IYQAHLELGNQ W AKIAKRLPGRTDNAIKNH W NSTMRRKYDVEERRI

Fig. 4. (a) Circular dichroism spectra of native and denatured proteins. The bands and 208 and 222 nm in the spectrum of the native protein are characteristic of a helical structure. (b) Sequence-wise prediction of the secondary structure elements. H indicates a helix, L indicates loop and indicates no definite secondary structure prediction.

of the quencher, F is the fluorescence in the presence of the quencher, $[Q]$ is the concentration of the quencher, f_a is the quenching fraction at infinite quencher concentration, k is the rate constant for the quenching and τ is the lifetime of the fluorophore (tryptophan) in the absence of the quencher. From the above equation it is clear that f_a represents the maximum quenchable fluorescence and thus indicates the number of tryptophans exposed to the solvent in the protein. A plot of $F_0/(F_0 - F)$ vs. $1/[Q]$ yields a straight line whose intercept on the y-axis yields the value of $1/f_a$. Thus, assuming that all the tryptophans have similar characteristics, one can determine the number of tryptophans exposed to the solvent in the protein. Fig. 3 shows these plots for the native R123 protein and for the denatured protein (8 M urea). It is seen that, while in the denatured protein essentially all the tryptophans are solvent exposed, in the native protein only a third of the tryptophans are exposed, i.e., two-thirds of the tryptophans are buried in the interior and contribute to the formation of the hydrophobic core.

4. Secondary structure characteristics of the R123 protein

The folded recombinant protein was investigated by circular dichroism spectroscopy to identify the secondary structure elements such as helices, sheets, etc. Fig. 4a shows the circular dichroism spectra of the protein in the native form and also in the dena-

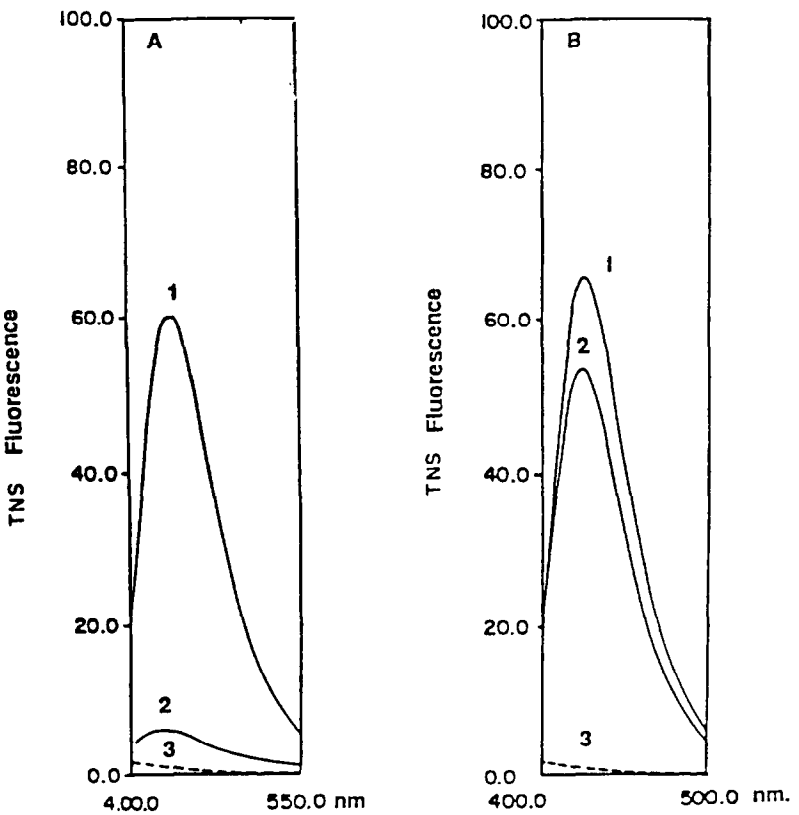


Fig. 5. TNS fluorescence in protein–TNS complex before (a) and after (b) addition of the same specific (A) and non specific (B) DNA targets as in Fig. 6. We notice that specific DNA displaces TNS from the complex but the non-specific DNA fails to do so. Excitation is at the absorption wavelength of TNS

tured form as a control. The spectrum shows two bands at 208 and 222 nm which is a characteristic feature of a helical segment. From the intensities of the bands we calculated the helical content in the protein to be approximately 58%. In order to get an insight into the possible locations of the helices in the protein structure, we used secondary structure prediction algorithms [29] and the results of these are schematically shown in Fig. 4b. Interestingly these results agree fairly well with the NMR derived results on the R23 domain in mouse and chicken [15,20]. The computed helical content above 80% confidence level in the calculation is approximately 35%.

