Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2006) 158–168
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Exploring the potential of complex formation between a mutant DNA and the wild type protein counterpart: A MM and MD simulation approach

Sujata Roy, Srikanta Sen *

Molecular Modeling Section, Chembiotek Research International, Bengal Intelligent Park Building, Tower B, Block EP & GP, Salt lake Electronics Complex, Calcutta 700091, West Bengal, India

Received 14 September 2005; received in revised form 20 October 2005; accepted 11 November 2005
Available online 6 January 2006

Abstract

We have demonstrated that the methods of molecular modeling and molecular dynamics simulation might be used to assess whether a specific mutation in the DNA would destabilize a known DNA–protein complex. The approach is based on probing the changes in the interaction that would be induced into the complex if within the already formed wild type complex the mutation could be introduced. We have used Hoxc8–DNA complex as a test system where it is known that the Hoxc8 binding affinity of the DNA is completely lost upon mutation of the DNA by replacing TAAT stretch to GCCG. Mutation was obtained by changing the relevant base pairs into the DNA of the model of the corresponding wild type complex developed by homology modeling and MD simulation in water for 2.0 ns. Comparison of the structure, dynamics and interactions between the hypothetical mutant model with those of the similarly refined wild type model shows that the loss of affinity of the mutant DNA to Hoxc8 has two different origins: (i) loss of several strong H-bonds as the direct consequences of mutation and (ii) reduced H-bonds in the common parts due to a net loss or inferior H-bonding geometry induced by the mutation as indirect effects. The net change in the interaction energy between the DNA and the protein in the best possible configuration indicated the experimentally observed destabilization effects. No significant change in the groove width was observed and no correlation was found between the water-bridges and the loss of affinity.

Keywords: Molecular recognition; DNA–protein complex; Interaction pattern; H-bonding; Molecular dynamics

1. Introduction

Protein–DNA complexes form a class of molecular system that performs important cellular tasks like gene regulation, replication, etc., and play a vital role in understanding the basic mechanisms of biomolecular recognition. Mutation in the recognition part of a DNA generally causes alterations in its protein binding affinity. There are well-established ways of estimating the changes in binding affinity based on free energy perturbation methods only for small mutations [1–4]. However, no prescription has been suggested for identifying whether a larger mutation might cause only mere alterations in the binding affinities or total destabilization of the complex. The objective of the present work is to demonstrate that the methods of molecular modeling and molecular dynamics simulation can be used to anticipate whether some mutation in the DNA will allow it to form a complex with the protein counterpart or not.

There are two main possibilities that may prevent the mutant DNA from forming a stable complex with the protein: (i) the mutation may cause considerable structural change particularly the groove width in the DNA and (ii) the interactions at the DNA–protein interface may be weakened considerably due to mutation. Modeling the hypothetical mutant-DNA–protein complex onto the model structure of the wild type DNA–protein complex as the template should highlight the differences in the interaction pattern at the DNA–protein interface. The model mutant complex should also indicate the alteration occurred in the groove width of the DNA due to mutation.

Being motivated by the above idea we have chosen DNA–Hoxc8 complex as a suitable working system because there are clear experimental evidences that the wild type DNA forms stable complex with the Hoxc8 protein while, a specific mutant DNA does not form a complex at all [5]. Hoxc8 is a repressor

* Corresponding author. Tel.: +91 33 2357 0340; fax: +91 33 2357 0342.
E-mail address: srikanta@chembiotek.com (S. Sen).

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.11.004

protein that modulates the expression of the osteopontin gene. It is a homeodomain protein and binds only to the response element TAAT having the immediate sequence context as in the sequence AGT**TAAT**GACATC. It has further been shown experimentally that if the TAAT stretch in the DNA is replaced by a GCCG stretch, the Hoxc8 binding activity of this DNA is totally lost [5].

Understanding the molecular basis of the molecular recognition requires the 3D structure. The crystal structure of the wild type DNA–Hoxc8 complex is not yet available. Recently, we have predicted a 3D structure of the Hoxc8–DNA complex based on homology modeling using the crystal structure of Antennapedia–DNA complex (9ant.pdb) [6] as the template. A detailed account of the method and the structural and energetic characteristics of the modeled complex have already been reported [7]. Our predicted structure of the wild type Hoxc8–DNA complex shows good agreement with the available experimental data on Hoxc8 and is consistent with the general features of other homeodomain–DNA complexes [7].

The basic idea is to probe the changes in interaction pattern that would be induced into the complex if within the already formed wild type complex the mutation could be introduced. Molecular modeling gives us an opportunity to generate a possible 3D structure of an unknown macromolecular system [7–9]. Similar approach can be employed to generate 3D structure representing the unstable mutant complex in order to identify the structural and interactional differences with reference to the wild type complex and to correlate them with the observed loss of stability.

Here, we have mapped the mutant complex onto this recently developed model of the wild type complex. The model mutant complex has been refined by performing MD simulation (2.0 ns) in aqueous solvent following an identical protocol that we used to refine the model wild type complex. This should reflect the effects of the mutation at the structural and interactional level. Comparison between the wild type and mutant complexes are then expected to highlight the interactional basis of the loss of binding affinity and their structural correlation. Considerable amount of work has been done demonstrates the importance of H-bonds and water-bridge in stability of biomolecular complex [10–12].

Analysis shows that considerable reduction in interaction energy occurred due to: (i) loss of a number of strong H-bonds as the direct consequences of mutation and (ii) reduced H-bonding interactions in the common parts either due to a net loss or inferior H-bonding geometry induced by the mutation as indirect effects. These findings are well correlated with the experimentally observed loss of binding affinity of the mutant DNA to Hoxc8. Interestingly, we have not found any correlation between the contribution of water-bridges and the loss of binding affinity.

We have not followed the standard way of comparing the relative binding affinity of the DNA protein complexes (wild type and mutant) in terms of free energy differences [1–4] because such a large change is not suitable for computing free energy difference by methods like thermodynamic perturbation. However, the present work additionally provides valuable insights regarding the molecular basis of loss of stability of the complex.

Finally, it should be pointed out that similar approach is not suitable for investigating the effects of even moderate mutations in proteins as that may alter the fold of the protein.

