Eur Biophys J (2012) 41:329–340
DOI 10.1007/s00249-012-0790-z

ORIGINAL PAPER

EBSA
Biophysics in Europe

Study of intermolecular contacts in the proline-rich homeodomain (PRH)–DNA complex using molecular dynamics simulations

Seifollah Jalili · Leila Karami

Received: 22 September 2011 / Revised: 3 January 2012 / Accepted: 16 January 2012 / Published online: 4 February 2012

Abstract The proline-rich homeodomain (PRH)–DNA complex consists of a protein with 60 residues and a 13-base-pair DNA. The PRH protein is a transcription factor that plays a key role in the regulation of gene expression. PRH is a significant member of the Q50 class of homeodomain proteins. The homeodomain section of PRH is essential for binding to DNA and mediates sequence-specific DNA binding. Three 20-ns molecular dynamics (MD) simulations (free protein, free DNA and protein–DNA complex) in explicit solvent water were performed to elucidate the intermolecular contacts in the PRH–DNA complex and the role of dynamics of water molecules forming water-mediated contacts. The simulation provides a detailed explanation of the trajectory of hydration water molecules. The simulations show that some water molecules in the protein–DNA interface exchange with bulk waters. The simulation identifies that most of the contacts consisted of direct interactions between the protein and DNA including specific and non-specific contacts, but several water-mediated polar contacts were also observed. The specific interaction between Gln50 and C18 and water-mediated hydrogen bond between Gln50 and T7 were found to be present during almost the entire time of the simulation. These results show good consistency with experimental and previous computational studies. Structural properties such as root-mean-square deviations (RMSD), root-mean-square fluctuations (RMSF) and secondary structure were also analyzed as a function of time. Analyses of the trajectories showed that the dynamic fluctuations of both the protein and the DNA were lowered by the complex formation.

Keywords Molecular dynamics simulations · Proline-rich homeodomain (PRH) · Protein–DNA binding · Specific and non-specific interactions · Hydrogen-bond interactions · Interfacial waters

S. Jalili (✉) · L. Karami
Department of Chemistry, K. N. Toosi University of Technology, P.O. Box 15875-4416, Tehran, Iran
e-mail: sjalili@kntu.ac.ir

S. Jalili
Computational Physical Sciences Research Laboratory, Department of Nano-Science, Institute for Studies in Theoretical Physics and Mathematics (IPM), P.O. Box 19395-5531, Tehran, Iran

Introduction

In biological systems, the binding of protein to DNA plays an important role in basic processes of living cells such as growth, cell division, differentiation, maturation and gene expression. Biological molecular recognition occurs through the interplay of non-covalent interactions, including direct and indirect (water-mediated) hydrogen bonds, van der Waals and electrostatic interactions. Proteins must be able to discriminate closely related DNA sequence and bind with high affinity to the proper DNA sequence.

Comprehension of the protein–DNA interaction details and binding mechanism is essential to understand the basic processes mentioned above. Otwinowski et al. (1988) first suggested a major function for water molecules in promoting macromolecular association in the trp repressor–DNA complex. Based on recent structural study about protein–protein and protein–DNA (Janin 1999; Schwabe 1997) recognition sites, water molecules exist in abundance at the interface. Due to the polarity of the surface described by the presence of phosphate groups on the DNA backbone and by charged groups on protein side-chains, water

molecules can play an important role in mediating protein–DNA interactions. Since X-ray crystallography, neutron diffraction and nuclear magnetic resonance (NMR) studies can only provide some information about macromolecular hydration and also are not able to provide a dynamical description, accurate molecular dynamics (MD) simulation procedures have been used to provide a wealth of knowledge on protein–DNA interaction and the dynamics of water molecules at atomic level and at the appropriate time scale. In recent years, many scientists have used the MD simulation method to investigate the molecular basis of protein–DNA interactions (Sen and Nilsson 1999; Reyes and Kollman 1999; Tsui et al. 2000).