5. Nature of the DNA binding site

The topography of the specific DNA binding site on the R123 protein has been investigated by fluorescence spectroscopy using a multiplicity of techniques. First of all we observed that the protein was able to distinguish between specific and non-specific targets; the protein fluorescence was quenched more significantly by the specific target than by the non-specific target. Secondly, addition of salt to 2 M lead to recovery of the protein fluorescence indicating that the stability of DNA–protein interaction has a large electrostatic contribution in both the cases.

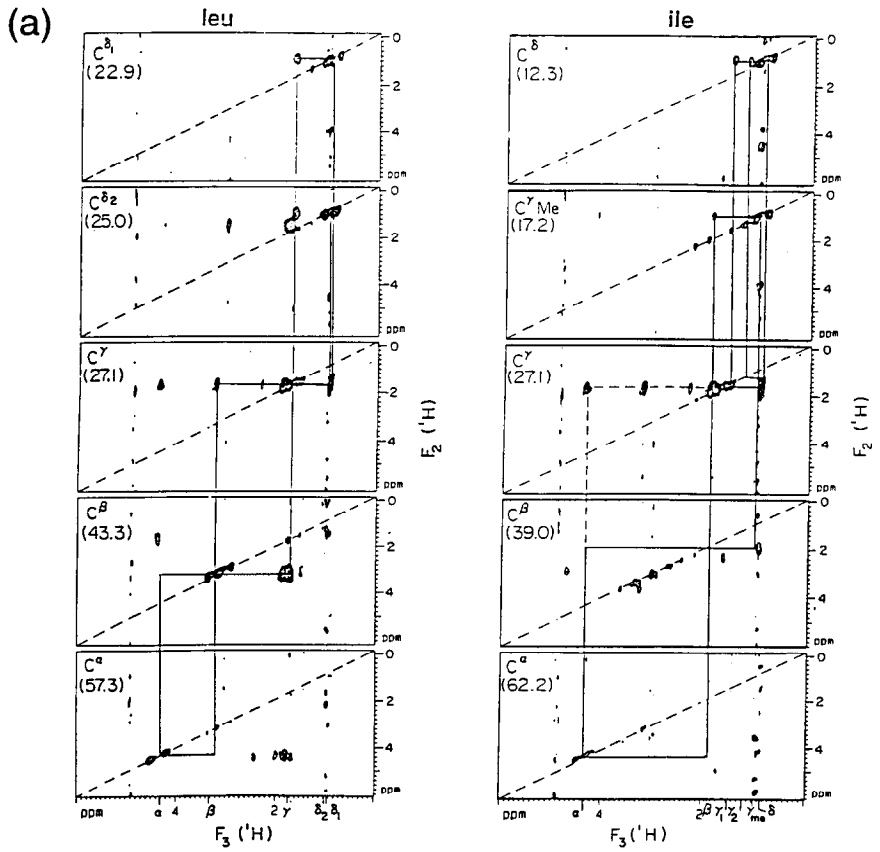
The involvement of the hydrophobic core in DNA binding has been inferred by competition experiments involving TNS. Fig. 5 shows TNS fluorescence in the protein–TNS complex, before and after addition of specific and non-specific DNA targets. We observed that addition of specific DNA displaces TNS from the complex leading to loss of TNS fluorescence. In contrast, the non-specific DNA sequence had no effect on the TNS fluorescence in the protein–TNS complex. This implies, in the case of specific DNA, that either both TNS and DNA bind at the same site but the DNA has a larger affinity compared to TNS or, the binding of DNA causes a conformational change in the protein which leads to loss of the hydrophobic pocket in the protein and thus liberates TNS into the solution. In separate experiments, we observed that the CD spectrum of the protein does not change significantly on DNA binding. Besides quantitative binding constant measurements indicated that specific DNA has a greater affinity than TNS for the protein. Thus, we believe that the binding sites for the two are the same. In other words, the specific DNA does indeed bind in or very near the hydrophobic core of the protein. In the case of non-specific DNA, it is clear that the DNA is not able to liberate TNS from the complex, and its mode of binding is likely to be different.