2. Methods

2.1. Construction of the initial model of the mutated DNA–Hoxc8 complex

We mapped the mutated system onto the energy minimized trajectory averaged structure of the wild type Hoxc8–DNA complex we obtained in our previous work [7] by changing the bases ^{4}TAAT^{7} of S2 strand to ^{4}GCCG^{7} and accordingly in the complementary strand also to preserve the base-pairs complementarities (Fig. 1). We first removed all the relevant bases from the DNA part retaining only the backbone parts. Then using the BUILD facility of the software CHARMM [13,14] the coordinates of the atoms of the missing bases were generated using the CHARMM topology and the geometry of the local backbone structure as the seed. The values of the angle chi and the sugar pucker were kept the same as in the template because our purpose was to generate the mutated base pairs close to the best possible complementary paired geometry. The resulting structure was then energy minimized in vacuum by 5000 steepest descent (SD) steps keeping the entire complex fixed excepting the newly added altered bases. This allowed these altered bases to accommodate themselves in a proper way into the structure defined by the rest of the complex. The end base pairs on both ends were constrained harmonically with a force constant of 10.0 kcal/Å^2/mol to their H-bonded base-paired configurations. Again, as we have modeled only a part (56 residues out of a total of 242 residues) of Hoxc8 protein, the distances between the Cα-atom of the residue Arg^{153} and the closest P atom S2-C^{6}:P of the DNA and also between the Cα-atoms of the other end residue Asn^{208} and the spatially nearest residue Leu^{169} were harmonically constrained by a weak force constant of 10.0 kcal/Å^2/mol to partially mimic the covalent continuity of the ends. It may be mentioned at this point that initially, we performed the MD simulation without this constraint but found that several artifacts in the structure and interaction pattern appeared due to the artificial freedom of the end residue Arg^{153} in the absence of the covalent continuation of Hoxc8. Use of this weak constraint removed those artifacts. It was then energy minimized again by 5000 SD steps removing all constraints excepting the constraints for reducing the end effects. All bond lengths involving hydrogen were kept fixed using SHAKE algorithm [15,16].

2.2. Solvation and system set up

Eighteen Na^+ counter-ions were added to the energy-minimized modeled complex for the electro-neutrality of the system. The counter-ions were placed following two different protocols. For the phosphate groups of the DNA, which are accessible to solvent, sodium ions were placed at a position of

Fig. 1. (a) Comparison of the aligned sequences of the template protein (9ant.pdb) and the human Hoxc8 protein. The conserved region of the Hoxc8 protein is shown in capital letter. Identical amino acid residues are shown with shade. (b) The aligned sequences of the template DNA and the Hoxc8 binding DNA are compared. The nucleotides that are identical with the template are shown with shade. (c) The sequences of the wild type Hoxc8 binding DNA duplex and mutant DNA duplex are shown separately. The mutated nucleotides are shown with shade.

6.0 Å from the phosphorous atom along the line starting from the phosphorous atom and bisecting the line joining the phosphate oxygen atoms. By this way we could place only eight Na^+ ions. The system was then solvated by placing it at the center of a pre-equilibrated TIP3P water sphere of radius 25.0 Å [17]. As the largest radius of the whole complex excepting the end base pairs of the DNA is 19.8 Å and it is 17.0 Å for only the protein, the water sphere of 25.0 Å effectively leaves a solvent layer of thickness 5.0 Å or more. Any water molecule having its oxygen atom within a distance of 2.8 Å from any non-H atom of the complex was deleted. Each of the other 10 sodium ions was placed by replacing the water molecule whose oxygen atom has the highest electrostatic energy. No sodium ion was placed closer than 5.0 Å to each other and all the counter ions were kept unconstrained during energy minimization and MD simulation. The effect of the bulk water that could be present outside the solvent sphere was taken into account by the mean-field effects [18]. It not only minimized the finite boundary effects but also prevented the evaporation of water molecules from the surface of the solvent sphere. The end base pairs were constrained harmonically with a force constant of 10.0 kcal/Å^2/mol to their H-bonded base-paired configurations in order to prevent unrealistic end effects like base-pair opening, buckling, etc. The constraints on the end residues of Hoxc8 were also continued during energy minimization and subsequent MD simulation in order to partially mimic the effect of covalent continuity of Hoxc8 as mentioned earlier. The resulting system was then energy minimized again by 5000 steepest descent steps keeping the Hoxc8–DNA complex fixed to allow only the water molecules to reorient themselves for eliminating any steric conflict with the complex. As Ewald method [19–21] cannot be employed in spherical systems we have used standard spherical cut-off methods with a cut-off values 12.0 Å along with the force shift method in handling the electrostatics where the interaction energy was smoothly shifted to zero at a cut-off distance of 11.0 Å [22,23]. This cut-off combination is known to produce stable and reliable trajectories [24] and thus appears to be suitable for relaxing the structure of a model complex. A dielectric constant 1.0 was used for the electrostatic interactions

during energy minimization and MD simulation. A further energy minimization for 5000 SD steps was performed without fixing any atom of the complex and the resulting structure was then relaxed by MD simulation at 300 K.

2.3. Dynamic simulation protocol for relaxing the model complex

MD simulation was performed using leapfrog algorithm [25] with an integration time step of 2.0 fs by employing the software CHARMM version 28 along with the combined nucleic acid and protein parameter set version 27 [26]. All bond lengths involving hydrogen were kept fixed using SHAKE algorithm [16,27]. The pair list of non-bonded interactions was updated every 20 steps. During the MD simulation we first thermalized the system by assigning velocities drawn from a Gaussian distribution at 100 K to the atoms of the system and then gradually heated the system over a time period of 10.0 ps by assigning velocities in steps at higher temperatures until it reached the target temperature 300 K. The system was subsequently equilibrated at 300 K over the next 10.0 ps by assigning velocities from a Gaussian distribution of velocities at 300 K. The simulation was then continued and atomic velocities were rescaled only if the average temperature was outside the window of ± 5 K to maintain the temperature around 300 K. The simulation was continued for a total of 2.0 ns and the coordinate frames were saved at intervals of 0.4 ps for analysis.

2.4. Analysis of the local relaxation and interaction pattern

Local structural relaxation of the mutant DNA–Hoxc8 complex in incorporating the mutation was monitored by the time evolution of the RMSD of the frames with respect to the initial structure. All the non-hydrogen atoms in the entire complex were considered for superposition and were also used for computing the RMSD for the complex.

Interaction energies between DNA and Hoxc8 protein were estimated by computing the electrostatic and the van der Waal components separately in order to identify the physical nature of the dominating interactions. The electrostatic interactions were estimated from trajectories using distance dependent dielectric constant as the trajectory was generated in the presence of explicit water as solvent. The interaction energies between DNA bases and the protein were computed for each frame of the trajectory over the last 600 ps and the average and the RMS fluctuations of the values were computed to characterize the interactions. In a similar fashion the residue-wise interaction of the protein residues with the DNA were also computed.

Direct H-bonding interactions between Hoxc8 and the DNA were analyzed by using the following criteria. A H-bond acceptor–donor pair is considered to form a hydrogen bond if the acceptor-hydrogen distance (d_{AH}) ≤ 2.4 Å and the donor–hydrogen-acceptor angle (θ_{DHA}) $\geq 135^{\circ}$. The hydrogen bonding statistics was calculated during the last 600 ps of the trajectory, using the coordinate frames at intervals of 0.4 ps and the average lifetimes of the H-bonds were computed by considering only the H-bonded events having duration longer than 5 ps.

A water-bridge is defined as an interaction between Hoxc8 and the DNA through hydrogen bonds mediated by a common water molecule. In identifying the water-bridge we have used the same distance and angle cut-offs and duration of the event length as used in hydrogen-bond analysis. H-bond donor acceptor declaration was added in the TIP3P topology of CHARMM to allow H-bonding considering the TIP3P oxygen atom acting as a H-bond donor as well as an acceptor.

