J Comput Aided Mol Des (2015) 29:283–295
DOI 10.1007/s10822-014-9828-x

Role of indirect readout mechanism in TATA box binding protein–DNA interaction

Manas Mondal · Devapriya Choudhury · Jaydeb Chakrabarti · Dhananjay Bhattacharyya

Received: 26 September 2014 / Accepted: 18 December 2014 / Published online: 10 January 2015

Abstract Gene expression generally initiates from recognition of TATA-box binding protein (TBP) to the minor groove of DNA of TATA box sequence where the DNA structure is significantly different from B-DNA. We have carried out molecular dynamics simulation studies of TBP–DNA system to understand how the DNA structure alters for efficient binding. We observed rigid nature of the protein while the DNA of TATA box sequence has an inherent flexibility in terms of bending and minor groove widening. The bending analysis of the free DNA and the TBP bound DNA systems indicate presence of some similar structures. Principal coordinate ordination analysis also indicates some structural features of the protein bound and free DNA are similar. Thus we suggest that the DNA of TATA box sequence regularly oscillates between several alternate structures and the one suitable for TBP binding is induced further by the protein for proper complex formation.

Keywords Molecular dynamics · Protein–DNA recognition · DNA bending analysis · Principal coordinate ordination · Minor groove width

Introduction

Molecular recognition in protein–DNA interaction plays a vital role in molecular biology as this process drives gene expression and various other biochemical processes [1–4]. In several such protein–DNA complexes the 'lock-and-key' mechanism [5] is utilized by nature where alpha helices of the gene regulatory proteins bind in the major groove of DNA by forming specific hydrogen bonds with the bases. This mechanism can also be termed as 'direct readout' [6, 7] as the proteins can make specific recognition by formation of hydrogen bonds with the bases in their native like form. In several situations, on the other hand, the protein or DNA needs to deform significantly for strong binding, which can be termed as 'induced-fit' binding [8]. All the semi-flexible biopolymers have large conformational subspace. Thus, other than the native and lowest energy conformation, thermally accessible conformational sub-states may also play important roles in molecular recognition [9, 10]. Another model, known as 'conformational selection model' [9, 11, 12], has been proposed where the weakly populated, higher energy conformational states are responsible for recognizing and binding to partners with subsequent population shift toward these conformers [13]. This model, with some experimental evidence [12], challenges the assumption that conformational differences between free and ligand bound states of a molecule automatically implicate an induced fit type mechanism of molecular recognition. This model assumes sampling of different conformations by the macromolecules as part of their inherent flexibility and

Electronic supplementary material The online version of this article (doi:10.1007/s10822-014-9828-x) contains supplementary material, which is available to authorized users.

M. Mondal · D. Bhattacharyya (✉)
Computational Science Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India
e-mail: dhananjay.bhattacharyya@saha.ac.in

D. Choudhury
School of Biotechnology, Jawaharlal Nehru University, New Mehrauli Road, New Delhi 110067, India

J. Chakrabarti (✉)
S.N. Bose National Center for Basic Sciences, Sector III, Salt Lake, Kolkata 700098, India
e-mail: jaydeb@bose.res.in

adaptation of one of them in the complex state. Also it is well understood that the distinction between induced fit and the conformational selection models is not absolute; indeed, an increasing number of cases show that conformational selection is often followed by conformational adjustment [14–16]. Extended conformational selection model [16] describe the general scenario, where both selection and adjustment-type steps follow each other. However the available time-average structures of different protein, DNA and their complexes obtained from crystallography or NMR spectroscopy are not sufficient to distinguish different modes of molecular recognition.

There are some sequence specific DNA binding proteins such as, TATA-box binding protein (TBP) or the histone octamer that can produce large scale deformation to DNA due to binding [17–23]. The biological role of these high energy 'activated' states is related to the improved accessibility of DNA base pairs during the course of DNA processing i.e. methylation, recombination, transcription etc. [24, 25]. These high energy activated forms, which are unfavourable in free and straight DNA, are presumably stabilized by the binding of the protein. The conserved C terminal region of TBP, which is important for transcription initiation throughout the biological kingdom, binds with the promoter region of TATA box and makes the transcription initiation complex [26–31]. Crystal structures of several TBP–DNA complexes show that TBP binds to the minor groove of the TATA box, unwinds the TATA box DNA and bends the DNA towards the major groove by an angle about 80° [26, 32–34]. Wu et al. [35, 36] reported from time resolved fluorescence emission in conjunction with fluorescence resonance energy transfer experiments that the geometries of the TBP-bound variants of TATA sequences in solution vary significantly and differ from their corresponding co-crystal structure. These solution conformations, based on a two kink bending model, are consistent with DNA bending angles ranging from 30° to 75°. Considering the fact that DNA double helices of most sequences can wrap around histone octamer to form nucleosome core particle, one can argue that DNA flexibility does not depend on base composition or sequence, but recent studies indicate it might depend on base sequence of DNA as well [37–40]. A correlation was observed between the bending of DNA and relative transcription efficiency in in vivo and in vitro analysis. Hawley and co-workers had observed that the induced degree of bending of TATA box mutants correlates with the stability of TBP–DNA complexes [41]. It was further indicated that depending upon the minor groove width, activity of the TATA element is also different for different TATA variants [42]. It was also reported that the width of the DNA minor groove varies with sequence and can be a major determinant of DNA shape recognition by proteins [43]. We had shown earlier, through an analysis of crystal structures of several TATA box–TBP complexes, that the deformed DNA attains thermodynamic stability upon protein binding [44]. The specific ways to facilitate DNA distortions are different for different proteins, but one feature always remains the same, the conformational preferences of the duplex itself. The above studies could not detect role of natural motions of the protein or DNA, which can be addressed by molecular dynamics (MD) simulations of the systems in physiological condition.

In this work, we have used MD simulations to study the role of sequence dependent flexibility of TATA box DNA sequence and role of indirect readout mechanism in TBP–DNA interaction, where DNA is deformed significantly after protein binding [21]. We have simulated the DNA and protein molecules both in free and bound states also to see any conformational preference of the DNA and protein. Here we report the results of these simulations in terms of flexibility of the DNA as well as the TBP protein and discuss the role of conformational selection as well as induced fit mode in TBP–TATA box recognition. Our study indicates that in case of TBP–DNA interaction conformational adjustment of DNA takes place i.e. most suitable and partially bent conformation of DNA are accepted for binding by increasing bending of the same.

Methodology

Molecular dynamics simulation

The coordinates of the TBP and TBP bound DNA have been taken from Protein Data Bank (PDB) [45] having PDB ID 1CDW [34]. The coordinates of regular B-DNA having the same sequence of 1CDW and two randomised sequences with same base composition were also generated using B-DNA fibre model [46]. The protein in 1CDW.pdb is C-terminal domain but for simplicity we numbered the 155th residue as the 1st. Initial set up and minimizations of the systems have been done using AMBER 14 [47] suite of programmes. Each of the systems were solvated in an orthorhombic water box containing TIP3P water molecules and required numbers of sodium ions to maintain the electro-neutrality. The systems were solvated in such a way that there were at least 15 Å thick layers of water molecules around the solute in all directions. These systems were then energy minimized for 20,000 steps using a combination of steepest descent and conjugate gradient algorithms and applying periodic boundary conditions. Lennard-Jones and short-range electrostatic interactions were truncated at 10 Å. Long-range electrostatic energy was calculated using the Particle Mesh Ewald summation method with 1 Å grid spacing and 10^{-6} convergence

criterion. We have also carried out analysis of the TATA box binding protein (TBP) alone in similar fashion. Details of the systems studied are shown in Table 1. The MD simulations were carried out using NAMD [48, 49] with AMBER parm94 [50] and ff14SB [47, 51] force fields. Initially for each of the minimized systems, heating to 300 K was carried out slowly during 30 ps with 1 fs time step. After that we have continued the simulations for the six systems up to a time as mentioned in Table 1 at constant temperature (300 K) and pressure (1 bar) using the Langevin-Piston algorithm and periodic boundary conditions. Translational and rotational movements of the centre of mass were removed at an interval of 5 ps. SHAKE constraints were applied to all bonds involving hydrogen atoms. Conformations obtained after every 1 ps were saved from the trajectories for further analysis.