The proline-rich homeodomain (PRH) [also known as haematopoietically expressed homeobox (Hex)] protein is a transcription factor that plays a key role in the regulation of gene expression in all eukaryotes (Martinez-Barbera et al. 2000; Crompton et al. 1992). PRH was first detected in avian haematopoietic and liver cells (Newman et al. 1997) and was then found to be conserved in humans, *Xenopus*, mice and rats (Tanaka et al. 1999). PRH protein is essential for the control of cell differentiation and cell proliferation (Martinez-Barbera et al. 2000; Guiral et al. 2001). PRH is involved in many processes such as embryonic development including embryonic patterning and the formation of many vital organs such as head, forebrain, thyroid, heart and liver (Swingler et al. 2004; Foley and Mercola 2005). Furthermore, PRH functions as a regulator of haematopoiesis in the adult (Guo et al. 2003; Jayaraman et al. 2000). PRH is a DNA binding protein that can activate and repress the transcription of its target genes using various mechanisms (Pellizzari et al. 2000; Kasamatsu et al. 2004). The PRH protein consists of three main domains: a proline-rich N-terminal domain (amino acids 1–136 in human PRH), a central homeodomain that is essential for binding to DNA and mediates sequence-specific DNA binding (amino acids 137–196 in human PRH) and an acidic C-terminal domain with unknown function (amino acids 197–270 in human PRH) (Crompton et al. 1992). Based on X-ray crystallographic and NMR spectroscopic studies on several homeodomain-containing proteins, central homeodomain of PRH is a helix–turn–helix-type DNA-binding domain with 60 amino acids that consists of an N-terminal arm and three α-helices that are separated by a short loop and short turn, respectively. The first and second helices lie parallel to each other and nearly perpendicular to the third helix. The sequence of homeodomain section (with 60 residues) of PRH protein is as follows: amino acids 10–22 = helix-1, amino acids 28–38 = helix-2, amino acids 42–58 = helix-3 (Crompton et al. 1992). The sequence-specific DNA binding is performed by insertion of third helix (also referred to as recognition helix) into the major groove and N-terminal arm into the adjacent minor groove of the DNA. Both N-terminal arm and recognition helix make sequence-specific contacts with base and backbone of DNA (Billeter 1996; Kissinger et al. 1990).

Previous experimental studies have shown that position 50 of the homeodomain located in the recognition helix plays a fundamental role in recognizing specific DNA–binding sites (Hanes and Brent 1991). This position is occupied most frequently by lysine and glutamine, but in some cases it can be occupied by cysteine, alanine or isoleucine. The homeodomain interacts with its conserved 5′-TAATNN-3′ binding sequence (Laughon 1991). Homeodomains with glutamine at position 50 (Q50), such as PRH, Antennapedia homeodomain and Engrailed homeodomain, recognize TAATTA, TAATTG or TAATGG sites, while those with lysine at position 50 (K50), such as Bicoid homeodomain and Pitx2 homeodomain, recognize TAATCC or TAAGCT (Billeter 1996; Schier and Gehring 1993).

The lack of a three-dimensional (3D) structure for the PRH–DNA complex in the Protein Data Bank (PDB) on the one hand, and its importance in biological systems on the other, motivated us to model the PRH–DNA complex structure and subsequently examine its mechanism of binding using molecular dynamics simulations. Application of the MD simulation method provides insight into details that are concealed from most experimental techniques. In this study, explicit water molecular dynamics simulations were carried out to further understanding of intermolecular contacts and the role of water molecules in protein–DNA binding.

Duan and Nilsson (2002) conducted molecular dynamics simulations on several wild-type and mutant homeodomain–DNA complexes to investigate the role of residue 50 in homeodomain–DNA interaction. They attempted to answer questions about the structural and dynamic properties of different side-chains (lysine, glutamine, serine and cysteine) of residue 50 in DNA recognition. In this study, we report calculations similar to those performed but only for glutamine at residue 50 to achieve information not only regarding atomic detail of direct and indirect protein–DNA contacts but also about participation of interfacial water in the protein–DNA contacts. We first performed MD simulations of free protein and free DNA and then compared the results with those for the protein–DNA complex.