3. Results

3.1. Comparison of the H-bonding patterns of the wild type complex and the initial model of the mutant complex

In order to see how different was the H-bonding patterns between the initial model of the mutant DNA–Hoxc8 complex and its wild type counterpart, we compared the H-bonding patterns of the wild type complex and the energy-minimized model of the mutant complex before its structural relaxation by MD simulation. Table 1 summarized the comparison. It is seen that the H-bonds involving the S1 strand remained almost the same and in the strand S2 the number of H-bonds remained the same even though due to mutation all the nucleotides involved are different. It may further be pointed out that there were two specific H-bonds in the wild type complex while in the mutant there was only one. Moreover, it was seen (see Table 3) that after structural relaxation many of these initially observed H-bonds were lost.

3.2. Relaxation of the model complex

Molecular dynamics simulation was required for allowing the local relaxation of the complex in order to incorporate the mutations into the 3D structure keeping the resulting structure as close as possible to the native complex. Thus, the relaxed structure represents the mutated complex where the mutation has been incorporated in the best possible way. The time evolution of the RMSD of the whole mutant DNA–Hoxc8 complex with respect to the initial structure is shown in Fig. 2. The reasonably stable behavior of the RMSD indicated that the local relaxations in the complex have been completed.

The trajectory averaged structure of the hypothetical mutant system was energy minimized by 5000 SD steps in vacuum and are shown in Fig. 3a (stereo-view). The corresponding structure of the wild type complex is presented in Fig. 3b (stereo-view).

3.3. Specific interactions (between DNA bases and Hoxc8)

Specific interactions are considered as the most important determinants of the molecular recognition process in DNA–protein complexes. Comparison of the average specific interaction of the DNA bases and Hoxc8 against the bases between the wild type and the mutant complexes indicates that there is no significant change in the interaction pattern in the

Table 1
Comparison of the H-bonding pattern of the wild type complex and the initial model of the mutant complex

Hydrogen bonded atoms	Wild type		Mutant	
	Distance (A–H) (Å)	Angle (A···D-H) (°)	Distance (A–H) (Å)	Angle (A···D-H) (°)
S1-T^4:O1P···ARG^{176}:HN	1.75	159.9	1.59	166.3
[a]S1-T^4:O2P···ARG^{179}:HH11	1.8	152.2		
S1-T^4:O2P···ARG^{179}:HH21	1.77	148.8	1.48	169.6
[b]S1-T^4:O5′···ARG^{179}:HH11			1.89	165.6
S1-G^5:O1P···LEU^{174}:HN	1.77	176.4	1.74	178.0
[a]S1-G^5:O5′···ARG^{201}:HH11	1.99	167.2		
S1-G^5:O1P···ARG^{201}:HH21	1.66	165.9	1.58	168.9
[b]S1-G^5:O2P···LYS^{194}:HZ3			1.51	161.3
S1-T^6:O1P···ARG^{201}:HH12	1.63	169.4	1.53	168.8
[b]S1-T^6:O2P···TYR^{173}:HH			1.5	177.0
[a]S1-C^7:O1P···LYS^{205}:HZ2	1.55	159.7		
[b]S1-C^7:O2P···LYS^{205}:HZ2			1.55	159.6
[b]S2-T^3:O2P···LYS^{206}:HZ3			1.49	176.5
[a]S2-T^4:O1P···LYS^{203}:HZ1	**1.53**	**165.0**		
[b]S2-G^4:O1P···LYS^{203}:HZ3			**1.46**	**174.9**
[b]S2-C^5:O2P···ASN^{199}:HD21			**1.66**	**170.3**
S2-A^6/C^6:O1P ···GLN^{154}:HN	**1.64**	**178.7**	**1.61**	**173.8**
[b]S2-C^6:O1P···GLN^{154}:HE21			**1.71**	**174.0**
[a]S2-A^6:O2P···GLN^{154}:HE21	**1.92**	**153.2**		
S2-A^6/C^6:O2P···TYR^{156}:HH	**1.64**	**170.6**	**1.47**	**175.3**
[a]S2-A^6:N7···ASN^{199}:HD22	**2.04**	**174.4**		
[a]S2-A^6:H61···ASN^{191}:OD1	**1.95**	**168.8**		
[b]S2-C^6:H42···ASN^{199}:OD1			**1.95**	**166.5**
S2-T^7/G^7:O2P···ARG^{191}:HH11	**1.54**	**173.1**	**1.45**	**177.0**
S2-T^7/G^7:O5′···ARG^{191}:HH21	**1.94**	**161.2**	**1.89**	**170.2**
[b]S2-G^8:O2P···ARG^{191}:HH22			1.52	172.0

The data for the recognition site S2:^{4}TAAT^7 in wild type and mutated site S2:^{4}GCCG^7 are shown in bold letter.
[a] Those are present in the wild type system only.
[b] Those are present only in the mutant system.

case of the S1 strand (Fig. 4a). It is also clear that the overall specific interactions are quite weak for the S1 strands in both the cases. On the other hand, in the case of the strand S2 there are marked differences in the interaction patterns between the wild type and the mutant systems (Fig. 4b). It is seen that the common base S2-T^3 has generated a significant unfavorable interaction with Hoxc8 that was practically absent in the wild type complex. The interactions with S2-G^4 and S2-C^5 in the altered part of the mutant system have shown some improvements in their favorable interaction with Hoxc8 compared to the interactions with the corresponding bases in the wild type complex. A drastic reduction in favorable interaction was observed for S2-C^6 in the mutant compared to that with the respective base S2-A^6 in the wild type complex. The last altered base S2-G^7 and the common base S2-G^8 are also found to become less favorable in the mutant compared to the wild type complex. The overall average specific interaction between the DNA and Hoxc8 in the mutant case (-5.36 ± 3.47 kcal/mol) appears significantly weaker compared to that of the wild type complex (-13.14 ± 4.58 kcal/mol). It may be pointed out that the interactions are considerably fluctuating in both types of complexes and appears to be the intrinsic features of the system as similar fluctuating interactions are observed in NMR experiments with the Antennapedia homeodomain–DNA complex [6]. The observed differences in the specific interaction patterns between the wild type and mutant systems seem to have two different origins: (i) loss of H-bonds due to the absence of some of the functional groups as a direct consequence of the mutation and (ii) loss or weakening of H-bonds in the conserved regions due to inferior H-bonding geometry induced as an indirect

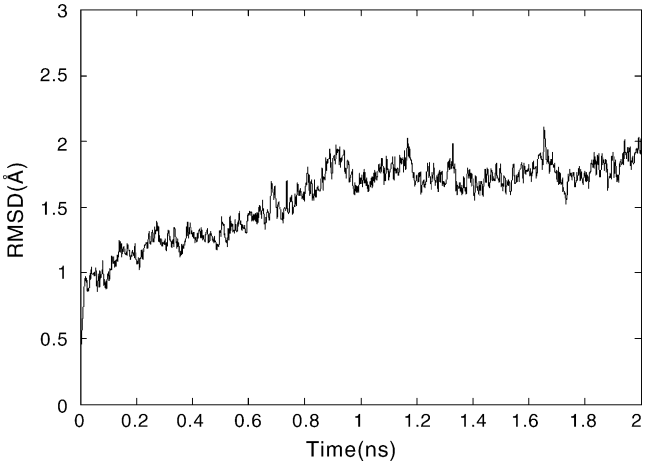

Fig. 2. Time evolutions of RMSD of the mutant complex with respect to the respective initial structures. H-atoms, the end base pairs of DNA and the end residues of Hoxc8 are excluded. In computing the RMSD values superposition of the entire complex (without H-atoms) was used. For the sake of clarity data are shown only at the interval of 2.0 ps.

