Nucleic Acids Research, 2001, Vol. 29, No. 2 **499–505**

# Solution structure and dynamics of GCN4 cognate DNA: NMR investigations

**Purnima Khandelwal, S. C. Panchal, P. K. Radha and R. V. Hosur***

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

Received August 23, 2000; Revised October 23, 2000; Accepted November 1, 2000

## ABSTRACT

**A 12 bp long GCN4-binding, self-complementary duplex DNA d(CATGACGTCATG)$_2$ has been investigated by NMR spectroscopy to study the structure and dynamics of the molecule in aqueous solution. The NMR structure of the DNA obtained using simulated annealing and iterative relaxation matrix calculations compares quite closely with the X-ray structure of ATF/CREB DNA in complex with GCN4 protein (DNA-binding domain). The DNA is also seen to be curved in the free state and this has a significant bearing on recognition by the protein. The dynamic characteristics of the molecule have been studied by $^{13}$C relaxation measurements at natural abundance. A correlation has been observed between sequence-dependent dynamics and recognition by GCN4 protein.**

## INTRODUCTION

Specific protein–DNA recognition is a basic step in genetic expression and regulation. Protein–DNA interactions have been found to be sequence specific and one of the following mechanisms is used to achieve this specificity (1,2): (i) hydrogen bonding and van der Waals interactions of DNA base pairs with protein side chains which support only the correct DNA sequence; (ii) a sequence-dependent ability of DNA to distort upon protein binding; (iii) reorganization of DNA into a distorted structure bound only by the appropriate protein. This shows the importance of knowledge of the precise structure and the inherent dynamics of the DNA molecule for the sequence-specific recognition process. Indeed, sequence-dependent dynamic variations in DNA have been reported in the literature (3,4)

The yeast transcriptional factor GCN4, belonging to a large family of eukaryotic bZIP proteins, recognizes cognate DNA via a motif (the leucine zipper) located at the C-terminus of its 60 residue long DNA-binding domain (5,6). The protein binds DNA as a dimer in aqueous solution and has similar affinities for two sequences: ATGACTCAT (AP1 site) and ATGACGTCAT (ATF/CREB site) (7,8). Crystal structures of complexes of the DNA-binding domain with both the above sequences have been published (9,10). The structures reveal that stabilization of the structure is attained by direct interaction of the protein with DNA. The protein holds the DNA with a forcep-like arrangement, in which the two arms of the fork are long helices, each interacting with ~3–4 bp in the DNA segment in a symmetrical manner. We notice that the two recognition sequences listed above are identical except for an additional G in the middle in the ATF/CREB sequence that displaces the two ATGA contact surfaces by an axial translation of 3.4 Å and a twist angle of 34°, thus providing a classic example in which to determine the extent of adaptability of the structures. It is observed that in the complex the CREB DNA is bent in the middle by ~20° and this brings the two half-sites in the DNA into the correct positions for protein interaction. The tertiary structure of the protein (relative orientation of the interacting arms) also undergoes changes as compared to that in the AP1–protein complex to accommodate an extra base pair, with a consequent twist in the orientations of the half-sites. The ability of GCN4 to bind to both the sequences as specific targets could be a result of flexibility of the α-helical segment of the protein or induced structural arrangements of one or both of the DNA targets or intrinsic deformation of the B-form structure of one target site. These observations suggest that the course of the DNA backbone, its deformability and dynamics help significantly in bringing about specific interactions. With this in mind we have here investigated by NMR the solution structure and dynamics of the ATF/CREB DNA sequence. A previous report on the solution structure of the CREB site DNA by NMR (11) was based on rather few NOEs for the size of the molecule and the structure essentially represented an unrefined model of the DNA. We also observed some differences from our data in the stereospecific assignments of the sugar H2′ and H2″ protons. Therefore, we have independently determined the structure with our own NMR data and this constitutes the basis of dynamics investigations by heteronuclear relaxation measurements at natural abundance. The results have provided an interesting insight into the role of local flexibility in the recognition process.

## MATERIALS AND METHODS

### Sample preparation

The oligonucleotide was synthesized on an Applied Biosystems 392 automated DNA synthesizer on the 10 μM scale using

*To whom correspondence should be addressed. Tel: +91 22 215 2971; Fax: +91 22 215 2110; Email: hosur@tifr.res.in
Present addresses:
Purnima Khandelwal, Biophysics Research Division, University of Michigan, 930 North University Avenue, Ann Arbor, MI 48109, USA
P. K. Radha, Department of Molecular Biology and the Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

solid phase β-cyanoethyl phosphoramidite chemistry, cleaved from the support and purified by standard procedures (12,13). Purity was checked by electrophoresis on a 20% polyacrylamide gel. The sample was dissolved in 10 mM phosphate buffer prepared in $H_2O$ or $D_2O$ to which 0.1 M NaCl and 0.1 mM EDTA were added. Sodium 3-trimethylsilyl-(2,2,3,3-$^2H_4$) propionate was added for internal referencing of $^1H$ chemical shifts.