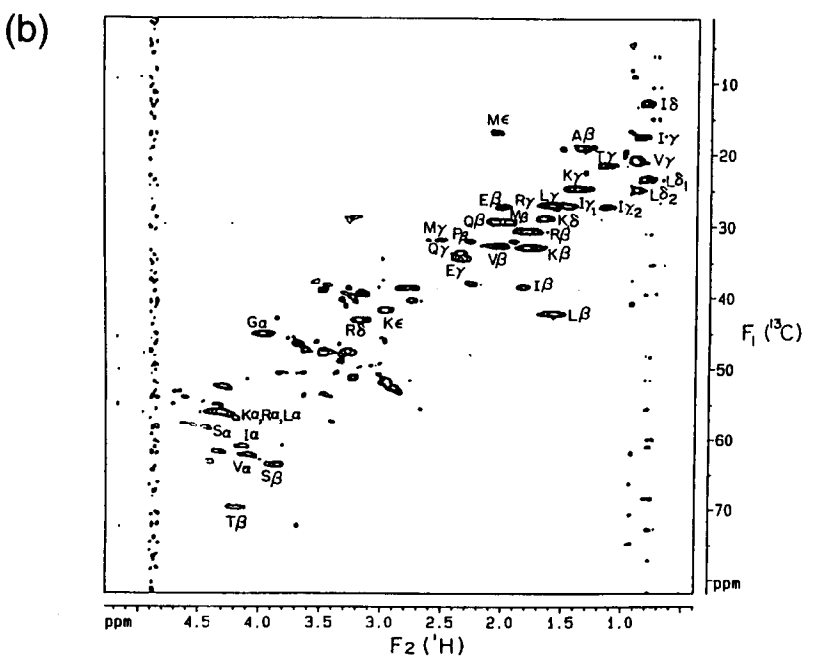
To further quantitate tryptophan involvement in DNA binding, we have carried out acrylamide quenching experiments on different DNA–protein complexes. The results are summarised in Table 1. We notice that the environments of the specific and non-specific DNA binding sites in terms of the number and types of tryptophans (buried or exposed) involved in the binding are clearly different. The specific binding site has a larger number of buried tryptophans in the environment, while the non-specific site has only a single exposed tryptophan in its vicinity. Thus the binding sites for specific and non-specific DNA targets are different on the R123 protein.

Table 1
Dynamic quenching of protein fluorescence by acrylamide in different protein DNA complexes

DNA	Fluorescence in absence of quencher (%)	Number of fluorophore tryptophans	Acrylamide quenching in different complexes (%)	Solvent-exposed fluorophore tryptophans
None	100	9	33	3
MRE-16	48	4	12	1
MRE-12	52	5	12	1
DSNs-12	88	8	24	2

MRE-16 = CCTAACTGACACACAT; MRE-12 = ACCGTTAACGGT; DSNS-12 = GGTACGCGTACC.


(a)
leu
ile
C^{δ_1} (22.9)
C^{δ_2} (25.0)
C^{γ} (27.1)
C^{β} (43.3)
C^{α} (57.3)
C^{δ} (12.3)
C^{γ}Me (17.2)
C^{γ} (27.1)
C^{β} (39.0)
C^{α} (62.2)
F_2 ('H)
F_3 ('H)
ppm


(b)
Mε
Aβ
Iδ
Iγ
Lγ
Vγ
Kγ
Lδ1
Lδ2
Eβ
Rγ
Qβ
Mβ
Kδ
Iγ1
Iγ2
Mγ
Pβ
Rβ
Qγ
Vβ
Kβ
Eγ
Iβ
Lβ
Gα
Rδ
Kε
Kα,Rα,Lα
Sα
Iα
Vα
Sβ
Tβ
F_1 (^{13}C)
F_2 ('H)
ppm