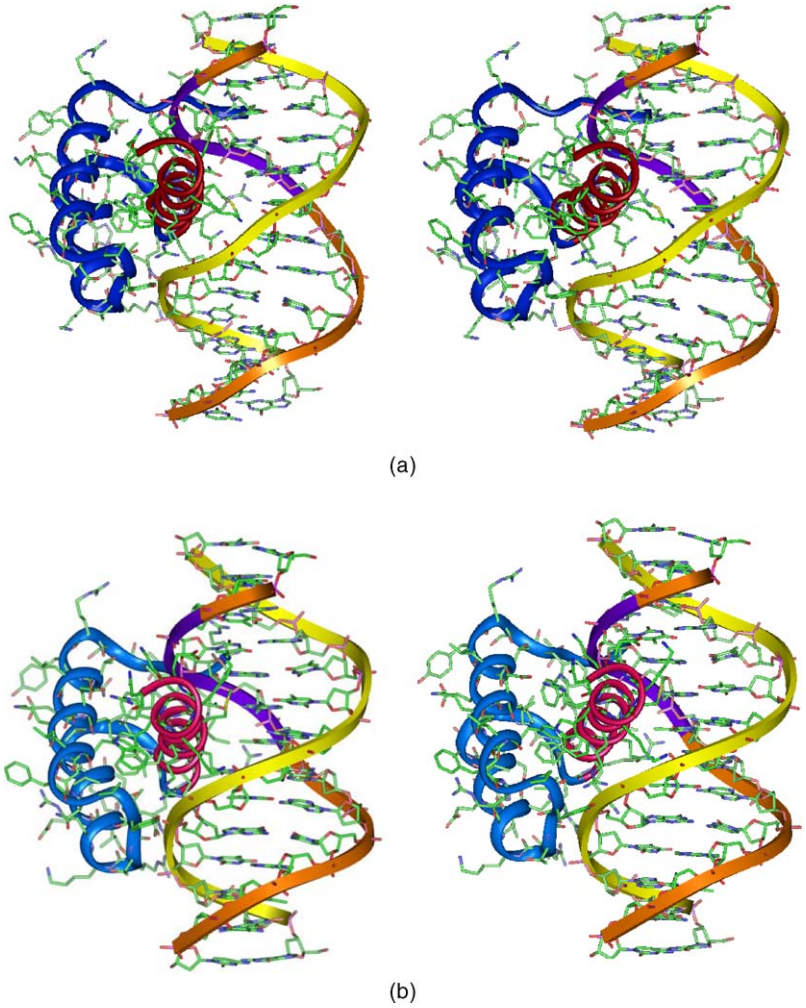

Fig. 3. The stereo views of the structure of (a) mutant Hoxc8–DNA complex and (b) wild type Hoxc8–DNA complex. The structures were obtained by energy minimization of the respective structures of each of the complexes averaged over the last 600 ps of the respective trajectory as described in the text. The H-atoms are not shown for clarity. The Hoxc8 protein is represented as solid oval blue ribbon where 'helix-3' is shown in red. The rectangular yellow ribbon represents the S1 strand of the DNA and the other strand S2 containing the TAAT stretch is shown by saffron ribbon. The violet parts of the rectangular ribbons represent the TAAT (in wild type) and GCCG (in mutant system) stretches on the S2 strand. Changes in the complementary stands are not marked.

consequence of the mutations through the structural rearrangements in the DNA and Hoxc8.

3.4. Non-specific interactions (between DNA backbone and Hoxc8)

Non-specific interactions are important for the overall stability of the complex and also determine the specificity in an indirect way. Comparison of the patterns of the non-specific interactions in the S1 strands of the wild type complex and the mutant system also indicates that there is a significant reduction of the interactions for most of the nucleotides excepting a few cases where the interactions remain comparable (Fig. 4c). On the other hand, in the case of the S2 strand, the non-specific interactions for the nucleotide S2-T^3 (common) and S2-C^5 (one of the mutated nucleotide) the interactions have become considerably favorable in the mutant case compared to the respective nucleotides in the wild type one (Fig. 4d). However, the interactions for other two nucleotides S2-C^6 and S2-G^7 have been reduced drastically in the mutant case with reference to that of the respective nucleotides in the wild type complex. The over all average non-specific interactions between the DNA and Hoxc8 over the last 600 ps of the respective trajectories for the mutant case (-263.0 ± 16.6 kcal/mol) appears to be much unfavorable compared to that in the wild type complex (-295.7 ± 20.6 kcal/mol). It is further interesting to note that the fluctuations in the non-specific interactions energies are considerably different for different nucleotides and in some of the cases the fluctuations in the mutant are larger than that in the wild type complex.

As the bases are not involved in the non-specific interactions the observed differences in the patterns in the non-specific interactions between the wild type and mutant systems correspond to the loss or weakening of H-bonds due to distorted H-bonding geometry induced as an indirect consequence of the mutations.

Comparison of the total (specific + non-specific) interaction energies between the wild type (-308.9 ± 19.5 kcal/mol) and

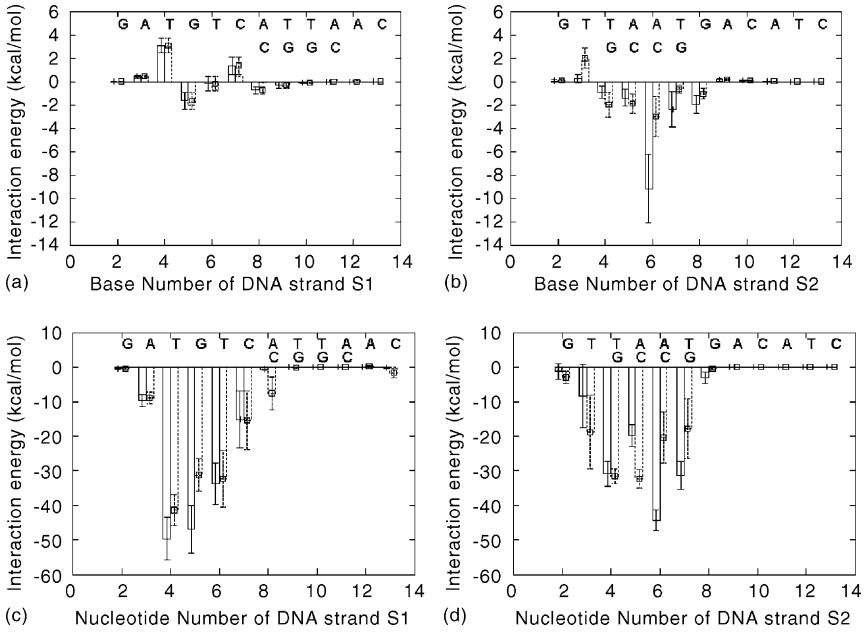

Fig. 4. Comparison of the specific interaction energies between Hoxc8 and the DNA against the base number for (a) strand S1 and (b) strand S2. The non-specific interaction energies between Hoxc8 and DNA for (c) strand S1 and (d) strand S2 are also presented. Bars with solid lines represent the wild type cases, while bars with dotted lines show the mutant cases. The error bars represent the RMS fluctuations of energies.

the mutant system (-268.4 ± 15.5 kcal/mol) clearly indicates that the overall stability of the mutant system is dramatically reduced.