Structure analysis

Root mean square displacement (RMSD) and root mean square fluctuation (RMSF) analyses of the trajectories are performed using CHARMM [52] and all the structural parameters of the DNA were calculated by NUPARM [53, 54]. The hydrogen bonding information between protein and DNA during MD runs of TBP–DNA complex are obtained by a modified version of pyrHBfind [55] using H-bond distance cut off 3.0 Å and angle greater than 150°. Detection of amino acid residues in alpha-helical or beta-sheet conformations were done using CHARMM that follows DSSP [56] algorithm for assignment of secondary structure.

We calculated bending angle as the angle between local helix axes ($\overline{Z}^*$) of two base pair steps using NUPARM [53]. The local helix axis ($\overline{Z}^*$) is defined by

$$\overline{Z}^* = \frac{(\bar{x}_1 - \bar{x}_2) \times (\bar{y}_1 - \bar{y}_2)}{|(\bar{x}_1 - \bar{x}_2)||(\bar{y}_1 - \bar{y}_2)|}$$

where $\bar{y}_i$ (long axes) are obtained as the vectors connecting C8 and C6 atoms of purines and pyrimidines, respectively, of the base pair, $\bar{z}_i$ are base pair normals and $\bar{x}_i$ are obtained by $\bar{x}_i = \bar{y}_i \times \bar{z}_i$. Angles between the last but one local helix axes are used to define bending angles of a double helical fragment and this can give a good estimate of bending or curvatures in oligo or polynucleotides [57, 58]. We, however, ignored the first two and last two base pairs for this analysis, as these often show fraying effect. It was observed that the TBP binds near the central region of the DNA double helix fragment in 1CDW and modulates the binding region while few base pairs of the two termini remain in B-DNA conformation. We have fitted two average helix axes to these B-DNA like conformations considering C1' atoms of 2nd to 5th base pairs as the first helix and 12th to 15th base pairs as the last helix. We have reduced the terminal fraying effect by not considering the first and the 16th base pairs. The helix axis determination algorithm follows that adopted by Dickerson in FREEHELIX program [59] and is also similar to that adopted in Curves [60]. We have finally calculated angle between these vectors as the second estimate of bending angle.

Principal component analysis (PCA)

Conformational space of a molecule is in reality a multi-dimensional hyper-surface but usually a few descriptors can simplify the landscape. Therefore reconstruction of the conformational landscape in terms of the so-called collective coordinates, which are also known as principal components, is more meaningful. The conformational

Table 1 Systems taken for study by molecular dynamics simulation

System taken for analysis	Initial structure	DNA sequence	Presence of TBP	MD Simulation run time	Average RMSD with respect to initial structure in Å
1 TBP–DNA complex	As in 1CDW	d(CTGCTATAAAAGGCTG)-3′	Yes	100 ns (parm 94)	2.02 (DNA)
					2.32 (TBP)
				100 ns (ff14SB)	2.42 (DNA)
					2.05 (TBP)
2 Deformed/Bent DNA	As in 1CWD	d(CTGCTATAAAAGGCTG)-3′	No	100 ns (parm94)	5.75
				150 ns (ff14SB)	6.23
3 Regular DNA	Fiber B-DNA	d(CTGCTATAAAAGGCTG)-3′	No	100 ns (parm94)	4.44
				150 ns (ff14SB)	2.69
4 Randomized sequence-1 DNA	Fiber B-DNA	d(GACGTCAGTTCGAATA)-3′	No	100 ns (parm94)	3.74
				50 ns (ff14SB)	2.50
5 Randomized sequence-2 DNA	Fiber B-DNA	d(CTGTAGATAGCTAGAC)-3′	No	100 ns (parm94)	4.40
				50 ns (ff14SB)	2.63
6 Only TBP without DNA	As in 1CDW	–	Yes	100 ns (ff14SB)	2.53

space can be approximated by those major principal components that have the largest eigenvalues [61]. We have studied these Principal Component motions using Bio3D suite of program [10, 62] from the MD simulation trajectories using R package. For simplicity we have considered only backbone P atoms of DNA and only backbone C^{α} atoms of protein trajectories.

Principal coordinate ordination (PCO)

PCO is another way to study the conformational space. Conceptually this method is very similar to Multidimensional Scaling methods as used in statistics or the method of Distance Geometry used in the building and refinement of molecular structures. The input to this method is an n × n symmetric matrix of distances between n points embedded in a p-dimensional space (R^p). The original positions of the points that give rise to the distance matrix are not known a priori but are to be determined by use of the method. In this sense, this method is akin to internal coordinate to external coordinate conversion, with the set of inter-point distances being the internal coordinates. Since distances (like all types of internal coordinate) do not specify the absolute position and orientation of the cloud of points, these have to be supplied externally. Thus in R^3, an n × n matrix of Euclidean distances will only provide 3n − 6 external coordinates for n points. In PCO, the output coordinate system is chosen in such a way that the origin is at the centroid of the points and orthogonal axes are chosen in such a way that the variance along each direction is maximized. The distance function, used to calculate the inter-point distances, should ideally be a Euclidean metric.

Now for a given a symmetric matrix of distances, one can carry out PCO using the following protocol [63, 64]:

1. Construct an n × n symmetric matrix A such that:

$$\begin{aligned} A_{ij} &= -1/2\left(d_{ij}^2\right) && \text{if } i \neq j \\ &= 0 && \text{if } i = j \end{aligned}$$

2. Center the A matrix using the formula:

$$A_{ij}^{*} = A_{ij} - \langle A_{ij}\rangle_i - \langle A_{ij}\rangle_j - 2\langle A_{ij}\rangle_{ij}$$

where $\langle \ldots \rangle_k$ refers to the mean over specified indices $k = \{i, j, ij\}$

3. Solve for

$$A^{*} = B\Lambda B^{T} \quad \text{or,} \quad B^{T}A^{*}B = \Lambda$$

where B is the matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues.

4. Scale each eigenvector (column vector of B) by the square root of the corresponding eigenvalue ($\sqrt{\lambda_i}$) such that $B^{T}B = \Lambda$. The rows of the scaled matrix B give the required ordination. Assuming that the eigenvalues are sorted in terms of decreasing magnitude, the first column of the matrix represents the axis of maximal variance; the second column represents the orthogonal axis having the second largest variance and so forth. When a particular eigenvalue approaches 0, the corresponding column can be neglected.

Alignment free distance measure, Distance RMSD (dRMSD), is used in this PCO method for distance evaluation. The dRMSD value is calculated using the formula:

$$\text{dRMSD} = \sqrt{\frac{2}{n(n-1)}\sum_{i=2}^{n}\sum_{j=1}^{n-1}(r_{ij}^{A} - r_{ij}^{B})^2}$$

where r_{ij}^{A} and r_{ij}^{B} denotes the distance between atom pairs i and j in structures A and B respectively and n being the total number of atom-pairs in the structures being compared. This dRMSD is known to satisfy the triangle inequality and thus qualifies as a true metric. A disadvantage with this measure is that atom-pairs that are farthest away from each other contribute most to the dRMSD. However since the molecular potential energy is distance dependent, atom pairs that contribute most to the dRMSD contribute the least to the molecular potential energy. Hence a large difference in dRMSD between two conformations need not translate into a correspondingly large potential energy difference between them. The use of dRMSD and PCO in the analysis of conformational energy landscapes of biomolecules and MD trajectory analysis was reported by several groups [9, 65].

Results and discussion

We have carried out molecular dynamics simulations for six systems, as indicated in Table 1—the TATA box-TBP complex (system 1), the free TATA box DNA sequence from two initial structures (system 2 and system 3), two random sequences (system 4 and system 5) and C-terminal domain of the free TBP protein (system 6). The RMSD values of DNA or protein in most of the systems sharply increase to their corresponding average values within 5 ns (Figure SI, Figure S2). However, the RMSD values of the free DNA systems continue to change significantly even after 5 ns. There can be several reasons for this—(1) the initial structure of system 2 is highly deformed and its relaxation needs longer time, (2) the system 3 also may adopt a significantly altered structure by overcoming some energy barriers, (3) the DNA is of length close to two full turns and it may be sufficiently deformable. Hence we have analysed the snapshots from the simulation trajectories after 30 ns for further studies.