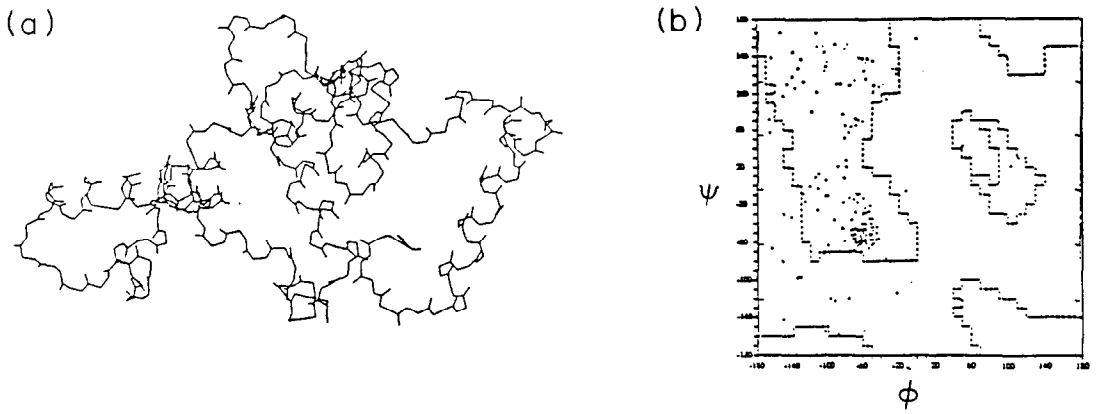


Fig. 7. (a) Backbone trace of the final energy minimized structure of the protein. (b) Ramchandran plot of backbone torsion angles for the structure. The MD calculations were performed with the protein immersed in a water matrix which was big enough to have at least two water molecules beyond the protein molecule in every direction.

Similar competition experiments involving reaction with the –SH group of the cysteine by the reagent dithionitrobenzoate, indicated that specific DNA protects the *cys* residue from reacting while the non-specific DNA is unable to do so. This again points to the same conclusion that the two binding sites have different environments.

6. NMR and MD studies on the R123 protein

The spectroscopic studies described above provided very valuable information regarding the topography and DNA binding properties of the protein. A complete molecular level understanding of the interaction requires however, a full knowledge of the three dimensional structures of the two molecules. NMR in conjunction with MD calculations is undoubtedly the most powerful technique for such purposes. As far as the protein is concerned our efforts in this direction are continuing with multidimensional experiments on labelled and unlabelled protein samples. Two dimensional NOE spectra gave indications to the presence of helical structures (NH–NH NOEs) and also of some long range order (aromatic protons to methyl proton NOEs) in the protein. However, the chemical shift dispersion in the protons was certainly not adequate to obtain specific assignments. Experiments with the labelled sample allowed use of the characteristic patterns for ^{13}C shifts for the different residues and we could obtain residue type spin system identifications by a combination of different two- and three-dimensional NMR experiments. Fig. 6a shows an illustrative assignment from a 3D HMQC–TOCSY [30] spectrum and Fig. 6b gives a summary of all the residue type identifications in a ^{13}C HMQC spectrum. We notice that the carbon chemical shift dispersion is also poor for any amino acid type. Similar ^{15}N resolved 3D experiments showed that the ^{15}N dispersion is somewhat better and spin systems of the glycines, alanines, serines, valines and a few other residues could be separately identified. However, it is clear to us that a much higher dispersion is required for a complete and unambiguous identification of all the spin systems. Three dimensional experiments such as HNCA,

Fig. 6. (a) An illustrative spin system identification using the different planes from a ^{13}C resolved ^{1}H–^{1}H TOCSY 3D spectrum of the R123 protein. (b) A two dimensional ^{13}C–^{1}H HMQC spectrum of the protein with residue type identifications marked.

HNCOCA, etc., [30] aimed at obtaining assignments along the backbone did not have enough resolution to separate the peaks belonging to individual residues.

We observed in the NMR spectra at different temperatures that the chemical shifts of the amide as well as aliphatic protons do not change dramatically as the temperature is increased. Progressive denaturation of the protein causes only a small change in the intensity patterns in the aliphatic regions, and progressive loss of the amide proton signals. This leads to better dispersion of the amide protons although only a fraction of them are seen at higher temperatures. The residual ones must be those belonging to residues which unfold later in the denaturation process. This leads to the speculation that these must belong to the hydrophobic core in the protein. It has been possible to identify these residue types in the spectra and consistent with the speculation the residues are all non-polar types, such as aromatic residues, valines, alanines, isoleucines, threonines, leucines etc.