3.5. Interaction pattern against Hoxc8 residue number

In order to identify the Hoxc8 residues that are important for interaction with the DNA the interaction patterns are plotted in Fig. 5. Comparison shows that the pattern of the DNA–Hoxc8 interaction against the residue number of Hoxc8 in the mutant complex differs significantly from that of the wild type complex (Fig. 5). It is found that the residue-wise interactions either have become less favorable in some cases or remained practically unchanged for most of the others, while only in one case the interaction energy has been increased significantly. In a few cases, considerable unfavorable interactions have also resulted in due to the mutation. For example, the interaction with Glu^{207} has changed from +6.5 kcal/mol in the wild type complex to +12.2 kcal/mol in the mutant. The observed major differences are summarized in Table 2.

3.6. H-bonding

H-bonds play an important role in molecular recognition and the overall stability of the complex. Comparison of the H-bonding patterns (Table 3) indicates that all the specific H-bonds that were present in the wild type complex are now completely missing in the mutant. These are due to the absence of the relevant functional groups of the bases, and are the direct consequences of the mutation. It is important to point out that all these missing H-bonds were very strong and stable in the wild type complex. The fact, that no specific H-bonds common to mutant and wild type are found, is directly correlated to the loss of specificity in binding affinity of the mutant DNA to Hoxc8. Some non-specific H-bonding interactions are also missing in the mutant and it seems to be due to the indirect consequences of the large mutation involving alteration of four base pairs. All these appear to be due to the indirect consequences of the mutation.

It is further seen that out of the seven non-specific H-bonds that are commonly present in the wild type and the mutant variant, three have been significantly weakened in terms of both the residence time as well as the H-bonding geometry.

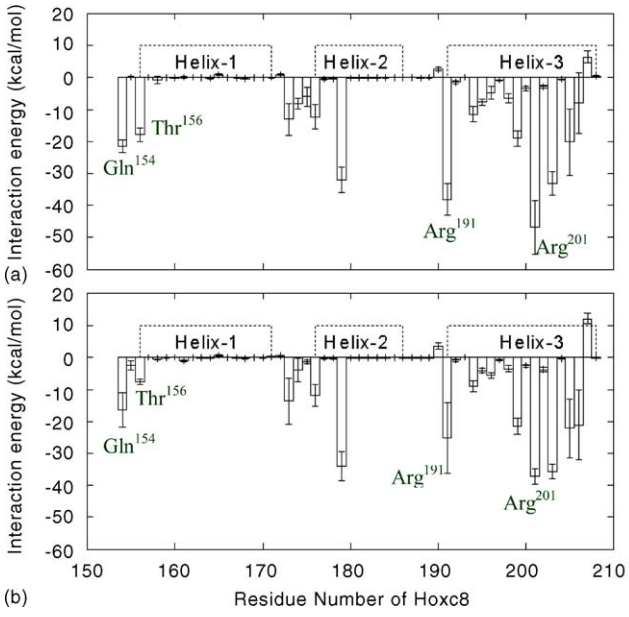

Fig. 5. The pattern of the interactions between DNA and Hoxc8 against the residue number of Hoxc8 in (a) the wild type complex and (b) the mutant. The RMS fluctuations in the interaction are shown as the error bars.

Table 2
Comparison of the interaction energies of some of the major residues of Hoxc8 with the DNA between the wild type and mutant complexes in the respective energy minimized trajectory averaged structures

Protein residues	Interaction energy (kcal/mol)	
	Wild type	Mutant
Gln^{154}	−21.4	−16.3
Thr^{156}	−17.9	−7.6
Arg^{190}	2.6	3.3
Arg^{191}	−38.1	−25.1
Lys^{194}	−11.4	−9.0
Ile^{195}	−7.6	−4.0
Gln^{198}	−6.4	−3.5
Arg^{200}	−3.3	−2.5
Arg^{201}	−46.9	−37.2
Lys^{203}	−7.9	−21.1
Glu^{207}	6.5	12.2

However, the other four have improved with longer residence time and slightly better H-bonding geometry.

Interestingly, seven new H-bonds, that were absent in the wild type complex, have been found in the mutant. However, only three of them have considerable residence time.

Above all, the total number of H-bonds obtained in the mutant is almost half of that in the corresponding wild type case and this correlates well with the observed loss of binding affinity of the mutant DNA to Hoxc8.

3.7. Water-bridge

Water-bridges are believed to be important for the specificity of the biomolecular recognition [12,28,29]. Table 4 summarizes the comparison of the water-bridges between the two cases of wild type and mutant DNA–hoxc8 complexes. Out of a total of 33 water-bridges in the wild type complex and 47 in the mutant complex obtained by analysis, we have listed only the water-bridges whose total residence time over the last 600 ps of the trajectories are more than 40.0 ps. It is observed that most of the neglected water-bridges are very short lived having total residence times in the range 5–25 ps only. Table 4 indicates that both the two base specific water-bridges (S1-C^7:H42···Gln^{198}:HE21, S2-T^7:O4···Asn^{199}:OD1) are missing in the mutant. The mutation has mixed effects on the water-bridging process. Some of the water-bridges that are common to both types of complexes are found to be weakened in the mutant

Table 3
The statistics of lifetimes and the average geometries of the H-bonds between Hoxc8 and DNA are compared between the cases of wild type and mutant systems