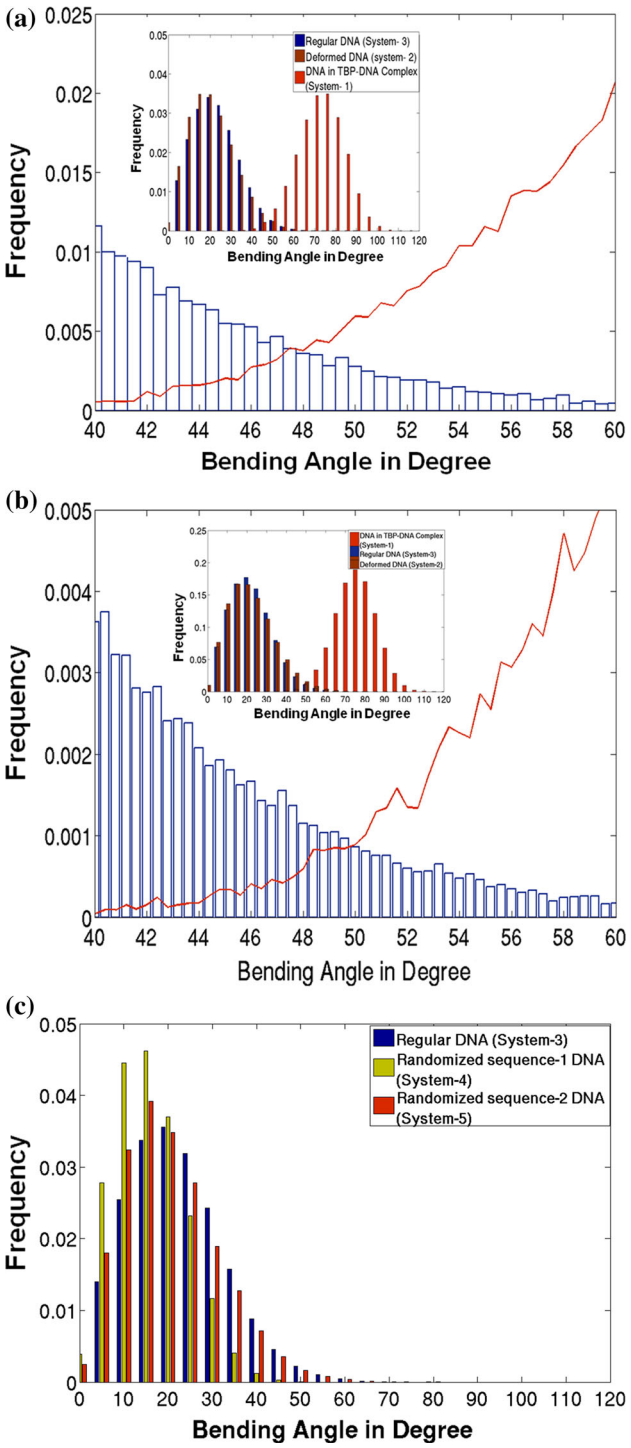

Fig. 1 Distribution of bending angle for free DNA of TATA-box ▶ sequence (system 3) (as *blue bars*) and in the TBP-TATA complex (*red line*) in their overlapping region. Distribution of bending angles over the whole range in the three TATA systems are shown in *inset* [*red* in Complex (system 1), *blue* regular DNA (system 3), *brown* deformed DNA (system 2)]. **a** The angles were calculated from local helix axes of the second and last but one dinucleotide steps, **b** the angles were calculated from average helix axes fitted to the two ends, leaving out the first and last base pairs. **c** Comparison of bending angle distributions for DNA of TATA sequence (system 3; *blue bars*) and for two random sequences (systems 4 and 5) are shown here. Bending angles were calculated from average helix axes fitted to the two ends, leaving out the first and last base pairs

Bending and flexibility of DNA

We attempted to compare the global bending of the DNA double helix upon complex formation and computed the end-to-end angle using the local helix axes of the terminal base pairs. For rigid and straight DNA of any form, such as B-DNA or even A-DNA with large roll angles for all steps, this angle becomes zero irrespective of the length of the double helix. According to our method value of this angle is 75° as in the crystal structure of TBP–DNA complex (1CDW). The distributions of this bending angle for the studied systems are shown in Fig. 1 and supplementary Figure S4 for parm94 force field. This indicates that the range of this bending angle in the TBP-bound state (system 1) varies between 30° and 110° with 76° being the most likely value (Fig. 1a). As DNA bending is crucial for TATA-box–TBP binding, we have calculated DNA bending angle from an alternative approach also. Here we fit a global helix axis through 2nd to 5th base pairs and another average global helix axis through the 12th to 15th base pairs following an algorithm adopted by FREEHELIX [59] which is also similar to that of CURVES [60] software. The distribution of angles between average helix axes (Fig. 1b) appear indistinguishable from that calculated from local helix axes. Thus the DNA sequence seems to have significant flexibility and shows different amounts of bending even after binding to the protein, as found from solution studies [35, 36] earlier. We have also analysed the curvature of the protein free DNA (system 2 and 3) and these show the bending angles of the free DNA from an initial bent conformation (system 2) and from an initial canonical B-DNA structure (system 3) varies within 0° and 70°.

In order to understand whether this bending flexibility of the DNA is related to its sequence, we have carried out MD simulations of two B-DNA molecules with randomized sequence (system 4 and 5). These can be considered as the mixed genomic sequences, which may not have the signature of TATA box. The bending angle analyses (Fig. 1c, Figure S4) for these mixed sequences indicate these are quite small indicating their rigid and straight form. It may also be mentioned that one of the random sequences (system 5) is seen to have somewhat greater (although minor) flexibility as compared to the other. This may indicate the possibility of TBP binding to some of the TATA-less promoter sequences also, as seen from genomic sequence analysis [66, 67]. As shown in Fig. 1a, b, simulation starting from canonical B-DNA conformation (system 3) gives a significant population of structures with

bending angle value above 50°, coinciding with bending angles of the complex (system 1). In the inset of the Fig. 1a, b we showed bending angle distributions over the whole range for the three systems having same DNA sequence. This indicates significant overlap of bending angle, especially for system 3 and system 1. The relevant regions, drawn as the main figure indicates that there are significant population of structures of the regular TATA sequence B-DNA (system 3) above 40°. These bending angles are also quite probable in the TBP bound form (as shown by line). This global bending feature is maintained in both the simulations using parm94 and ff14SB force field of AMBER (Fig. 1 and Figure S4). However as ff14SB force field is more accepted in recent days, therefore for further analysis we have considered the trajectories corresponding to ff14SB force field only. The structure and bending pattern in this overlapping region for regular and TBP bound DNA is shown in Fig. 2. Our study indicates that the deformed bent DNA can get back to its regular conformation without TBP due to its inherent flexibility and some deformed conformations of regular B-DNA might be adjusted in TBP bound state. Furthermore, significant overlap between the bending histograms of regular and TBP bound TATA box DNA sequences may indicate a high inherent bending deformability of this sequence that can lead to deformed bound like conformation.

DNA trajectory comparison with principal component ordination (PCO)

PCA analysis of the backbone P atoms of the DNA trajectories indicates that regular TATA sequence DNA (system 3) can access a wide conformational space as compared to others, as first two Principal components (PCs) are not sufficient to captured most of the conformational variation in this case (Figure S5). Therefore to identify any conformational overlap in conformational space among the regular TATA sequence DNA conformations in unbound state (system 2 and 3) and TBP bound state (system 1) we have compared these three trajectories using PCO analysis [10]. The three trajectories were concatenated in the order: (a) Complex (system 1) (b) Deformed (system 2) (c) Regular (system 1). In case of system 1 from the last 80 ns trajectory we have considered the snapshots after every 20 ps and for system 2 and system 3 we have considered the snapshots after every 30 ps from last 120 ns trajectories. Calculations of dRMS were done in three different ways depending on the atoms chosen for RMS calculation. This resulted in three different sets of output data. The datasets were:

1. *Phosphorus set:* RMS calculations were done only on the coordinates of the phosphorus atom.