Next, with a view to identify the energetically most favoured conformation for the protein, which we anticipated might help in the interpretation of complex NMR spectra, which in turn might help in refining the structure, we carried out molecular dynamic simulations on the protein in a water matrix using the X-plor force field and standard parameters for various energy functions [31]. Starting with a predicted secondary structure, keeping the disordered portions in a random coil state the protein was heated to 1000 K taking care that the helices are not broken, and then cooled slowly to 300 K. Then incorporating a coupling to the temperature bath, constant temperature calculation was carried out until a stable structure with stable energy was reached. The total dynamics time in these calculations was about 500 ps. The average of the last few ps struc-

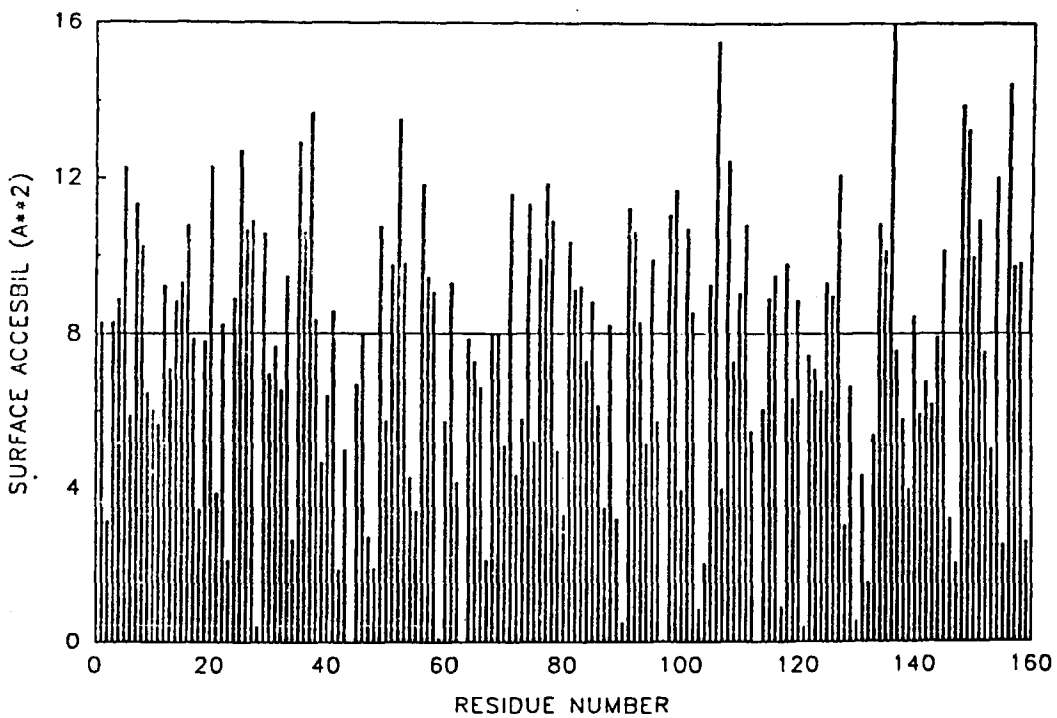


Fig. 8. Surface accessibility of the individual residues in the energy minimised R123 structure shown in Fig. 7a, calculated according to the Lee–Richards algorithm [32]. An accessible area of 16 A^2 for a rolling water implies complete accessibility to the solvent. From this data we noticed that the solvent accessibility of the tryptophans is in accord with the results of the acrylamide quenching experiments.

tures was energy minimised and this has been analysed for specific details. Fig. 7a shows a backbone trace of the above structure and Ramachandran plot of the backbone torsion angles of the individual residues is shown in Fig. 7b. Fig. 8 shows an analysis of the surface accessibility of the individual residues calculated using the Lee–Richard rolling water model [32]. These in a way reflect the hydrophobicity of the residues and thus are useful in estimating their participation in the formation of the hydrophobic core in the protein. An accessible area of 16 A^2 implies complete accessibility to the solvent, an area of 8 A^2 implies 50% solvent accessibility, etc. We notice that about 50% of the residues are less than 50% accessible and these have contributions from all the three repeats. Specifically the results for the tryptophans are in accord with those derived from acrylamide quenching experiments.

7. NMR structure of *myb* cognate DNA sequence

The *myb* cognate DNA sequence was originally identified to be YAACKG [8] where Y refers to a pyrimidine and K refers to a G/T. Recently this has been extended to YAACKGHH [12] with H representing A/T/C. Noting that H is not really a very specific base, its role is implicated to be mostly in non-specific interaction and in enhancing the affinity of the interaction. AAC is the core recognition sequence and so also is its complementary sequence GTT. Therefore, to understand the atomic level structure of this sequence, we have determined the structure of a 12 mer oligonucleotide D-ACCGTTAACGGT by NMR in conjunction with restrained energy minimization. The procedure involves quantitative interpretation of J-correlated E. COSY [33] and NOE correlated [34] spectra by spectral simulation [35] and relaxation matrix [36] calculation respectively. Fig. 9 shows typical simulations of cross peak patterns in E. COSY spectra and also quantitative comparison of calculated and experimental NOE intensities for the final derived structure of the molecule. In Fig. 10a, the structure of the molecule is shown in stereo view. An interesting feature of the structure is that it exhibits three center H-bonds at AC and AA steps in both the strands (Fig. 10b). This results in a distortion of the base plane geometry at