### NMR data acquisition

*$^1H$ experiments*. NMR data were obtained either on a Varian Unity-*plus* 600 MHz or a Bruker AMX 500 MHz spectrometer. Standard presaturation techniques were used for solvent suppression. Phase-sensitive NOESY (14) and TOCSY (15) spectra in $D_2O$ were recorded with mixing times of 80, 125, 160, 225, 275 and 375 ms for NOESY and 80 ms for TOCSY at 35°C with 2048 × 512 data points and 16 scans. High resolution E-COSY (16) and DQF-COSY (17) spectra were acquired with 4096 × 600 data points for $^1H$-$^1H$ coupling constants estimation. Zero filling to 4096 × 2048 real data points and a sine-squared bell window function shifted by 45° along both dimensions were used for data processing by Felix v.97.0 software (MSI).

*Heteronuclear experiments*. $^1H$-$^{13}C$ gradient HSQC (18) and gradient $^{13}C$-edited TOCSY with sensitivity enhancement with a TOCSY mixing time of 60 ms were recorded with 32–80 scans having different spectral widths and offsets to cover the entire $^{13}C$ frequency range and achieve maximum resolution along the $F_1$ axis. Moreover, a spectrum with a full $F_1$ spectral width of 24 140 Hz was also acquired for referencing purposes. Two-dimensional $^1H$-$^{31}P$ heteroTOCSY, hetero/homoTOCSY and heteroTOCSY-NOESY experiments (19) were recorded at 308 K with different TOCSY mixing times and a NOESY mixing time of 200 ms with an $F_1$ spectral width of 270 Hz, and 320 scans were used for averaging. Pulse sequences with different cross-polarization mixing sequences, like DIPSI-2, DIPSI-3 and DIPSI-2rc (20), using gradients and cw presaturation or multiple pulse presaturation were tested. DIPSI-2rc with multiple pulse presaturation yielded the best results.

### Structure calculations

*Experimental restraints*. Cross-peaks in the NOESY spectra were integrated and intensities in the low mixing time spectrum were translated into inter-proton distances using H5-H6 cross-peaks of Cyt as the reference (2.46 Å). A range of ±0.3 Å was allowed for the distances to account for any errors in integration. Pseudo-atom corrections were used for methyl and other equivalent protons. Force constants of 100, 10 and 1 kcal/mol/rad were fixed for strong, medium and weak peaks. Two hydrogen bond constraints per hydrogen bond between base pairs were also used. Further constraints were derived from analysis of the E-COSY spectrum, which provided dihedral angle ranges for protons in the sugar rings and indicated that the sugar geometries were in the S-domain.

*Molecular modeling and simulated annealing*. Since the coupling constant data and the sequential NOE correlations indicated that the topology of the duplex belongs to the B-DNA family, we decided to perform a conformational search within this domain only, rather than wasting computational effort doing an *ab initio* search from random conformations. An initial model of double helical B-DNA structure was generated using the Biopolymer module of InsightII v.97.0 (MSI) on a Silicon Graphics Indigo2 workstation. The 3′- and 5′-ends were capped to terminate biopolymer building. The energy of the molecule was minimized first with 100 steps each of steepest descents and conjugate gradients to remove any short contacts in the starting structure using the Amber force field in Discover v.97.0 (MSI). A distance-dependent dielectric constant (1.0*$r$) was used. A conformational search was performed by the following simulated annealing–restrained molecular dynamics (rMD) protocol. The molecule was 'heated' to a temperature of 1000 K and equilibrated at this temperature for 5 ps. Dynamics was continued for 50 ps during which 15 structures were saved at regular intervals. Each of these was then slowly cooled to 300 K in 50 K steps. At each temperature step the molecule was equilibrated for 5 ps. At the end of simulated annealing all the structures were energy minimized by steepest descents followed by conjugate gradients until a predefined convergence limit (RMSD < 0.001) was reached.

*Relaxation matrix refinement*. These structures were then selected to perform relaxation matrix refinement using the IRMA (iterative relaxation matrix analysis) protocol in Discover (21). The input data included NOE intensities from 80, 125, 160, 225 and 275 ms spectra, chemical shift assignments and p.p.m. values for all assigned protons. Distance restraints for pseudoatoms were not included in IRMA calculations. The isotropic correlation time as derived from the relaxation measurements described before was further refined to give the best NOE build-up curve fit, which yielded a value of 3 ns. The relaxation matrix approach entails the generation of a theoretical relaxation matrix and a set of theoretical NOE intensities based on an initial structure. The experimental NOE intensities measured as a function of mixing time are supplemented with the theoretical NOE values. A new relaxation matrix is calculated from this mixed NOE intensity matrix and a new set of distance restraints is deduced. These relaxation matrix calculations take into account multi-spin effects. The refined distances obtained from the above set of calculations are then used in a short rMD protocol to generate the starting structure for the next IRMA cycle. The full relaxation matrix calculations are carried out again and the distances are refined further. These second round refined distances are used once more in a rMD protocol, completing the IRMA loop for a second time. The cycle is then repeated until convergence is reached. Convergence is determined via the $R$ factors (a measure of the difference between the theoretical and experimental NOE intensities) which are calculated during each IRMA loop. Once an $R$ factor does not decrease significantly over two or three cycles, the IRMA procedure is complete. The most important $R$ factor is $R1$, which is defined as

$$R1 = [\sum_{i,j} W_{ij}(\tau_m) \mid A_{ij}^{calc}(\tau_m) - A_{ij}^{exp}(\tau_m) \mid ] / [\sum_{i,j} W_{ij}(\tau_m) \mid A_{ij}^{exp}(\tau_m) \mid ]$$

where $A_{ij}$ is the intensity of the cross-peak between the two protons $i$ and $j$ for a given NOESY mixing time $\tau_m$ and $W_{ij}$ are the weight factors included to make allowance for errors in experimental intensities. RMSD analysis and restraints violation statistics were generated for all the structures along with measurements of dihedral angles and other structural parameters.