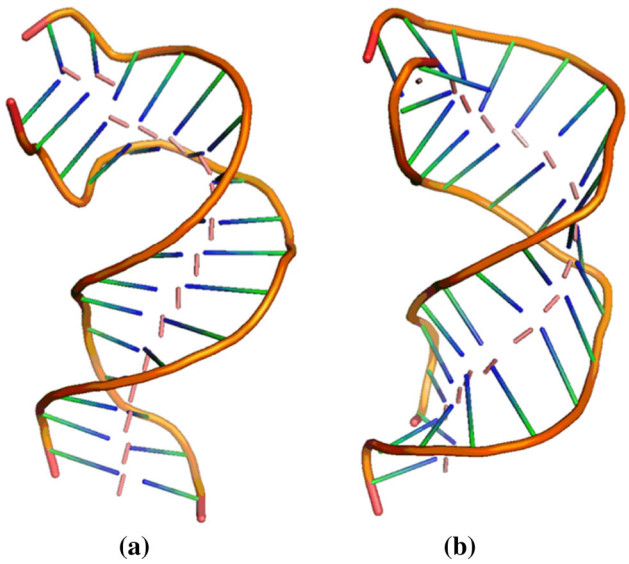

Fig. 2 Conformation of **a** regular DNA (system 3) and **b** DNA in Complex (system 1) having similar bend angle along with their local helix axes corresponding to each dinucleotide step

2. *Base set:* RMS calculations were done using the ring nitrogen atoms of the bases (N1 and N3 for pyrimidine, N1, N3, N7 and N9 for purines).
3. *All atom set:* All non-hydrogen atoms were used for RMS calculations. In this case only every alternate structure in the concatenated trajectories was used for RMS and subsequent PCO calculation.

The plot of the dRMSD matrix obtained from this analysis is shown in Fig. 3. It is clear from the figure that regular and deformed DNA trajectories are not distinctly different from each other as their dRMSD values are less, but with respect to the DNA in complex state the dRMSD values are comparably large in all the three sets. However Fig. 3 indicates that there are few structures in the complex state having lower dRMSD values with respect to regular and deformed DNA (system 3 and 2).

The PCO analysis of all the three data sets indicates that the first two components accounts for approximately 80 % of the total conformational variation in the three trajectories and the first component alone accounting for about 75 % of the conformational variation. The contribution of 2nd and 3rd is around 4 %. However few higher components are also required, in order to capture overall 90 % explained variance. The plot of the explained variances with the number of principal components for set (c) is shown in Fig. 4a. Corresponding PCO score plots in the space of two of the first three principal components obtained from the combined distance matrix are shown in Fig. 4b–d. It is clear from the plot that in terms of PC1 versus PC2 and PC1 versus PC3, the conformational spaces of the deformed and regular DNAs (system 2 and 3) are

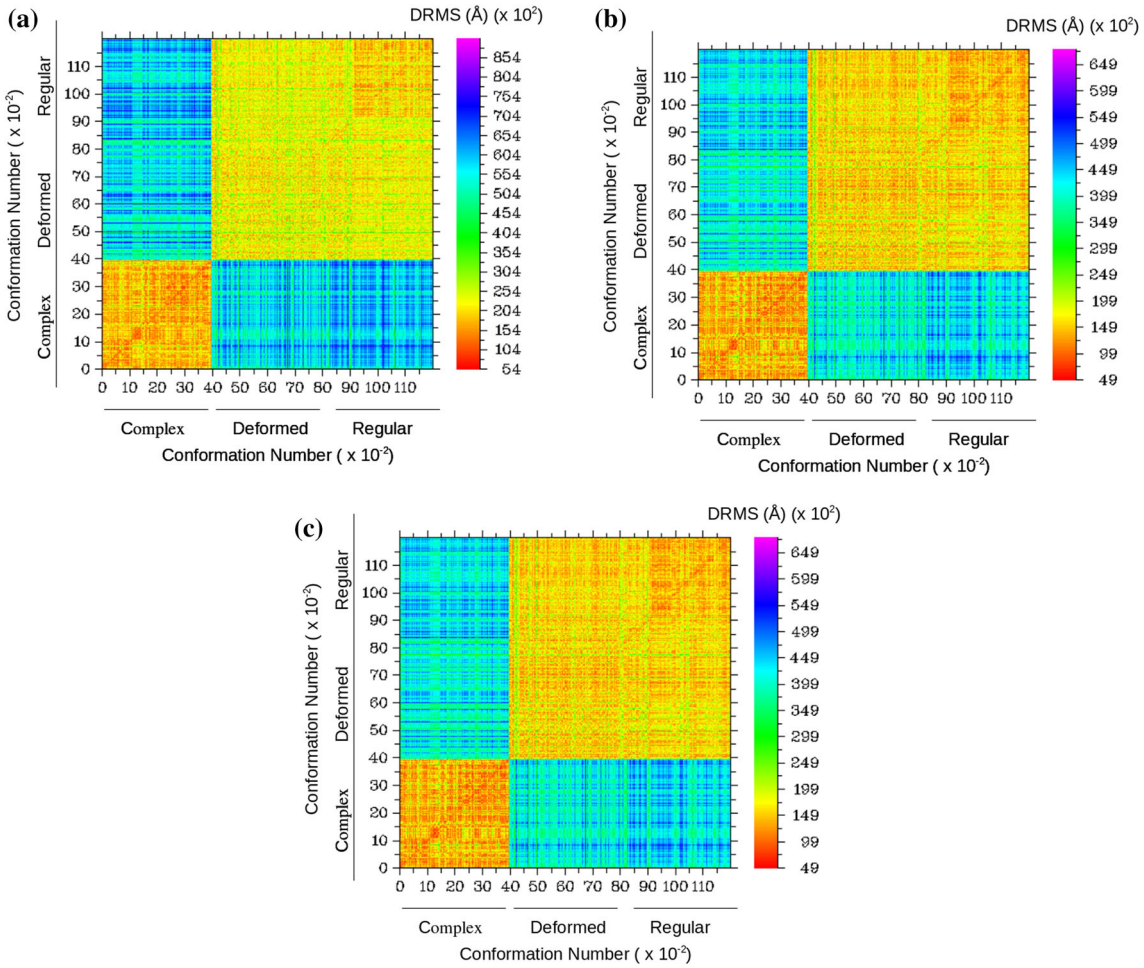

Fig. 3 Principal Coordinate Ordinate (PCO) analyzed DNA trajectories, in the form of dRMS plot for the three systems **a** phosphorus set, **b** base set and **c** all atom set

essentially identical and both of which are distinct from the conformational space traversed by the DNA molecule complexed with the protein. However the PC2 versus PC3 plot of the free DNA systems (systems 2 and 3) show strong overlap with the same for the TBP bound TATA box (system 1) in the conformational space. Therefore the first three principal components indicate that the observed overlap of the conformational spaces in these three dimensions accounts for a significant amount of conformational similarities between the three DNA trajectories. This indicates that there are few principal motions of the free TATA box sequence DNA, which are very similar to that of the TBP-TATA box complex. It may be indicated here that RMSD values for the snapshots of system 3 with respect to system 1 reduces from an initial value of 9–6.5 Å, indicating its trend of transition towards the deformed and bound conformation (Figure S3). This also might indicate a tendency of the regular DNA structure to alter towards the TBP bound form.