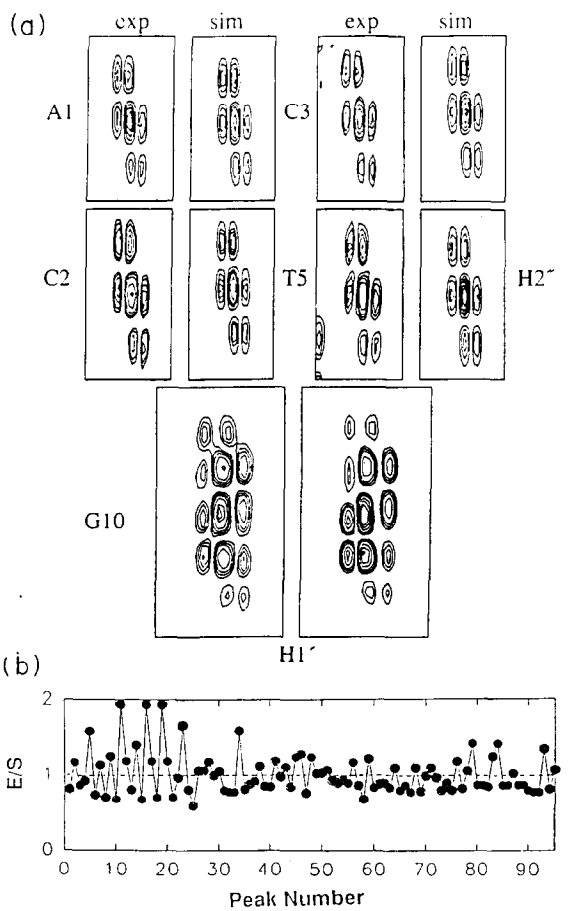


Fig. 9. (a) Typical spectral simulations of J-correlated spectra using the program SICOS [35]. (b) Intensity fits between calculated and experimental NOEs using the program SIMNOE [37] for the final structure of the 12 mer DNA segment.

these locations and also causes a bend in the backbone of the DNA. These structural features may be of significance in specific *myb*-DNA interaction.

8. Concluding remarks

We have described in this paper the salient features of the DNA binding domain of *Drosophila myb* protein, the topographical and secondary structure characteristics as determined by fluorescence and circular dichroism spectroscopic techniques, the DNA binding properties of the recombinant protein, some insight into the hydrophobic core of the protein