### Relaxation measurements

$^{13}C$ $T_1$, $T_2$ and steady-state {$^1H$}-$^{13}C$ NOEs were measured following established methods (22–24), which use pulse field gradients for coherence transfer pathway selection combined with sensitivity enhancement (25,26). Quadrature detection along the indirectly detected dimension was achieved by the States-TPPI method (27). $T_1$ and $T_2$ spectra were recorded as 100 × 2048 complex matrices with 96 scans per complex $t_1$ point and spectral widths of 3500 Hz and 8000 Hz in the $F_1$ and $F_2$ dimensions with recycle delays of 2 and 1.5 s (including acquisition time), respectively. For $T_1$ measurements 13 experiments with inversion recovery delays of 10.016, 20.032, 40.064, 60.096, 90.144, 110.176, 140.224, 220.320, 290.464, 390.624, 490.684, 590.944 and 691.104 ms were recorded. For $T_2$ measurements 10 experiments with nine transverse relaxation delays of 8.382, 16.765, 25.147, 33.530, 41.912, 58.677, 67.059, 83.824, 92.206, 108.971 and 117.354 ms were recorded. Two spectra, one with and one without $^1H$ saturation, with 256 transients for each complex $t_1$ point were recorded to yield steady-state {$^1H$}-$^{13}C$ NOEs. Spectral width along the $F_1$ and $F_2$ dimensions and the matrix sizes were the same as used in the $T_1$ and $T_2$ experiments. A recycle delay of 5 s was used for the spectrum recorded in the absence of proton saturation, whereas a 2 s recycle delay followed by a 3 s period of proton saturation was used in the NOE experiment. $^1H$ saturation was achieved with a burst of 120° $^1H$ pulses at 5 ms intervals (28).

Relaxation rate constants and NOE enhancements were calculated from peak heights in the cross-sections through the HSQC spectra. The relaxation rate constants were obtained from non-linear fits of monoexponential functions for longitudinal and transverse relaxation using the Levenburg–Marquardt algorithm. The measured peak intensities for different delays from the $T_1$ and $T_2$ experiments were fitted to

$$I(t) = a + b \exp(-R_{1,2}t)$$

where $I(t)$ is the intensity at time $t$, $a$ is the intensity corresponding to the steady-state value of the magnetization, $a + b$ is the magnetization value at the beginning of the relaxation period $t$ and $R_1 = 1/T_1$ and $R_2 = 1/T_2$ are the phenomenological spin-lattice and spin-spin rate constants. The errors in estimating $R_1$ and $R_2$ were obtained as standard errors from the Levenburg–Marquardt algorithm. The steady-state {$^1H$}-$^{13}C$ NOEs were calculated as

$$\mathrm{NOE} = I_{sat}/I_{eq}$$

where $I_{sat}$ and $I_{eq}$ are the intensities of a peak from the spectra with and without $^1H$ saturation, respectively. The uncertainty in the NOE values was obtained by recording duplicate spectra and analyzing in a similar way.

### Analysis of relaxation data and model-free parameters

The relaxation of a methine $^{13}C$ nucleus at natural abundance at high field strengths is dominated by the dipolar interaction with the directly attached proton and by the chemical shift anisotropy (CSA) mechanism (29). The dipolar relaxation rate constants and steady-state NOE are described in terms of five spectral density functions $J(\omega)$ at frequencies $\omega_H$, $\omega_C$, $\omega_H + \omega_C$, $\omega_H - \omega_C$ and $J(0)$. The CSA ($\sigma_{||} - \sigma_{\perp}$) value depends on the type and hybridization of the atom. For a $sp^3$-hybridized C1′ carbon the value is 41 p.p.m. (30). To calculate these five unknown spectral density functions five known experimental values are required. However, generally the experimental relaxation data available are insufficient and, to circumvent this problem, Lipari and Szabo proposed a model-free formalism (31,32), which was subsequently extended by Clore *et al.* (33) to obtain better fits to experimental data. The spectral density function is described in terms of a certain number of dynamic variables: the order parameter ($S^2$), internal correlation time ($\tau_e$) and overall correlation time ($\tau_m$). Conformational exchange rates are also included where needed. Depending upon the symmetry of the molecule the diffusion tensors are different and they also contribute to the spectral densities. A duplex DNA molecule that has a cylindrical shape can be aptly described by an axially symmetrical diffusion tensor. Therefore, model-free calculations were performed assuming an axially symmetrical diffusion tensor using Model-free v.4.0 and the associated suite of programs provided by Dr A. G. Palmer. We also calculated the model-free parameters under the assumption of isotropic rotational diffusion for the purpose of comparison with results from the axially symmetrical tensor.