Structural features of DNA during MD simulations

Global bending of DNA structure is associated with the base pair step geometry and deformability. Therefore, for TATA box containing DNA sequences we have studied the base pair step deformability in terms of roll, twist, slide, etc. for each base pair step in TBP bound and unbound states (system 1, system 2, system 3). In the binding region (6th–11th step) of system 1 unusually large roll, positive slide and very small twist values are observed for most of the base pair steps as reported in our previous study as well [44, 68]. Due to partial intercalation of a pair of phenylalanine residues of the TBP protein, the roll values for the first and last base pair steps of binding region (i.e. step-5 and step-11) are significantly high. However in these two steps the variation of tilt and slide values are similar to the unbound DNA (systems 2 and 3). In system 2 and system 3, positive roll associated with lower twist values are observed for TA/TA steps in the TBP binding region of

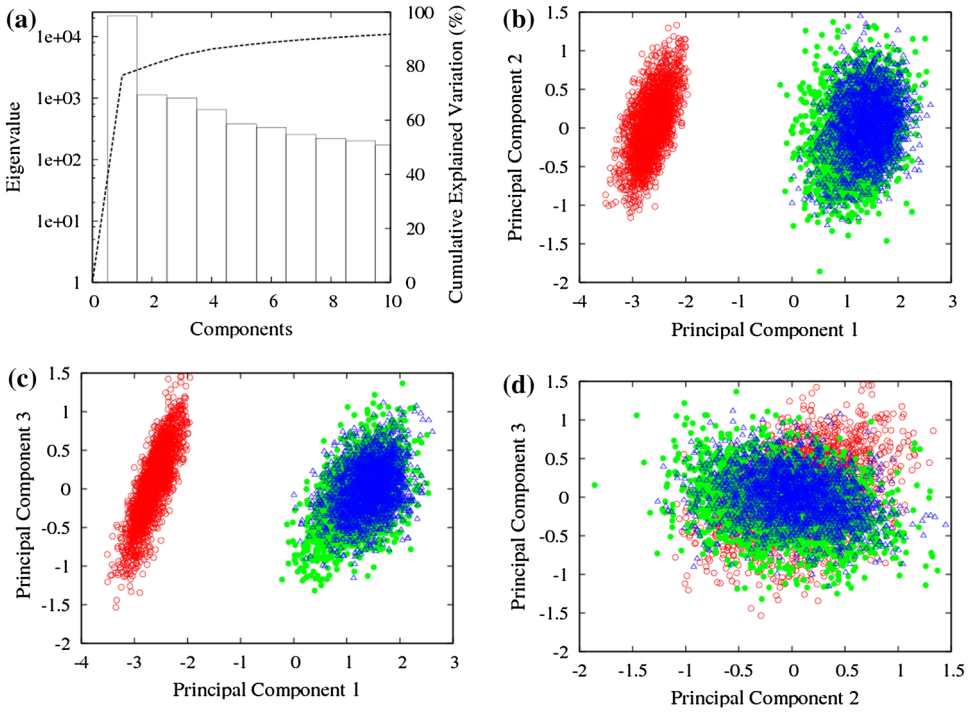

Fig. 4 **a** Plot of the eigenvalues after PCO. The *dashed line* plots the cumulative explained variation using axis on the *right side* of the figure. **b–d** PCO score plot in the space of the principal components. The *symbols* refer to conformations taken from the Complex (system 1) (*red*), deformed (system 2) (*green*) and regular (system 3) (*blue*) DNA trajectories

TATA box sequence. Flexibility of roll is also higher for these steps. The variations of these base pair parameters and torsion angles with their standard deviations for different systems are given as supplementary materials (Table S1-S9). Structural rigidity of the TBP bound TATA box DNA from the base-pair level, however, is not observed within this time scale. Variations of torsion angle values for all the systems indicate these become more restricted upon TBP binding. In system 1 (TBP bound DNA) the δ torsion angles in the binding region are in the 90° range, corresponding to C3′-endo sugar puckers. Therefore the DNA in the complex (system 1) adopts unusual conformation in the binding region and this leads to the overall bending deformation in DNA.

TBP–DNA interaction and variation of groove widths

Bending deformability of the duplex DNA assists the interactions of TBP at the binding interface along the shallow and wide minor groove of DNA. However the variation in disposition of the functional groups are significantly less in the minor groove [69]. It was observed that in TBP–DNA complex maximum number of H-bonds is formed between phosphate groups of DNA and a few amino acid side chains. Few amino acid residues also make hydrogen bonds and large van der Waals contacts with DNA bases. We have studied the stability of these hydrogen bonds in system 1 (Table S10). This indicates all the hydrogen bonds are not static. Hydrogen bonding partners of few of them are seen to change quite often. Hydrogen bonds involving DNA base atoms where Asn-9 forms a pair of hydrogen bonds with the O2 atoms of Thy-24 and Thy-25 are very stable. The H-bonds between Lys-37, Lys-60, Arg-136, Arg-141 with strand 1 of DNA and between Arg-38, Arg-45, Lys-151 and 2nd strand of DNA are quite stable and are similar to electrovalent bonds. Fluctuations of these hydrogen bonded TBP residues in presence and absence of DNA (Figure S6) indicate they are quite rigid compared to other residues in DNA bound state. Fluctuations of corresponding hydrogen bonded phosphate groups of DNA bases are also less in TBP bound state (Figure S7).

As some of the positively charged residues binding to phosphate groups across minor groove of DNA remain static and TBP binding causes widening of minor groove, we have analysed variations of minor groove width of the DNA systems by calculating Euclidian distances between phosphate atoms. All these relevant distances are schematically shown in a cylindrical coordinate's plot of the phosphates (Fig. 5). The distances between P-8 … P-29, P-9 … P-28, P-10 … P-27, P-11 … P-26 and P-12 … P-25 along the minor groove of the binding region of DNA in the complex (system 1) are found to be large and their standard deviations are also significantly less as compared to the other free DNA systems. The corresponding distances along the major groove shows a decrease but their standard deviations remain same with those of the others.

However, few distances between P…P atoms along the minor groove of binding region are highly flexible in the unbound states (systems 3) and they have a tendency to adopt higher values. Analyses of frequency histogram of these minor groove distances in the free DNA system (system 3) also indicate that these have wide variability and their tail regions coincide with the P–P distances of the TBP-TATA box complex systems (Fig. 6). Similar overlap of histograms was also seen in the bending angle analysis.

Structural stabilities of protein

We have also analysed the trajectory of TBP in presence and absence of DNA to see any conformational preference or alteration of the TBP intended for binding. As shown in Table 1, the average RMSD value of the free protein starting from DNA bound conformation (system 6) is only 0.5 Å higher than the DNA bound state (system 1). Furthermore no major conformational transition and secondary structure alteration of TBP is observed in either case. The residues near the ends of the β-strands sometimes show a transition to unstructured turn or coil conformation (Figure S8). This may be due to loss of inter-residue hydrogen bonds or alteration of ϕ/ψ torsion angles of the protein backbone. The amino acid residues in this region also make hydrogen bonds or van der Waals contacts with DNA backbone in the DNA bound state as mentioned before and they have smaller RMSF values in DNA bound state. Crystal data of several TBP–DNA complexes also indicates that the overall secondary structure of the DNA binding region of TBP is same as the free TBP (1VOK [27]) and their RMSD values are also small (Figure S8). PCA analysis of the C^{α} atoms of TBP also shows that, around 90 % of the total variance of the atomic fluctuations was captured along the first three principal components (PCs) both for free and DNA bound state (Figure S9). However for TBP in DNA bound state, contribution of 1st PC in total variance is higher. The side chain torsion angle (χ_1, χ_2, etc.) becomes more restricted in the DNA bound state for few amino acid residues, namely Arg-38, Phe-39, Arg-45, Lys-60, Gln-98, Asn-99, Lys-151, Val-152 and Lys-158, all of which are in the unstructured loop regions and within 5 Å of DNA. Some of these residues are found to form strong H-bonds with phosphate groups across the minor groove of the TATA box DNA. Therefore though some of the key protein residues get rigidity in DNA bound state, the overall conformation and conformational variability of TBP in configurational space remains same in presence and absence of DNA.

As indicated earlier that positively charged amino acid residues in the binding side interact with the phosphate