Hydrogen bonded atoms	Wild type				Mutant			
	Total residence time (ps)	Average life time (ps)	Distance (A–H) (Å)	Angle (A···D-H) (°)	Total residence	Average life time (ps)	Distance (A–H) (Å)	Angle (A···D–H) (°)
[a]S1-T^4O1P···ARG^{176}:HN	266.0	12.1	1.9 ± 0.2	153.4 ± 10.1				
S1-T^4:O2P···ARG^{179}:HH11	534.0	35.6	1.8 ± 0.2	159.0 ± 10.2	215.2	11.3	1.9 ± 0.2	156.1 ± 10.8
S1-T^4:O2P···ARG^{179}:HH21	72.0	10.3	1.9 ± 0.2	150.4 ± 9.5	178.4	14.9	1.8 ± 0.2	156.7 ± 10.6
[a]S1-G^5:O5′···ARG^{201}:HH11	58.0	8.3	2.1 ± 0.2	159.8 ± 11.1				
[a]S1-G^5:O1P···ARG^{201}:HH21	148.0	21.1	1.9 ± 0.2	158.3 ± 10.5				
S1-G^5:O1P···LEU^{174}:HN	586.0	73.3	1.9 ± 0.2	162.7 ± 9.4	212.8	9.7	2.0 ± 0.2	165.5 ± 8.4
[a]S1-T^6:O2P···TYR^{173}:HH	164.0	41.0	1.9 ± 0.2	163.3 ± 8.3				
S1-T^6:O1P···ARG^{201}:HH12	576.0	115.2	1.8 ± 0.1	162.1 ± 8.5	597.6	85.4	1.8 ± 0.1	164.4 ± 8.0
[b]S1-T^6:O2P···TYR^{173}:HH					201.6	67.2	1.8 ± 0.1	162.8 ± 8.4
[a]S1-C^7:O1P···LYS^{205}:HZ1	56.0	11.2	1.8 ± 0.1	159.0 ± 11.1				
[b]S1-C^7:O2P···LYS^{205}:HZ2					57.6	19.2	1.8 ± 0.1	157.6 ± 10.8
[b]S2-T^3:O2P···LYS^{2O6}:HZ1					51.2	12.8	1.7 ± 0.1	160.4 ± 10.0
[b]S2-T^3:O2P···LYS^{206}:HZ2					82.4	10.3	1.7 ± 0.1	159.3 ± 10.7
S2-T^3:O2P···LYS^{206}:HZ3	48.0	24.0	1.7 ± 0.1	160.9 ± 10.5	63.6	21.2	1.7 ± 0.1	163.0 ± 9.0
[a]S2-T^4:O1P···LYS^{203}:HZ2	166.0	23.7	1.8 ± 0.1	160.4 ± 10.2				
[a]S2-T^4:O1P···LYS^{203}:HZ3	180.0	22.5	1.8 ± 0.1	161.6 ± 10.4				
[a]S2-T^4:O1P···LYS^{203}:HZ1	152.0	30.4	1.8 ± 0.1	162.2 ± 9.8				
[b]S2-G^4:O2P···LYS^{203}:HZ2					590.4	45.4	1.7 ± 0.1	163.9 ± 8.7
S2-A^5/C^5:O2P···ASN^{199}:HD21	76.0	8.4	2.1 ± 0.2	155.1 ± 9.4	428.0	12.6	2.0 ± 0.2	162.8 ± 9.1
[b]S2-C^5:O3′···GLN^{154}:HN					69.6	6.3	2.2 ± 0.1	161.7 ± 9.3
[a]S2-A^6:H62···ASN^{199}:OD1	96.0	16.0	2.0 ± 0.2	162.2 ± 8.9				
[a]S2-A^6:O2P···GLN^{154}:HE21	516.0	25.8	2.0 ± 0.2	157.4 ± 10.0				
[a]S2-A^6:N7···ASN^{199}:HD22	380.0	22.4	2.1 ± 0.1	162.9 ± 8.8				
[a]S2-A^6:O1P···GLN^{154}:HN	596.0	198.7	1.8 ± 0.1	163.4 ± 8.5				
[b]S2-C^6:O1P···GLN^{154}:HE21					306.0	14.0	1.9 ± 0.2	162.3 ± 9.4
[a]S2-A^6:H61···ASN^{199}:OD1	352.0	27.1	2.0 ± 0.2	161.3 ± 9.1				
[a]S2-A^6:O2P···TYR^{156}:HH	598.0	299.0	1.8 ± 0.1	164.5 ± 7.8				
S2-T^7/G^7:O2P···ARG^{191}:HH21	334.0	23.9	1.9 ± 0.2	156.0 ± 12.2	390.0	24.4	1.8 ± 0.1	161.6 ± 9.0
[a]S2-T^7:O5′···ARG^{191}:HH21	194.0	11.4	2.2 ± 0.1	155.3 ± 10.0				
[a]S2-T^7:O2P···ARG^{191}:HH11	352.0	25.1	1. 8 ± 0.2	160.3 ± 10.0				

Only the cases with H-bonding with total residence time longer than 40.0 ps have been listed here.
[a] Those are present in the wild type system only
[b] Those are present only in the mutant system.

Table 4
Comparison of the statistics of the residence time of the water-bridges between wild type and mutant system

The atoms which are connected through water-bridge	Wild type			Mutant		
	Total residence time (ps)	Average life time (ps)	No. of water molecules involved in water-bridge	Total residence time (ps)	Average life time (ps)	No. of water molecules involved in water-bridge
S1-T^4:O2P···Arg^{179}:HH21	180.00	8.18	6	48.80	6.10	7
S1-T^4:O1P···Arg^{176}:HN	72.40	9.05	3	77.60	8.62	3
S1-G^5:O2P···Arg^{201}:HH21	98.00	8.17	1	90.00	6.92	1
S1-G^5:O2P···Arg^{179}:HH12	310.40	10.35	1	288.40	10.68	1
S1-G5:O2P···Leu^{174}:O	248.80	9.95	1	77.20	9.65	1
S1-G^5:O2P···Arg^{201}HH11	66.80	6.68	2	193.60	9.68	1
S1-G^5:O2P···Gln^{198}:OE1				53.60	8.93	1
S1-G^5:O5′···Arg^{201}:HH11	57.20	6.36	2			
S1-T^5:O2P···Tyr^{173}:OH				52.80	7.54	4
[a]S1-C^7:H42···Gln^{198}:HE21	214.00	9.73	1			
S1-C^8:O2P···Lys^{205}:HZ3				54.00	6.75	3
S1-C^8:O2P···Lys^{205}:HZ1				55.60	9.27	3
S2-T^3:O1P···Glu^{207}:OE2	68.80	17.20	1	82.80	9.20	6
S2-T^3:O1P···Glu^{207}:OE1	78.00	7.09	4	396.80	9.92	11
S2-G^4:O2P···Glu^{207}:OE1				152.00	11.69	4
S2:C^5:O1P···Tyr^{156}:HN				295.20	9.23	1
S2-C^5:O1P···Trp^{196}:HE1				307.60	9.05	1
S2-C^6:O2P···Tyr^{156}:HH				474.80	18.26	1
S2-C^6:O2P···Arg^{191}:HH22				247.20	11.24	1
[a]S2-T^7:O4···Asn^{199}:OD1	134.00	9.57	1			
S2-G^8:O2P···Arg^{191}:HH22	153.20	13.93	6			

Water-bridges with the total residence time longer than 40 ps have been listed here.
[a] Water-bridges are specific type and absent in the mutant system.

with a shortening of total residence time while there are some others that have improved in qualities. Moreover, in the mutant few additional water-bridges have been developed and some of them are considerably strong. Surprisingly, the total number of water-bridges in the mutant is more compared to that in the wild type complex. Thus, we could not correlate the water-bridging phenomenon to the observed loss of binding affinity in any meaningful way.