We analyzed the relaxation data following the protocol of Mandel *et al.* (34), according to which five different dynamic models can be used to fit the data. Initially the diffusion tensor [or overall rotational correlation time ($\tau_m$) in the case of the isotropic model] was evaluated using the solution structure of the molecule and the $R_2/R_1$ ratios; for the isotropic case the solution structure is not required. Then, for each residue a model with the minimum number of parameters required to fit the experimental $R_1$, $R_2$ and NOE data was selected. After this, the diffusion tensor (correlation time in the case of the isotropic model) parameters were optimized along with the motional parameters using the Powell minimization algorithm. At each step the fit was monitored by the $\chi^2$ function given by

$$\chi^2 = \sum[\{(R_{1i} - R_{1i}^c)^2/\sigma_{1i}^2\} + \{(R_{2i} - R_{2i}^c)^2/\sigma_{2i}^2\} + \{(\mathrm{NOE}_i - \mathrm{NOE}_i^c)^2/\sigma_{\mathrm{NOE}i}^2\} = \sum_i \Gamma_i$$

where the index $i$ stands for the spin number, $\sigma_{1i}$, $\sigma_{2i}$ and $\sigma_{NOEi}$ are the uncertainties in the experimental relaxation parameters $R_{1i}$, $R_{2i}$ and $\mathrm{NOE}_i$, $R_{1i}^c$, $R_{2i}^c$ and $\mathrm{NOE}_i^c$ are the respective calculated relaxation parameters and $\Gamma_i$ is the sum squared error for spin number $i$. Monte Carlo simulations using 300–500 randomly distributed synthetic data sets for $R_{1i}^c$, $R_{2i}^c$ and $\mathrm{NOE}_i^c$ were used to assess the goodness of fit and to calculate the statistical properties, like random error in the model-free parameters arising from experimental uncertainties.

## RESULTS AND DISCUSSION

### Resonance assignments

Exchangeable and non-exchangeable, except for H5′ and H5″, resonance assignments were obtained from standard protocols (35,36) based on TOCSY and NOESY experiments. Stereospecific assignments of the H2′/H2″ protons were obtained on the basis that the H1′–H2″ NOE is always stronger than the H1′–H2′ NOE at short mixing times, since the former corresponds to a shorter distance than the latter, irrespective of the sugar geometry. Assignment of the H5′ and H5″ protons was hampered due to extensive overlap. These were obtained using $^1H$-$^{31}P$ correlation experiments. $^1H$-$^{31}P$ TOCSY experiments yield bond sequential assignments of $^{31}P$ nuclei in DNAs (37): H3′($i$)→$^{31}$P($i$ + 1)→[H5′, H5″, H4′($i$ + 1)]→H3′($i$ + 1). Hence,

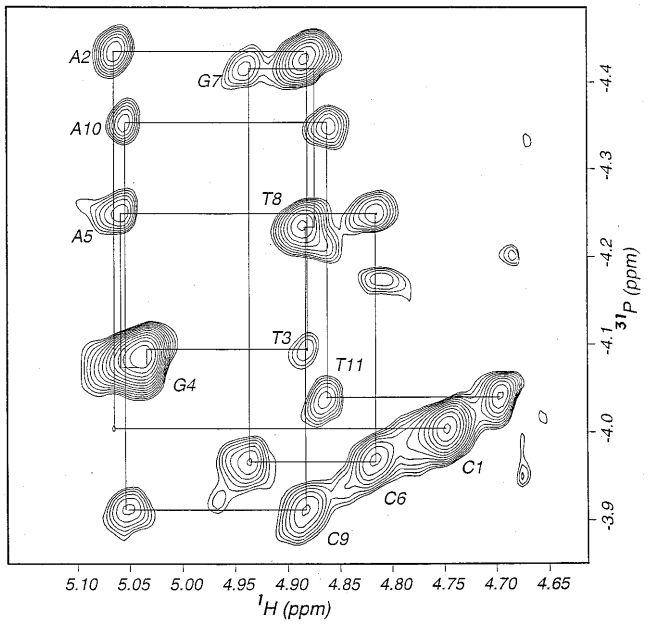


**Figure 1.** $^1$H-$^{31}$P TOCSY spectrum displaying sequential walk via H3′-$^{31}$P correlation. The labels identify correlations from each H3′ to its neighboring phosphorus atom at the 3′-end.

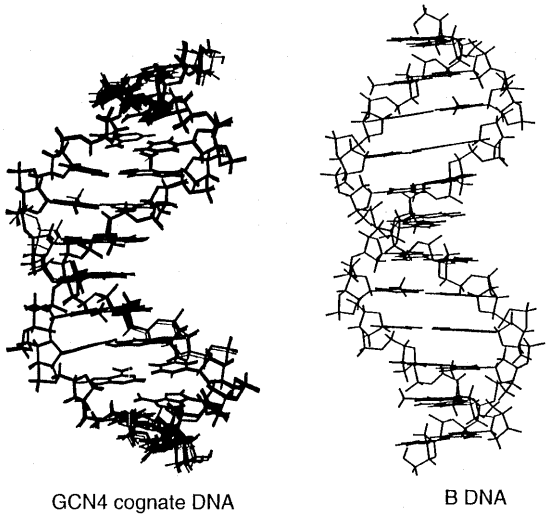


**Figure 2.** Visual comparison of the solution structure (15 final structures superimposed) of the cognate DNA (left) with a standard B-DNA structure (right) of the same sequence. The bend in the solution structure is clearly seen.

correlation is obtained between H3′ protons on the adjacent residues through the in-between $^{31}$P nucleus. These connections in the present DNA are shown in Figure 1. These were then utilized in a 7 ms heteroTOCSY experiment to obtain assignments of the 3′-end directly coupled H4′ protons, which in turn provide H5′/H5″ assignments. Assignments of H5′/H5″ resonances do not have stereospecificity and H5′ is assumed to resonate downfield from H5″ (38). The heteroTOCSY-NOESY experiment provided additional confirmation of all the assignments via the sequential H1′ connectivities (data not shown).