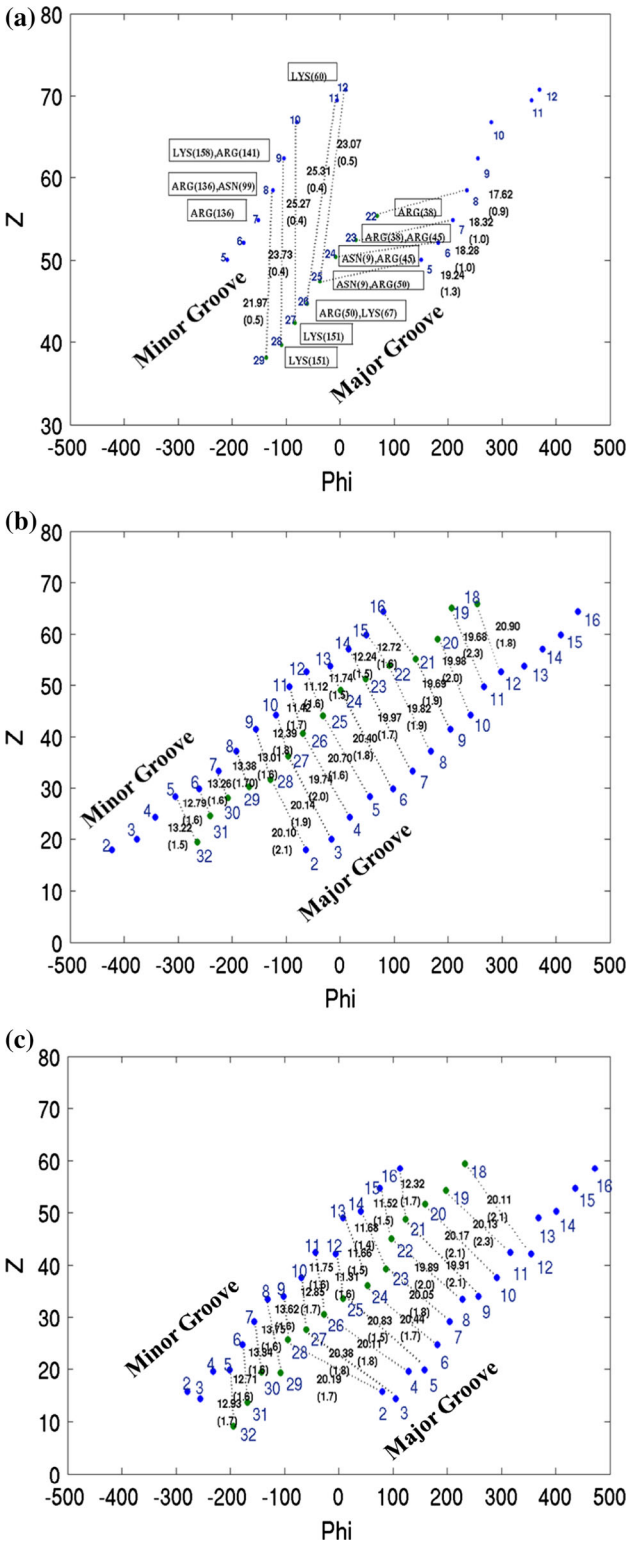

Fig. 5 Groove widths, as measured by inter-strand P–P distances along major and minor grooves of **a** binding region of TBP bound DNA (system 1), **b** regular DNA (system 3), **c** deformed DNA (system 2)

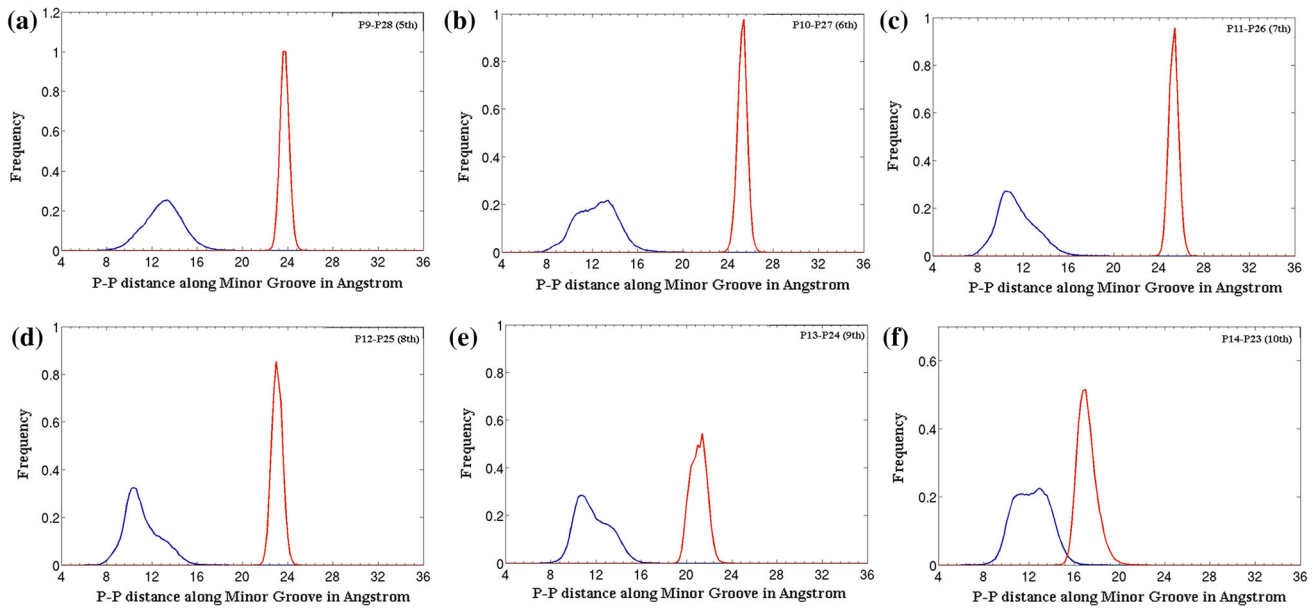

Fig. 6 Distribution of P–P distance along Minor Groove for **a** P9–P28 (5th), **b** P10–P27 (6th), **c** P11–P26 (7th), **d** P12–P25 (8th), **e** P13–P24 (9th) and f P14–P23 (10th) [*blue* DNA from fiber model (system 3), *red* DNA in TBP–DNA complex (system 1)]

groups across minor groove (Table S10) and move the phosphate groups far apart. Therefore to see any positional preference of these active site residues, we have calculated distances between these charged residues from molecular dynamics simulations of the free as well as DNA bound proteins (Table 2). These residues can be organized into two sets, one binding to the first strand (amino acid residues 37, 60, 136, 141) and another binding to the second strand (residues 38, 45, 50, 67 and 151). Some of the distances between these two sets as mentioned before remain the same in both the systems (DNA bound and free TBP; indicated by italics) and these have smaller standard deviations. It is found that in most cases variations in their pair-wise distances are larger in case of the free protein as compared to those in the complex, e.g. distance between Lys-37 and Lys-151, where standard deviation reduces from 1.91 to 0.87 Å. On the other hand the distances between Lys-60 and Arg-50 remain fixed at 20.0 Å with small standard deviation in both the cases. Similar constant distances are also seen for Lys-60…Lys-67, Arg-136…Lys-67 and Arg-141…Lys-67 residue pairs. These distances are also very similar to the average DNA minor groove width in the TBP-bound state. This indicates that these residues are very static and if the DNA sequence has enough flexibility to bind to these residues, the DNA may deform and bind to the protein. Rigidity of these key residues of protein in the binding side may also help to hold the DNA in a bent conformation.

Conclusion

Our results indicate that TATA box containing DNA sequences have an inherent tendency of bending and minor-groove widening in terms of minor groove P–P distances. The TATA-box binding protein, on the other hand is quite rigid, especially its phosphate group binding basic residues. Variations of minor groove widths can thus be the critical determinant mediating overall shape complementarity between the DNA and protein surface as the binding takes place through interaction between phosphate groups of DNA and the rigid positively charged residues of the TBP. Due to the inherent bending flexibility of regular TATA-box sequence, DNA can adopt large number of thermally accessible conformations in multidimensional conformational space. The weakly populated, higher energy conformational sub-states of regular TATA sequence DNA are energetically favourable for binding to TBP with subsequent population shift toward the more deformed conformation in TBP–DNA complex and successively the binding interface is optimized for specific interactions via induced fit mechanism. In this way bending flexibility indirectly modulates the TBP binding activity

Table 2 Distance between DNA binding protein residues. The values in Italics font indicate distances which remain conserved in DNA-bound and free protein