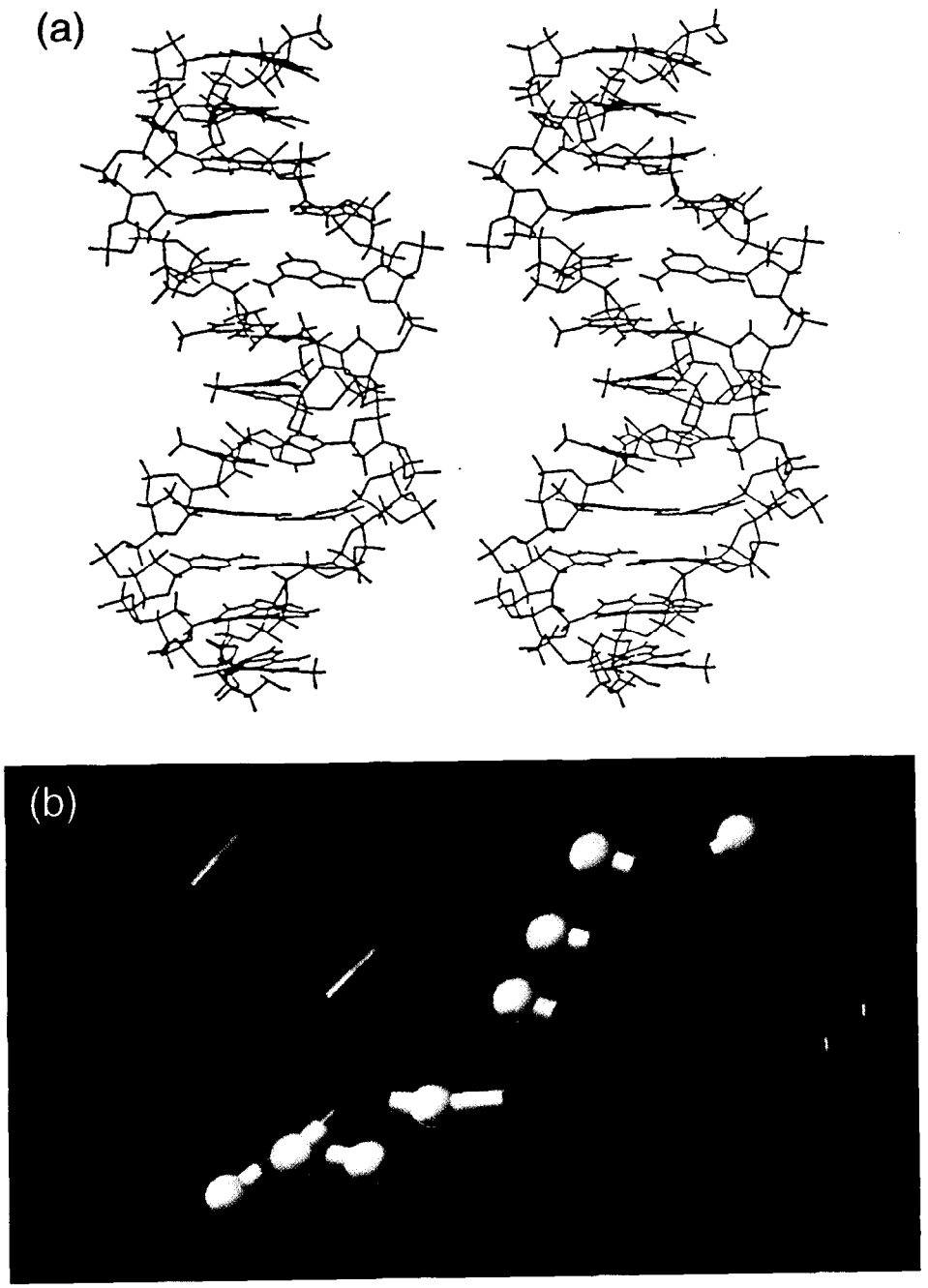


Fig. 10. (a) Stereo view of the structure of the 12 mer DNA segment. (b) A three center H-bond observed at the AC step.

as determined by fluorescence, NMR and MD simulations and finally the NMR structural features of the recognition DNA sequence. It is worth mentioning at this stage that the *myb* proteins from different sources such as mouse, chicken, human and *Drosophila* possess some intrinsic differences. This may not be

surprising considering that in the R1 repeat the sequence homology is rather poor, and that important structural differences among proteins having only a few amino acid differences are not rare in the literature. Even so the fact that the proteins recognise the same DNA sequence suggests that the protein has the ability to adjust to the dictates of the DNA structure. The DNA binding experiments with different DNA sequences have indicated that the protein does bind to non-specific targets as well, but the sites for the specific and non-specific targets are physically different.

Acknowledgements

The support provided by the National Facility for High Field NMR supported by the Government of India and located at TIFR are gratefully acknowledged. We thank Drs. M.V. Hosur and K.K. Kannan at the Bhabha Atomic Research Centre for the use of the X-plor package.