$^{13}$C assignments, which are necessary for dynamics studies, were obtained through HSQC and $^{13}$C-edited TOCSY experiments, where the latter could resolve some of the overlapping peaks. For instance, Cyt C6 carbon chemical shifts were immediately assignable because of magnetization transfer from H6 to H5 protons, which fall in a unique region of the HSQC spectrum. Moreover, C2′ carbon chemical shifts could be easily assigned on the basis of H1′ proton chemical shifts from the H2′/H2″ to H1′ cross-peaks. Table S1 in Supplementary Material summarizes the complete set of proton, carbon and phosphorus assignments obtained for the DNA.

## Sugar geometries

Sugar geometry information for the individual nucleotide units was obtained from the analysis of E-COSY spectra. It was observed that the H1′-H2′, H1′-H2″, H2′-H3′ and H2″-H3′ coupling constant values were closer to the values expected for the S-domain (B-DNA) rather than the N-domain (A-DNA) geometry. On this basis, qualitative ranges derived for the sugar geometries were considered sufficient as constraints and any selection of the dominant geometries from NOE-based simulations would have to satisfy these limits.

## Structure of the DNA

The structure of the DNA was calculated using simulated annealing–rMD and the IRMA refinement protocol as described in Materials and Methods. In two cycles of IRMA calculations the NOE $R1$ factor decreased to ~0.35 and did not change significantly thereafter. The input and structure convergence parameters are listed in Table 1. Figure 2 shows a superposition of the final 15 structures and a comparison with the standard B-DNA structure for the same sequence. It is interesting to see that the DNA is curved. Various structural parameters like global and local axis, base axis, base pair axis, base–base, intra-base and inter-base pair parameters, global axis, backbone and groove parameters were determined for the minimized average structure using the CURVES 5.0 program (39; data not shown). The sugar geometries for the different nucleotides show variations, but within the S-domain. The major groove is compressed and the minor groove is widened, which is a result of curvature in the DNA. Such a 'bending' of the DNA in the free state as seen here has also been observed by Paolella *et al.* (2) using a circular permutation assay and phase-dependent effects on electrophoretic mobility. Their results indicate that in solution the center of the CREB binding sequence is bent towards the major groove in the complex by a bend angle of 13°. Figure 3 shows a superposition of the average solution state structure over that in the X-ray structure of the ATF/CREB site DNA–GCN4 protein complex. The two structures are largely similar except for small differences in the trace of the backbone in the middle. This could be a consequence of complex formation.

## Dynamics of the DNA

We have investigated the dynamics of the DNA by relaxation measurements. Figure 4 shows the H1′-C1′ cross-peaks used to monitor relaxation at C1′ sites in the DNA and Figure 5 shows illustrative $R_1$ and $R_2$ fitting curves. We analyzed the data using

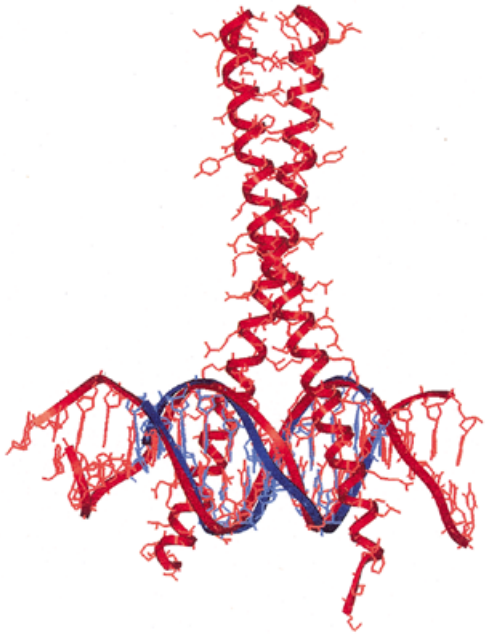

**Figure 3.** Superposition of solution structure of GCN4 cognate DNA (blue) over the DNA in the ATF/CREB–GCN4 complex.

both isotropic and axially symmetrical rotational diffusion tensors as described in Materials and Methods. Interestingly, we observed that the general qualitative conclusions with regard to order parameters were essentially the same in both cases. There were, however, small differences in the internal motional and exchange characteristics between the two cases. This is in accordance with earlier results from simulations by Schurr *et al.* (40), which showed that neglecting anisotropies results in overestimation of internal motions and exchange contributions to the relaxation data. In the following we discuss only the results from the axially symmetrical diffusion tensor.