Amino acid residue	DNA binding strand and % of occurrence	State	Lys(37) (1st)	Lys(60) (1st)	Arg(136) (1st)	Arg(141) (1st)	Arg(38) (2nd)	Arg(45) (2nd)	Arg(50) (2nd)	Lys(67) (2nd)	Lys(151) (2nd)
Lys(37)	1st	Bound		18.71	34.50	34.14	3.87	20.28	26.83	30.56	39.03
	31.75			(0.46)	(0.86)	(0.90)	(0.05)	(0.38)	(0.47)	(0.57)	(0.87)
		Free		19.00	35.33	35.13	3.87	20.26	26.82	31.08	38.96
				(0.62)	(1.81)	(1.71)	(0.05)	(0.64)	(0.85)	(1.11)	(1.91)
Lys(60)	1st	Bound			30.98	26.78	18.08	16.05	*20.20*	*21.48*	36.00
	16.00				(0.45)	(0.62)	(0.30)	(0.30)	(*0.28*)	(*0.30*)	(0.29)
		Free			31.01	26.99	17.96	16.00	*20.32*	*21.68*	35.78
					(0.83)	(0.90)	(0.54)	(0.30)	(*0.28*)	(*0.30*)	(0.52)
Arg(136)	1st	Bound				7.50	31.64	31.95	29.32	*23.23*	16.33
	93.20					(0.29)	(0.77)	(0.39)	(0.43)	(*0.38*)	(0.28)
		Free				7.45	32.36	32.55	29.97	*23.49*	16.17
						(0.29)	(1.64)	(0.63)	(0.62)	(*0.54*)	(0.31)
Arg(141)	1st	Bound					31.58	30.31	27.72	*21.02*	27.72
	22.76						(0.82)	(0.54)	(0.55)	(*0.48*)	(1.58)
		Free					32.36	30.95	28.46	*21.31*	28.07
							(1.56)	(0.68)	(0.63)	(*0.60*)	(2.12)
Arg(38)	2nd	Bound						17.67	23.76	27.36	35.36
	37.83							(0.53)	(0.61)	(0.63)	(0.85)
		Free						17.61	23.70	27.76	35.29
								(0.95)	(1.13)	(1.29)	(1.87)
Arg(45)	2nd	Bound							7.80	14.74	29.40
	59.31								(0.19)	(0.35)	(0.49)
		Free							7.80	15.26	29.74
									(0.21)	(0.38)	(0.73)
Arg(50)	2nd	Bound								8.41	24.02
	5.54									(0.46)	(0.53)
		Free								9.04	24.47
										(0.48)	(0.81)
Lys(67)	2nd	Bound									18.54
	2.15										(0.41)
		Free									18.71
											(0.64)
Lys(151)	2nd	Bound									
	26.65										
		Free									

that control transcription efficiency. As DNA flexibility can be large for various other sequences, some non-TATA-box sequence based promoters can also undergo the TBP induced conformational change, but TATA-box sequence surely is more preferred. The conformational heterogeneity due to inherent flexibility of biomolecules plays an important role in many protein-nucleic acid, protein–protein and protein–ligand interactions where large scale deformation or alteration of structures takes place after binding. Consideration of this conformational heterogeneity of the biomolecules can also lead to better understanding of flexible docking, enzyme catalysis, drug design and molecular evolution.

References

1. Deremble C, Lavery R (2005) Macromollecullar recognition. Curr Opin Struct Biol 15(2):171–175. doi:10.1016/j.sbi.2005.01.018
2. Sarai A, Kono H (2005) Protein–DNA recognition patterns and predictions. Annu Rev Biophys Biomol Struct 34:379–398. doi:10.1146/annurev.biophys.34.040202.144537

3. Rhodes D, Schwabe JWR, Chapman L, Fairall L (1996) Towards an understanding of protein–DNA recognition. Philos Trans R Soc Lond Ser B 351(1339):501–509. doi:10.1098/rstb.1996.0048
4. Paillard G, Lavery R (2004) Analyzing protein–DNA recognition mechanisms. Structure 12(1):113–122. doi:10.1016/j.str.2003.11.022
5. Fischer E (1894) Einfluss der configuration auf die wirkung der enzyme. Ber Dtsch Chem Ges 27:9
6. Gromiha MM, Siebers JG, Selvaraj S, Kono H, Sarai A (2004) Intermolecular and intramolecular readout mechanisms in protein–DNA recognition. J Mol Biol 337(2):285–294. doi:10.1016/j.jmb.2004.01.033
7. Gromiha MM, Slebcrs JG, Selvaraj S, Kono H, Sarai A (2005) Role of inter and intramolecular interactions in protein–DNA recognition. Gene 364:108–113. doi:10.1016/j.gene.2005.07.022
8. Koshland DE (1958) Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci USA 44(2):98–104. doi:10.1073/pnas.44.2.98
9. Ma BY, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. Protein Eng 12(9):713–720. doi:10.1093/protein/12.9.713
10. Troyer JM, Cohen FE (1995) Protein conformational landscapes—energy minimization and clustering of a long molecular-dynamics trajectory. Proteins Struct Funct Genet 23(1):97–110. doi:10.1002/prot.340230111
11. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254(5038):1598–1603. doi:10.1126/science.1749933
12. Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol 5(11):789–796. doi:10.1038/nchembio1209-954d
13. Kumar S, Ma BY, Tsai CJ, Sinha N, Nussinov R (2000) Folding and binding cascades: dynamic landscapes and population shifts. Protein Sci 9(1):10–19. doi:10.1110/ps.9.1.10
14. Grunberg R, Leckner J, Nilges M (2004) Complementarity of structure ensembles in protein–protein binding. Structure 12(12):2125–2136. doi:10.1016/j.str.2004.09.014
15. Wlodarski T, Zagrovic B (2009) Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. Proc Natl Acad Sci USA 106(46):19346–19351. doi:10.1073/pnas.0906966106
16. Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci 35(10):539–546. doi:10.1016/j.tibs.2010.04.009
17. Klug A, Jack A, Viswamitra MA, Kennard O, Shakked Z, Steitz TA (1979) Hypothesis on a specific sequence-dependent conformation of dna and its relation to the binding of the lac-repressor protein. J Mol Biol 131(4):669–680. doi:10.1016/0022-2836(79)90196-7
18. Dixit SB, Andrews DQ, Beveridge DL (2005) Induced fit and the entropy of structural adaptation in the complexation of CAP and lambda-repressor with cognate DNA sequences. Biophys J 88(5):3147–3157. doi:10.1529/biophysj.104.053843
19. Dragan AI, Read CM, Makeyeva EN, Milgotina EI, Churchill MEA, Crane-Robinson C, Privalov PL (2004) DNA binding and bending by HMG boxes: energetic determinants of specificity. J Mol Biol 343(2):371–393. doi:10.1016/j.jmb.2004.08.035
20. Crothers DM (1998) DNA curvature and deformation in protein–DNA complexes: a step in the right direction. Proc Natl Acad Sci USA 95(26):15163–15165. doi:10.1073/pnas.95.26.15163
21. Davis NA, Majee SS, Kahn JD (1999) TATA box DNA deformation with and without the TATA box-binding protein. J Mol Biol 291(2):249–265. doi:10.1006/jmbi.1999.2947
22. Cherstvy AG (2009) Positively Charged Residues in DNA-binding domains of structural proteins follow sequence-specific positions of DNA phosphate groups. J Phys Chem B 113(13):4242–4247. doi:10.1021/jp810009s
23. Garcia-Perez M, Pinto M, Subirana JA (2003) Nonsequence-specific arginine interactions in the nucleosome core particle. Biopolymers 69(4):432–439. doi:10.1002/bip.10389
24. Werner MH (1996) Intercalation, DNA kinking, and the control of transcription. Science 271(5250):778–784. doi:10.1126/science.271.5250.778
25. Privalov PL, Dragan AI, Crane-Robinson C (2009) The cost of DNA bending. Trends Biochem Sci 34(9):464–470. doi:10.1016/j.tibs.2009.05.005
26. Kim YC, Geiger JH, Hahn S, Sigler PB (1993) Crystal-structure of a yeast TBP TATA-box complex. Nature 365(6446):512–520. doi:10.1038/365512a0
27. Nikolov DB, Burley SK (1994) 2.1-Angstrom resolution refined structure of a TATA box-binding protein (TBP). Nat Struct Biol 1(9):621–637. doi:10.1038/nsb0994-621
28. Osheagreenfield A, Smale ST (1992) Roles of TATA and initiator elements in determining the start site location and direction of rna polymerase-II transcription. J Biol Chem 267(2):1391–1402
29. Singer VL, Wobbe CR, Struhl K (1990) A wide variety of DNA-sequences can functionally replace a yeast TATA element for transcriptional activation. Genes Dev 4(4):636–645. doi:10.1101/gad.4.4.636
30. Wang Y, Stumph WE (1995) RNA-polymerase II/III transcription specificity determined by TATA box orientation. Proc Natl Acad Sci USA 92(19):8606–8610. doi:10.1073/pnas.92.19.8606
31. Wobbe CR, Struhl K (1990) Yeast and human TATA-binding proteins have nearly identical DNA-sequence requirements for transcription invitro. Mol Cell Biol 10(8):3859–3867. doi:10.1128/MCB.10.8.3859
32. Kim JL, Nikolov DB, Burley SK (1993) Co-crystal structure of TBP recognizing the minor-groove of a TATA element. Nature 365(6446):520–527. doi:10.1038/365520a0
33. Kim JL, Burley SK (1994) 1.9-Angstrom resolution refined structure of TBP recognizing the minor-groove of TATAAAAG. Nat Struct Biol 1(9):638–653. doi:10.1038/nsb0994-638
34. Nikolov DB, Chen H, Halay ED, Hoffmann A, Roeder RG, Burley SK (1996) Crystal structure of a human TATA box-binding protein/TATA element complex. Proc Natl Acad Sci USA 93(10):4862–4867. doi:10.1073/pnas.93.10.4862
35. Wu J, Parkhurst KM, Powell RM, Brenowitz M, Parkhurst LJ (2001) DNA bends in TATA-binding protein center dot TATA complexes in solution are DNA sequence-dependent. J Biol Chem 276(18):14614–14622. doi:10.1074/jbc.M004402200
36. Wu J, Parkhurst KM, Powell RM, Parkhurst LJ (2001) DNA sequence-dependent differences in TATA-binding protein-induced DNA bending in solution are highly sensitive to osmolytes. J Biol Chem 276(18):14623–14627. doi:10.1074/jbc.M004401200
37. Geggier S, Vologodskii A (2010) Sequence dependence of DNA bending rigidity. Proc Natl Acad Sci USA 107(35):15421–15426. doi:10.1073/pnas.1004809107
38. Du Q, Kotlyar A, Vologodskii A (2008) Kinking the double helix by bending deformation. Nucleic Acids Res 36(4):1120–1128. doi:10.1093/nar/gkm1125
39. Vafabakhsh R, Ha T (2012) Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. Science 337(6098):1097–1101. doi:10.1126/science.1224139
40. Kannan S, Kohlhoff K, Zacharias M (2006) B-DNA under stress: over- and untwisting of DNA during molecular dynamics simulations. Biophys J 91(8):2956–2965. doi:10.1529/biophysj.106.087163
41. Starr DB, Hoopes BC, Hawley DK (1995) DNA bending ss an important component of site-specific recognition by the TATA-binding protein. J Mol Biol 250(4):434–446. doi:10.1006/jmbi.1995.0388