3.8. Conformational changes due to mutation

A detailed examination of the local structure and orientation of the important side chains over the stable parts of the trajectories indicate that the mutation has caused large-scale local conformational changes to some of the residues of Hoxc8. Obviously, this has an impact on the interaction patterns. Major changes have been observed for the residues Arg^{191}, Ile^{195}, Gln^{198} and Met^{202}. In the wild type complex Arg^{191} was interacting strongly with S2-T^7 through two strong H-bonds (see Table 3). Due to mutation S2-T^7 → S2-G^7, a change from pyrimidine to purine has occurred which has pushed away the side chain of Arg^{191} with a change in the dihedral angle (CG-CD-NE-CZ) from 271° to 152° (Fig. 6a). As a consequence, a H-bond between this side chain and the backbone of S2-G^7 is lost. An orientational change of the side chain of Ile^{195} has been observed to occur due to a change in the dihedral angle θ (CG2-CB-CG1-CD) from 298° to 64°. The time evolution of θ shows that the side chain flipped from one conformation with $\theta = 298°$ to another with $\theta = 64°$ and remained there for the rest of the simulation (Fig. 6b). Gln^{198} plays an important role in the recognition process due to its interaction with S2-G^8 of the DNA in the wild type complex mainly through electrostatic interaction [7]. It is observed that due to the mutation, the distance between the atoms Gln^{198}:NE2 and S2-G^8:O6 has

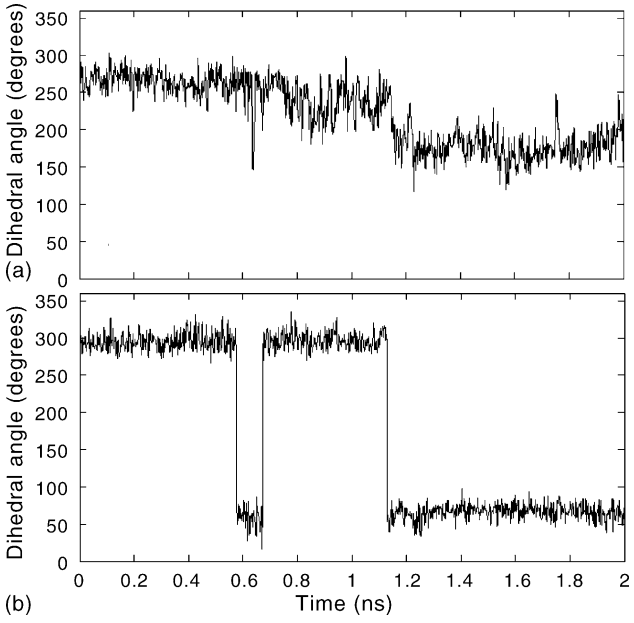

Fig. 6. The time evolutions of the dihedral angles (a) CG-CD-NE-CZ of the residue Arg^{191} and (b) CG2-CB-CG1-CD of the residue Ile^{195} are shown. For clarity data are shown only at the interval of 2.0 ps.

increased from 3.7 Å in the wild type to 5.6 Å in the mutant causing a reduction in the specific interaction (Fig. 4b). In another case, the conformation of the side chain of the residue Met^{202} has also been altered considerably due to a change in the dihedral angle (CB-CG-SD-CE) form 291° to 176°. It is interesting to point out that all these residues are in the close vicinity of the site of mutation.

3.9. Comparison of the global structures of the wild type and mutant DNA duplexes

The base-sequence dependent structural modulation in the DNA is often considered as a source of biomolecular recognition involving DNA [30–32]. Thus, in the present case, there is a possibility that the mutation may alter the groove size of the DNA in the binding site and prevents the Hoxc8 binding. In order to check such a possibility we have compared the minor groove widths of the wild type and mutant DNA duplexes. Fig. 7 dictates that the minor groove of the 3′-side of the mutation region has been widened slightly while at the other positions of the DNA duplexes the widths appear similar. It is believed that the AT-rich regions are characterized by a narrower minor groove compared to that of GC-rich regions [33]. The observed widening of the minor groove in the mutation region due to the mutation TAAT → GCCG is thus consistent with this fact. However, compared to the groove widths in the wild type DNA the widening is not very significant and considering the fact that the molecular binding occurs under dynamical conditions, the observed small modulations of the DNA grooves due to the mutation is quite unlikely to prevent Hoxc8 binding of the DNA completely. Significant sequence dependences are generally observed for longer sequence stretches while in the present case a change of sequence occurred for a short stretch of four base pairs only. Moreover, the perfect base-pair complementarities between the two strands of the mutant DNA rule out the possibility of any destabilization of the mutant DNA, rather the GC rich substitution stabilizes the double helix further. It may be noticed in Fig. 7 that there is significant change in the minor groove width in the region of the duplex containing the nucleotides numbered 3–7 of the strand S1 of the Hoxc8 binding wild type DNA and this region of the DNA has the maximum sequence dissimilarity (Fig. 1) compared to the template sequence.

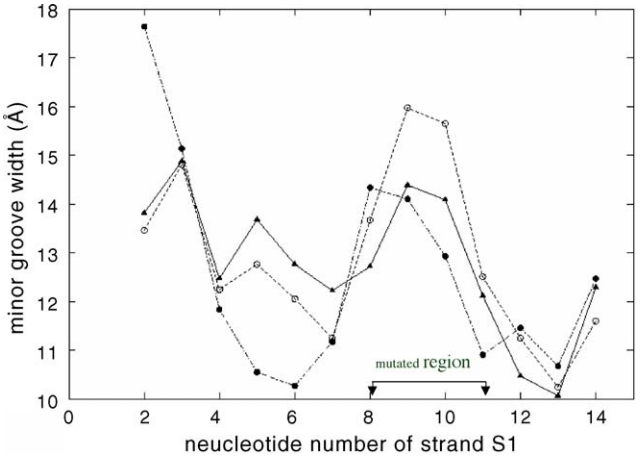

Fig. 7. Minor groove width profile for the template (●), wild type (▲) and mutant (○) DNA duplexes are shown here. The numbering of the nucleotides is according to strand S1. The mutated region (S1–8 to S1–11) are shown as labeled.

4. Discussions

The objective of this work was to use the methods of molecular modeling and molecular dynamics simulation to assess whether a specific mutation in the DNA would destabilize a known DNA–protein complex and also to gain insight in understanding the molecular mechanism of the loss of binding affinity of the Hoxc8–DNA complex due to a mutation (TAAT → GCCG). We have employed the techniques of molecular modeling and MD simulation to probe the changes in the interaction pattern that would be induced into the complex if within the already formed wild type complex the mutation could be introduced.

Our approach is based on the concept that, in a solution, formation of molecular complexes may be viewed as diffusion controlled reaction process where the molecular components (DNA and Hoxc8 in the present case) approach each other and interact through the non-bonded interactions to form a complex. The residence time to that state is determined by the strength of the interaction between them at a given temperature. When the interaction is strong enough such that they remain into the bound state for sufficiently long time, we consider that a stable complex has been formed. In the present case, a Hoxc8–DNA (mutant) complex similar to the wild type Hoxc8–DNA complex could be formed but only for a very short time period and then dissociate into free DNA and Hoxc8 protein due to weak interaction between them.

In the present work, we are only interested in the details of the altered atomistic interactions and structure induced by the mutation and thus causing the observed differences in binding affinity. No attempt has been made to compare the absolute binding affinities. Estimating the binding affinities in the wild type and mutant cases requires additional MD simulations of the free wild type and mutant DNA in water.