**Table 1.** NMR restraints and convergence statistics for GCN4 cognate DNA

| | |
|---|---|
| Constraints | |
| NOE intensities | |
| all | 290 (145 per strand) |
| intra-nucleotide | 206 |
| sequential | 84 |
| hydrogen bond distances | 60 |
| Number of convergent structures | 15 |
| Average *R*1 factor | 0.35 ± 0.02 |
| RMSDs with average structure (Å) | 0.17 ± 0.15 |
| Average pairwise RMSD (Å) | 0.31 ± 0.11 |

Table 2 lists the relaxation parameters (A) and the model-free parameters (B) extracted as described in Materials and Methods for the case of the axially symmetrical diffusion

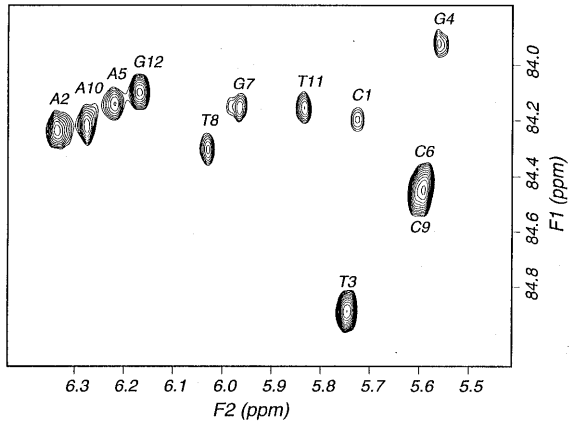


**Figure 4.** $^{1}H$-$^{13}C$ HSQC spectrum displaying cross-peaks used for monitoring relaxation at C1′ sites in the GCN4 cognate DNA.

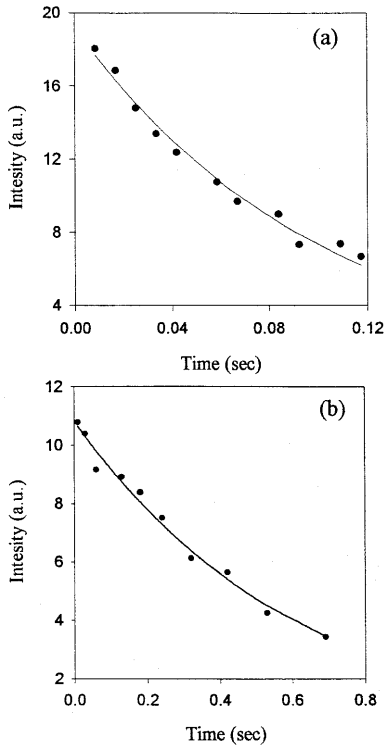


**Figure 5.** $^{13}C$ relaxation data for the GCN4 DNA: (**a**) $R_2$ relaxation decay curve for residue T11 and (**b**) $R_1$ relaxation decay curve for residue T3.

tensor. The sites of C6 and C9 could not be monitored due to overlap of the two peaks in the $^{1}H$-$^{13}C$ HSQC spectrum and for C1 NOE data are absent. The global $\tau_m$ of the 12mer DNA was estimated to be 3.15 ns. With regard to other model-free parameters we see a sequence-dependent microheterogeneity in the dynamics of the DNA. On average C1′ atoms are constrained on slow time scale motion. Different motional models (34) were required for the different residues, indicating that there are differences in the motional behaviors along the

**Table 2.** C1′ relaxation and model-free parameters for GCN4 cognate DNA

| (A) Primary relaxation data | | | | | |
|---|---|---|---|---|---|
| Nucleotide | $R_1$ ($s^{-1}$) | $R_2$ ($s^{-1}$) | NOE | | |
| C1 | 3.36 ± 0.81 | 15.73 ± 1.40 | – | | |
| A2 | 2.82 ± 0.53 | 21.17 ± 1.30 | 2.08 ± 0.19 | | |
| T3 | 1.53 ± 0.42 | 9.65 ± 0.43 | 1.58 ± 0.23 | | |
| G4 | 3.02 ± 0.63 | 15.60 ± 1.60 | 1.76 ± 0.26 | | |
| A5 | 2.49 ± 0.52 | 20.26 ± 1.20 | 1.21 ± 0.05 | | |
| G7 | 3.18 ± 0.88 | 15.43 ± 0.62 | 1.12 ± 0.17 | | |
| T8 | 1.64 ± 0.43 | 15.13 ± 0.47 | 1.93 ± 0.77 | | |
| A10 | 3.48 ± 0.42 | 18.10 ± 1.30 | 1.78 ± 0.04 | | |
| T11 | 1.65 ± 0.51 | 14.21 ± 0.86 | 1.37 ± 0.21 | | |
| G12 | 2.19 ± 0.36 | 14.22 ± 0.69 | 1.38 ± 0.19 | | |
| (B) Model-free parameters | | | | | |
| Nucleotide | $S^2$ (slow) | $S^2$ (fast) | $\tau_e$ (ps) | $R_{ex}$ ($s^{-1}$) | $\Gamma_i$ |
| C1 | – | | – | – | – |
| A2 | 0.34 ± 0.00 | | 200 ± 0 | 12.5 ± 0.0 | 0.24 |
| T3 | 0.68 ± 0.13 | 0.59 ± 0.09 | 200 ± 218 | | 0.75 |
| G4 | 0.56 ± 0.00 | | 200 ± 0 | 4.6 ± 0.0 | 0.74 |
| A5 | 0.92 ± 0.09 | | | 6.7 ± 1.8 | 0.00 |
| G7 | 0.86 ± 0.04 | | 0.86 ± 56 | | 2.37 |
| T8 | 0.65 ± 0.03 | | 84 ± 73 | | 0.07 |
| A10 | 0.55 ± 0.03 | | 200 ± 1 | 8.7 ± 1.3 | 3.27 |
| T11 | 0.62 ± 0.17 | | | 4.7 ± 2.6 | 0.29 |
| G12 | 0.80 ± 0.09 | 0.89 ± 0.09 | 200 ± 436 | | 0.99 |