42. Qian XL, Strahs D, Schlick T (2001) Dynamic simulations of 13 TATA variants refine kinetic hypotheses of sequence/activity relationships. J Mol Biol 308(4):681–703. doi:10.1006/jmbi.2001.4617
43. Hancock SP, Ghane T, Cascio D, Rohs R, Di Felice R, Johnson RC (2013) Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. Nucleic Acids Res. doi:10.1093/nar/gkt357
44. Samanta S, Chakrabarti J, Bhattacharyya D (2010) Changes in thermodynamic properties of DNA base pairs in protein–DNA recognition. J Biomol Struct Dyn 27(4):429–442. doi:10.1080/07391102.2010.10507328
45. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28(1):235–242. doi:10.1093/nar/28.1.235
46. Chandrasekaran R, Arnott S (1996) The structure of B-DNA in oriented fibers. J Biomol Struct Dyn 13(6):1015–1027. doi:10.1080/07391102.1996.10508916
47. Case DA, VB JTB, Betz RM, Cai Q, Cerutti DS, Cheatham III TE, Darden TA, Duke RE, Gohlke H, Goetz AW, Gusarov S, Homeyer N, Janowski P, Kaus J, Kolossvary I, Kovalenko A, Lee TS, LeGrand S, Luchko T, Luo R, Madej B, Merz Jr KM, Paesani F, Roe DR, Roitberg A, Sagui C, Salomon-Ferrer R, Seabra G, Simmerling CL, Smith WL, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Kollman PA (2014) The FF14SB force field. AMBER 14 Reference Manual, pp 29–31
48. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) NAMD2: greater scalability for parallel molecular dynamics. J Comput Phys 151(1):283–312. doi:10.1006/jcph.1999.6201
49. Nelson M, Humphrey W, Kufrin R, Gursoy A, Dalke A, Kale L, Skeel R, Schulten K (1995) MDSCOPE—a visual computing environment for structural biology. Comput Phys Commun 91(1–3):111–133. doi:10.1016/0010-4655(95)00045-h
50. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117(19):5179–5197. doi:10.1021/ja00124a002
51. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys J 92(11):3817–3829. doi:10.1529/biophysj.106.097782
52. Foloppe N, MacKerell AD (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. J Comput Chem 21(2):86–104. doi:10.1002/(SICI)1096-987X(20000130)21:2<86:AID-JCC2>3.0.CO;2-G
53. Bansal M, Bhattacharyya D, Ravi B (1995) NUPARM and NUCGEN—software for analysis and generation of sequence-dependent nucleic-acid structures. Comput Appl Biosci 11(3):281–287
54. Mukherjee S, Bansal M, Bhattacharyya D (2006) Conformational specificity of non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. J Comput Aided Mol Des 20(10–11):629–645. doi:10.1007/s10822-006-9083-x
55. Mukherjee S, Majumdar S, Bhattacharyya D (2005) Role of hydrogen bonds in protein–DNA recognition: effect of nonplanar amino groups. J Phys Chem B 109(20):10484–10492. doi:10.1021/Jp0446231
56. Kabsch W, Sander C (1983) Dictionary of protein secondary structure—pattern-recognition of hydrogen-bonded and geometrical features. Biopolymers 22(12):2577–2637. doi:10.1002/bip.360221211
57. Bhattacharyya D, Bansal M (1989) A self-consistent formulation for analysis and generation of non-uniform DNA structures. J Biomol Struct Dyn 6(4):635–653. doi:10.1080/07391102.1989.10507727
58. Kanhere A, Bansal M (2003) An assessment of three dinucleotide parameters to predict DNA curvature by quantitative comparison with experimental data. Nucleic Acids Res 31(10):2647–2658. doi:10.1093/nar/gkg362
59. Dickerson RE (1998) DNA bending: the prevalence of kinkiness and the virtues of normality. Nucleic Acids Res 26(8):1906–1926. doi:10.1093/nar/26.8.1906
60. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K (2009) Conformational analysis of nucleic acids revisited: curves+. Nucleic Acids Res 37(17):5917–5929. doi:10.1093/nar/gkp608
61. Skjaerven L, Martinez A, Reuter N (2011) Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. Proteins Struct Funct Bioinf 79(1):232–243. doi:10.1002/prot.22875
62. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD (2006) Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics 22(21):2695–2696. doi:10.1093/bioinformatics/btl461
63. Becker OM (1998) Principal coordinate maps of molecular potential energy surfaces. J Comput Chem 19(11):1255–1267. doi:10.1002/(sici)1096-987x(199808)19:11<1255:aid-jcc5>3.3.co;2-h
64. Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338. doi:10.1093/biomet/53.3-4.325
65. Becker OM (1997) Geometric versus topological clustering: an insight into conformation mapping. Proteins Struct Funct Genet 27(2):213–226. doi:10.1002/(sici)1097-0134(199702)27:2<213:aid-prot8>3.0.co;2-g
66. Brandl CJ, Struhl K (1990) A nucleosome-positioning sequence is required for GCN4 to activate transcription in the absence of a TATA element. Mol Cell Biol 10(8):4256–4265. doi:10.1128/MCB.10.8.4256
67. Anish R, Hossain MB, Jacobson RH, Takada S (2009) Characterization of transcription from TATA-less promoters: identification of a new core promoter element XCPE2 and analysis of factor requirements. PLoS One. doi:10.1371/Journal.Pone.0005103
68. Mondal M, Mukherjee S, Bhattacharyya D (2014) Contribution of phenylalanine side chain intercalation to the TATA-box binding protein-DNA interaction: molecular dynamics and dispersion-corrected density functional theory stidies. J Mol Model 20(11):2499. doi:10.1007/s00894-014-2499-7
69. Calladine CR, Drew HR, Luisi BF, Travers AA (2004) Understanding DNA, the molecule and how it works, 3rd edn. Elsevier, London