References

[1] G.L.C. Shen-Ong, Biochem. Biophys. Acta 1032 (1990) 39–52.
[2] B. Luscher, R.N. Eisenman, Genes Dev. 4 (1990) 2235–2241.
[3] Y. Nishima, H. Nakagoshi, F. Imamoto, T.J. Gonda, S. Ishii, Nucl. Acids. Res. 17 (1989) 107–117.
[4] K. Weston, J.M. Bishop, Cell 58 (1989) 85–93.
[5] H. Nakagoshi, T. Nagase, Y. Ueno, S. Ishii, Nucl. Acids. Res. 17 (1989) 7315–7324.
[6] S.A. Ness, A. Marknell, T. Graf, Cell 59 (1989) 1115–1125.
[7] H. Beidenkapp, U. Borgmeyer, A.E. Sippel, K.-H. Klempnauer, Nature 335 (1988) 835–837.
[8] H. Nakagoshi, T. Nagase, C. Kanei-Ishii, Y. Ueno, S. Ishii, J. Biol. Chem. 265 (1990) 3479–3483.
[9] H. Sakura, C. Kanei-Ishii, T. Nagase, H. Nakagoshi, T.J. Gonda, S. Ishii, Proc. Natl. Acad. Sci. USA 86 (1989) 5758–5762.
[10] T. Oehler, A. Hannelore, H. Beidenkapp, K.-H. Klempnauer, Nucl. Acids. Res. 18 (1990) 1703–1710.
[11] K.M. Howe, C.F.L. Reakes, R.J. Watson, EMBO J. 9 (1990) 161–169.
[12] K.M. Howe, K. Watson, Nucl. Acids. Res. 19 (1991) 3913–3919.
[13] K. Weston, Nucl. Acids. Res. 20 (1992) 3043–3049.
[14] K. Ogata, H. Hojo, S. Aimoto, T. Nakai, A.S. Nakamura, S. Ishii, Y. Nishimura, Proc. Natl. Acad. Sci. USA 89 (1992) 6428–6432.
[15] N. Jamin, O.S. Gabrielsen, N. Gilles, P.-N. Lirsac, F. Toma, Eur. J. Biochem. 216 (1993) 147–154.
[16] K. Ogata, S. Morikawa, H. Nakamura, A. Sakikawa, T. Inoue, H. Kanai, A. Sarai, S. Ishii, Y. Nishimura, Cell 79 (1994) 639–659.
[17] A. Ebneth, O. Schweers, H. Thole, U. Fagin, C. Urbanke, G. Maass, H. Wolfes, Biochemistry 33 (1994) 14586–14593.
[18] J. Frampton, T.J. Gibson, S.A. Ness, G. Doderlein, T. Graf, Protein Eng. 4 (1991) 891–901.
[19] M.D. Carr, U. Wollborn, P.B. Mcintosh, T.A. Frenkiel, J.E. McCormick, C.J. Bauer, K.H. Klempnauer, J. Feeney, Eur. J. Biochem. 235 (1996) 721–735.
[20] K. Ogata, S. Morikawa, H. Nakamura, H. Hojo, S. Yoshimura, R. Zhang, S. Aimoto, Y. Ametani, Z. Hirata, A. Sarai, S. Ishii, Y. Nishimura, Nature Struct. Biol. 2 (1995) 309–320.
[21] A. Madan, R.V. Hosur, L.C. Padhy, Biochemistry 33 (1994) 7120–7126.
[22] P.K. Radha, A. Madan, L.C. Padhy, R.V. Hosur, Curr. Sci. 66 (1994) 287–294.
[23] A. Madan, P.K. Radha, R.V. Hosur, L.C. Padhy, Eur. J. Biochem. 232 (1995) 150–158.
[24] A. Madan, P.K. Radha, A. Srivastava, L.C. Padhy, R.V. Hosur, Eur. J. Biochem. 230 (1995) 733–740.
[25] P.K. Radha, A. Madan, L.C. Padhy, R.V. Hosur, Proc. Ind. Acad. Sci. Chem. Sci. 106 (1994) 1537–1549.
[26] P.K. Radha, A. Madan, R. Nibedita, R.V. Hosur, Biochemistry 34 (1995) 5913–5922.
[27] H. Mach, C.R. Middugh, R.V. Lewis, Anal. Biochem. 200 (1992) 74–80.
[28] S.S. Lehrer, Biochemistry 10 (1971) 3254–3262.
[29] B. Rost, C. Sander, J. Mol. Biol. 232 (1993) 584–599.
[30] T.L. James, N.J. Oppenheimer (Eds.). Methods in Enzymology Vol. 239 (1994) Part C, Academic Press, New York.
[31] A.T. Brunger, X-Plor version 2.1 Manual, fellows of Harvard University, New York, 1990.
[32] B. Lee, F.M. Richards, J. Mol. Biol. 55 (1971) 379–400.
[33] C. Griesinger, G. Otting, K. Wuthrich, R.R. Ernst, J. Am. Chem. Soc. 110 (1988) 7870–7872.
[34] J. Jeener, B.H. Meier, P. Bachmann, R.R. Ernst, J. Chem. Phys. 71 (1979) 4546.
[35] A. Majumdar, R.V. Hosur, Prog. NMR. Spectrosc. 24 (1992) 109–158.
[36] S. Macura, R.R. Ernst, Molec. Phys. 41 (1980) 95–117.
[37] R. Nibedita, R. Ajay Kumar, A. Majumdar, R.V. Hosur, J. Biomol. NMR 2 (1992) 467–476.