sequence of the molecule. There is a clear gradation in the order parameters, being low for the terminal three or four residues and high for the interior residues. Exchange contributions, which reflect μs time scale motions, are seen for residues A2, G4, A5, A10 and T11. Likewise, ps time scale motions are seen for residues A2, T3, G4, G7, T8, A10 and G12.

The half-site recognition sequence of GCN4 is ATGA and it was found that T3 plays a key role in the recognition process by GCN4 protein (9). It is interesting to note in our results that A2, T3 and G4 and their complementary units are quite dynamic in nature, as reflected by the low order parameter values. Mutagenesis studies (41) have shown that proteins mutated at conserved sites still bind to these DNA sequences, indicating the flexible nature of the DNA enabling it to make new contacts. Our results concur with this observation.

## CONCLUSIONS

In this paper we have described detailed NMR investigations on the high resolution structure and residue level dynamics of GCN4 cognate DNA. This system is unique in the sense that the protein has nearly similar affinities for two DNA sequences termed AP1 and ATF/CREB, respectively, which differ in only one base pair in the middle; the ATF/CREB sequence has an extra base pair in the center of the recognition sequence. The NMR structure determined with high precision indicated that the ATF/CREB DNA is inherently bent in the middle. This feature is probably what helps GCN4 protein recognize ATF/CREB as well as AP1 sequences without having to undergo large structural changes itself. The relaxation dynamics results have shown that the DNA has a fair amount of flexibility and this coincides with the recognition half-sites. These would introduce substantial variability in DNA deformation and thus would be of significance for recognition and regulation processes.

## SUPPLEMENTARY MATERIAL

Table S1 of chemical shifts of all assigned nuclei is available as Supplementary Material at NAR Online.

## ACKNOWLEDGEMENTS

We thank the National Facility for High Field NMR at the Tata Institute of Fundamental Research, supported by the Department of Science and Technology, India, for all the facilities. We also thank Dr A. Majumdar and Mr G. S. Lakshmikanth for the help provided during the initial stages of the work.