The present work demonstrated that the mutation did not introduced significant changes in the groove widths but caused considerable changes in the interaction pattern between the DNA and Hoxc8 at the interface. These alterations were caused by local structural changes and loss of functional groups due to mutation.

5. Concluding remarks

Present investigation demonstrated the capability of the method in assessing the potential of the destabilizing effects of a DNA–protein complex due to mutations in the DNA based on the changes in the enthalpy part of the DNA protein interactions. No significant change in DNA groove was observed.

It also provides significant insight about the molecular basis of the effective loss of binding affinity in the case of the mutant DNA. Modeling the mutant DNA–Hoxc8 complex on to the wild type complex model indicated that the reduction in affinity of the mutant DNA towards Hoxc8 was due to: (i) direct consequences of the mutation and (ii) indirect effects on the common parts induced by the mutation.

Both the specific and non-specific interactions of the DNA with Hoxc8 became considerably unfavorable in the hypothetical mutant compared to that in the wild type complex.

The number of inter molecular H-bonds was reduced significantly in the hypothetical mutant compared to the wild type complex that was consistent with the observed loss of binding affinity of the mutant DNA.

It is found that the loss of the capabilities of forming several base-specific H-bonds (S2-A^6:H61···ASN^{199}:OD1 and S2-A^6:N7···ASN^{199}:HD22) between the DNA bases and Hoxc8 as a direct consequence of the mutation reduced the specificity as well as the stability of the mutated DNA–Hoxc8 complex. As the indirect consequences of the mutation some of the strong non-specific H-bonds (for example, S2-A^6:O2P···TYR^{156}:HH, S2-A^6:O2P···GLN^{154}:HE21) were also lost in the mutant form of the complex.

All these changes in the DNA–Hoxc8 interaction are found to be correlated to the changes in the conformation of the side chains of several major residues of Hoxc8 in the close proximity of the DNA binding site.

References

[1] M. Eriksson, L. Nilsson, Structure, thermodynamics and cooperativity of glucocorticoid receptor DNA–DNA-binding domain in complex with different response elements: molecular dynamics and free energy perturbation study, J. Mol. Biol. 253 (1995) 453–472.

[2] J.W. Essex, D.L. Severance, J. Tirado-Rives, W.L. Jorgensen, Monte Carlo simulations for proteins: binding affinities for trypsinbenzamidine complexes via free energy perturbations, J. Phys. Chem. 101 (1997) 9663–9669.

[3] S. Sen, L. Nilsson, Free energy calculations and molecular dynamics simulations of wild-type and variants of the DNA–*Eco*RI complex, Biophys. J. 77 (1999) 1801–1810.

[4] C.M. Reyes, P.A. Kollman, Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change, J. Mol. Biol. 297 (2000) 1145–1158.

[5] X. Shi, X. Yang, D. Chen, Z. Chang, X. Cao, Smad1 interacts with homeobox DNA-binding proteins in bone morphogenic protein signaling, J. Biol. Chem. 274 (1999) 13711–13717.

[6] E. Fraenkel, C.O. Pabo, Comparison of X-ray and NMR structures for the Antennapedia homeodomain–DNA complex, Nat. Struct. Biol. 5 (1998) 692–697.

[7] S. Roy, S. Sen, Homology modeling and MD simulation study of Hoxc8–DNA complex in water: role of context bases outside TAAT stretch, J. Biomol. Struct. Dyn. 22 (2005) 707–718.

[8] S. Sen, L. Nilsson, Molecular dynamics of duplex systems involving PNA: structural and dynamical consequences of the nucleic acid backbone, J. Am. Chem. Soc. 120 (1998) 619–631.

[9] C.S. Tung, Structural study of homeodomain protein–DNA complexes using a homology modeling approach, J. Biomol. Struct. Dyn. 17 (1999) 347–354.

[10] N.C. Seeman, J.M. Rosenberg, A. Rich, Sequence-specific recognition of double helical nucleic acids by proteins, Proc. Natl. Acad. Sci. U.S.A. 73 (1976) 804–808.

[11] Y. Mandel-Gutfreund, O. Schueler, H. Margalit, Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles, J. Mol. Biol. 253 (1995) 370–382.

[12] J.W. Schwabe, The role of water in protein–DNA interactions, Curr. Opin. Struct. Biol. 7 (1997) 126–134.

[13] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minmimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[14] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.

[15] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327–341.

[16] W.F. van Gunsteren, H.J.C. Berendsen, Algorithms for macromolecular dynamics and constraint dynamics, Mol. Phys. 34 (1977) 1311–1327.

[17] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.

[18] C.L. Brooks III, M. Karplus, Deformable stochastic boundaries in molecular dynamics, J. Chem. Phys. 79 (1983) 6312–6325.

[19] T.E. Cheatham III, J.L. Miller, T. Fox, T.A. Darden, P.A. Kollman, Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA and proteins, J. Am. Chem. Soc. 117 (1995) 4193–4194.

[20] T.A. Daren, D.M. York, L.G. Pedersen, Particle mesh Ewald: an $N\cdot\log(N)$ method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.

[21] U. Essman, L. Perela, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8592.

[22] P.J. Steinbach, B.R. Brooks, New spherical-cutoff methods for long-range forces in macromolecular simulation, J. Comput. Chem. 15 (1994) 667–683.

[23] P. Auffinger, D.L. Beveridge, A simple test for evaluating the truncation effects in simulations of systems involving charged groups, Chem. Phys. Lett. 234 (1995) 413–415.

[24] J. Norberg, L. Nilsson, On the truncation of long-range electrostatic interactions in DNA, Biophys. J. 79 (2000) 1537–1553.

[25] R.W. Hockney, The potential calculation and some applications, Methods Comput. Comput. Phys. 9 (1970) 135–211.

[26] N. Foloppe, A.D. MacKerell Jr., All-atom empirical force field for nucleic acids: 1) parameter optimization based on small molecule and condensed phase macromolecular target data, J. Comput. Chem. 21 (2000) 86–104.

[27] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327–341.

[28] J. Jenin, Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition, Struct. Fold Des. 7 (1999) 277–279.

[29] S. Jones, P. van Heyningen, H.M. Berman, J.M. Thornton, Protein–DNA interactions: a structural analysis, J. Mol. Biol. 287 (1999) 877–896.

[30] W. Saenger, Principles of Nucleic Acid Structure, Springer Verlag, New York, 1984.

[31] V.P. Chuprina, A.A. Lipanov, O.Y. Fedoroff, S. Kim, A. Kintanar, B.R. Reid, Sequence effects on local DNA topology, Proc. Natl. Acad. Sci. U.S.A. 88 (1991) 9087–9091.

[32] E. Gavathiotis, G.J. Sharman, M.S. Searle, Sequence-dependent variation in DNA minor groove width dictates orientational preference of Hoechst 33258 in A-tract recognition: solution NMR structure of the 2:1 complex with d (CTTTTGCAAAAG)(2), Nucleic Acids Res. 28 (2000) 728–735.

[33] D.G. Alexeev, A.A. Lipanov, I.Y. Skuratovskii, Poly(dA)-poly(dT) is a B-type double helix with a distinctively narrow minor groove, Nature 325 (1987) 821–823.