## REFERENCES

1. Steitz,T.A. (1990) Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophys.*, **23**, 205.

2. Paolella,D.N., Palmer,C.R. and Schepartz,A. (1994) DNA targets for certain bZIP proteins distinguished by an intrinsic bend. *Science*, **264**, 1130–1133.
3. Paquet,F., Gaudin,F. and Lancelot,G. (1996) Selectively $^{13}$C-enriched DNA: evidence from $^{13}$C1′ relaxation rate measurements of an internal dynamics sequence effect in the *lac* operator. *J. Biomol. NMR*, **8**, 252–260.
4. Spielmann,H.P. (1998) Dynamics in psoralen-damaged DNA by $^{1}$H-detected natural abundance $^{13}$C NMR spectroscopy. *Biochemistry*, **37**, 5426–5438.
5. Hinnebusch,A.G. (1984) Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc. Natl Acad. Sci. USA*, **81**, 6442–6446.
6. Thireos,G., Driscoll,P.M. and Greer,R. (1984) 5′ Untranslated sequences are required for the translational control of a yeast regulatory gene. *Proc. Natl Acad. Sci. USA*, **81**, 5096–5100.
7. Weiss,M.A., Ellenberger,T., Wobbe,C.R., Lee,J.P., Harrison,S.C. and Struhl,K. (1990) Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature*, **347**, 575–578.
8. Suckow,M., von Wilcken-Bergmann,B. and Muller-Hill,B. (1993) Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP and TAF-1 that are involved in specific DNA binding. *EMBO J.*, **12**, 1193–1200.
9. Keller,W., Konig,P. and Richmond,T.J. (1995) Crystal structure of a bZIP/GDNA complex at 2.2 Å: determinants of DNA specific recognition. *J. Mol. Biol.*, **254**, 657–667.
10. Ellenberger,T.E., Brandl,C.J., Struhl,K. and Harrison,S.C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: crystal structure of the protein-DNA complex. *Cell*, **71**, 1223–1237.
11. Mauffret,O., Hartmann,B., Convert,O., Lavery,R. and Fermandjian,S. (1992) The fine structure of two DNA dodecamers containing the cAMP responsive element sequence and its inverse: Nuclear Magnetic Resonance and molecular simulation studies. *J. Mol. Biol.*, **227**, 852–875.
12. Beaucage,S.L. and Iyer,R.P. (1992) Advances in the synthesis of oligonucleotides by the phosphoramidite approach. *Tetrahedron*, **48**, 2223–2311.
13. Sinha,N.D., Biernat,J., McMamis,J. and Koster,H. (1984) Polymer support oligonucleotide synthesis XVIII: use of β-cyanoethyl-*N*,*N*-dialkylamino-*N*-morpholino phosphoramidite of deoxynucleosides for the synthesis of DNA fragments simplifying deprotection and isolation of the final product. *Nucleic Acids Res.*, **12**, 4539–4557.
14. Jeener,J., Meier,B.H., Bachmann,P. and Ernst,R.R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.*, **71**, 4546–4554.
15. Braunschweiler,L. and Ernst,R.R. (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.*, **53**, 521–528.
16. Griesinger,C., Sorensen,O.W. and Ernst,R.R. (1985) Two-dimensional correlation of connected NMR transitions. *J. Am. Chem. Soc.*, **107**, 6394–6396.
17. Piantini,U., Sorensen,O.W. and Ernst,R.R. (1982) Multiple quantum filters for elucidating NMR coupling networks. *J. Am. Chem. Soc.*, **104**, 6800–6801.
18. Kay,L.E., Keifer,P. and Saarinen,T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.*, **114**, 10663–10665.
19. Kellogg,G.W. and Schweitzer,B.I. (1993) Two- and three-dimensional $^{31}$P-driven NMR procedures for complete assignment of backbone resonances in oligodeoxyribonucleotides. *J. Biomol. NMR*, **3**, 577–595.
20. Shaka,A.J., Lee,C.J. and Pines,A. (1988) Iterative schemes for bilinear operators; application to spin decoupling. *J. Magn. Reson.*, **77**, 274–293.
21. MSI (1997) *NMRchitect v.97.0 Online Manual* at http://www.msi.com/doc/insight970/nmrchitect/nmr970TOC.doc.html
22. Kay,L.E., Torchia,D.A. and Bax,A. (1989) Backbone dynamics of proteins as studied by $^{15}$N inverse detected heteronuclear NMR spectroscopy: application to Staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.
23. Skelton,N.J., Palmer,A.G., Akke,M., Kördek,J., Rance,M. and Chazin,W.J. (1993) Practical aspects of two-dimensional proton-detected $^{15}$N spin relaxation measurements. *J. Magn. Reson.*, **B102**, 253–264.
24. Farrow,N.A., Muhandiram,R., Singer,A.U., Pascal,S.M., Kay,C.M., Gish,G., Shoelson,S.E., Pawson,T., Forman-Kay,J.D. and Kay,L.E. (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by $^{15}$N NMR relaxation. *Biochemistry*, **33**, 5984–6003.
25. Cavanagh,J., Palmer,A.G., Wright,P.E. and Rance,M. (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear relay spectroscopy. *J. Magn. Reson.*, **91**, 429–436.
26. Palmer,A.G., Cavanagh,J., Wright,P.E. and Rance,M. (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy *J. Magn. Reson.*, **93**, 151–170.
27. Marion,D., Ikura,M., Tshudin,R. and Bax,A. (1989) Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.*, **85**, 393–399.
28. Peng,J.W. and Wagner,G. (1994) Investigation of protein motions via relaxation measurements. *Methods Enzymol.*, **239**, 563–596.
29. Abragam,A. (1968) *The Principles of Nuclear Magnetism*. Clarendon Press, Oxford, UK.
30. Gaudin,F., Paquet,F., Chanteloup,L., Beau,J.M., Nguyen,T.T. and Lancelot,G. (1995) Selectively $^{13}$C-enriched DNA: dynamics of the C1′-H1′ vector in d(GCGAAATTTGCG)$_2$. *J. Biomol. NMR*, **5**, 49–58.
31. Lipari,G. and Szabo,A. (1982) Model-free approach to the interpretation of Nuclear Magnetic Resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.*, **104**, 4546–4559.
32. Lipari,G. and Szabo,A. (1982) Model-free approach to the interpretation of Nuclear Magnetic Resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.*, **104**, 4559–4570.
33. Clore,G.M., Szabo,A., Bax,A., Kay,L.E., Driscoll,P.C. and Gronenborn,A.M. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.*, **112**, 4989–4991.
34. Mandel,A.M., Akke,M. and Palmer,A.G. (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.*, **246**, 144–163.
35. Wuthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York, NY.
36. Hosur,R.V., Govil,G. and Miles,H.T. (1988) Application of two-dimensional NMR spectroscopy in the determination of solution conformation of nucleic-acids. *Magn. Reson. Chem.*, **26**, 927–944.
37. Kellogg,G.W., Szewczak,A.A. and Moore,P.B. (1992) Two-dimensional hetero-TOCSY-NOESY. Correlation of $^{31}$P resonances with anomeric and aromatic $^{1}$H resonances in RNA. *J. Am. Chem. Soc.*, **114**, 2727–2728.
38. Varani,G. and Tinoco,I.,Jr (1991) RNA structure and NMR spectroscopy. *Q. Rev. Biophys.*, **24**, 479–532.
39. Lavery,R. and Sklenar,H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.*, **6**, 63–91.
40. Schurr,J.M., Babcock,H.P. and Fujimoto,B.S. (1994) A test of the model-free formulas: effects of anisotropic rotational diffusion and dimerization. *J. Magn. Reson.*, **B105**, 211–224.
41. Madan,A. (1994) PhD thesis, Mumbai University, India.