Computational details

The starting coordinates of the PRH protein were obtained from PDB (PDB ID: 2E1O). The starting coordinates of the PRH–DNA complex were built by superimposing the structure of Msx-1 homeodomain–DNA complexes (PDB ID, 1IG7) (Hovde et al. 2001) on the PRH structure using the Strap program (Gille 2006), based on 3D alignment and structural similarity at the recognition helix of PRH and

Msx-1 homeodomain. To create the starting DNA structure of 5′-CACTAATTGAAGG-3′, the mentioned sequence was made by the HyperChem (release 7, 2002) modelling program. This is the same sequence present in the Msx-1 homeodomain–DNA complex. The alignment of residues between PRH and Msx-1 homeodomain protein is indicated in Fig. 1. In this figure, the sequence of homeodomain section of PRH and Msx-1 protein as well as DNA sequence present in Msx-1-DNA complex are shown. All systems are in neutral pH 7.0. Both N- and C-terminal arms of protein are consistent with zwitterionic state ($NH3^+$ and COO^-, respectively). Three 20-ns MD simulations were carried out: (1) simulation of free protein, (2) simulation of free DNA and (3) simulation of the protein–DNA complex. All of the MD simulations were carried out with the GROMACS package (version 4.0.7, van der Spoel et al. 2005). The AMBER 03 force field (Duan et al. 2003) was applied for protein, DNA and protein–DNA complex. Each starting structure was placed in a cubic box. In all cases, the minimum distance from any solute atom to the edge of the box was 10 Å. The boxes were then filled with TIP3P water molecules (Jorgensen et al. 1983). For protein, DNA and complex systems, 6,157, 7,915 and 7,406 water molecules were needed, respectively. In all MD simulations, periodic boundary conditions were applied in all directions. Ten Cl^- (free protein), 24 Na^+ (free DNA) and 14 Na^+ (protein–DNA complex) ions were added to achieve electroneutrality by substituting individual water molecules. This also provides a more realistic environment. The initial velocities were taken from a Maxwell distribution at 310 K. The temperature was maintained by coupling to a reference temperature of 310 K with a Nosé–Hoover thermostat (Nosé 1984; Hoover 1985). To maintain the systems at constant pressure of 1 bar, a Parrinello–Rahman barostat (Parrinello and Rahman 1981) was applied. Coupling time of 0.1 and 1.0 ps was used for thermostat and barostat, respectively. The value of the isothermal compressibility was set to 4.5×10^{-5} bar^{-1}. To treat long-range electrostatic interactions, the particle mesh Ewald (PME) algorithm (Essman et al. 1995) was applied. The cutoff distance for van der Waals interactions was set to 10 Å. To constrain all bonds involving hydrogen atoms, the SHAKE algorithm (Ryckaert et al. 1977) was used. MD simulation was performed by the leapfrog algorithm (Hockney 1970) with time step of 2 fs. In all MD simulations, the following protocol was used: The energy minimization included 5,000 steps of steepest descent, followed by 5,000 steps of the conjugate gradient method; Afterward, 1 ns of equilibration molecular dynamics was performed in the *NVT* ensemble to relax the system; During the equilibration phase, harmonic restraints with force constants of 1,000 kJ mol^{-1} nm^{-2} were applied on the coordinates of the protein and DNA; Finally, a production run was performed for 20 ns under an *NPT* ensemble. The atomic coordinates were saved every 4 ps for analysis.

Results and discussion

In this study, RMSD, RMSF (as a measure of the structural properties) and secondary structure analysis with the DSSP algorithm were investigated. Using these physical properties, the stability of the PRH protein can be qualitatively compared during simulation time.

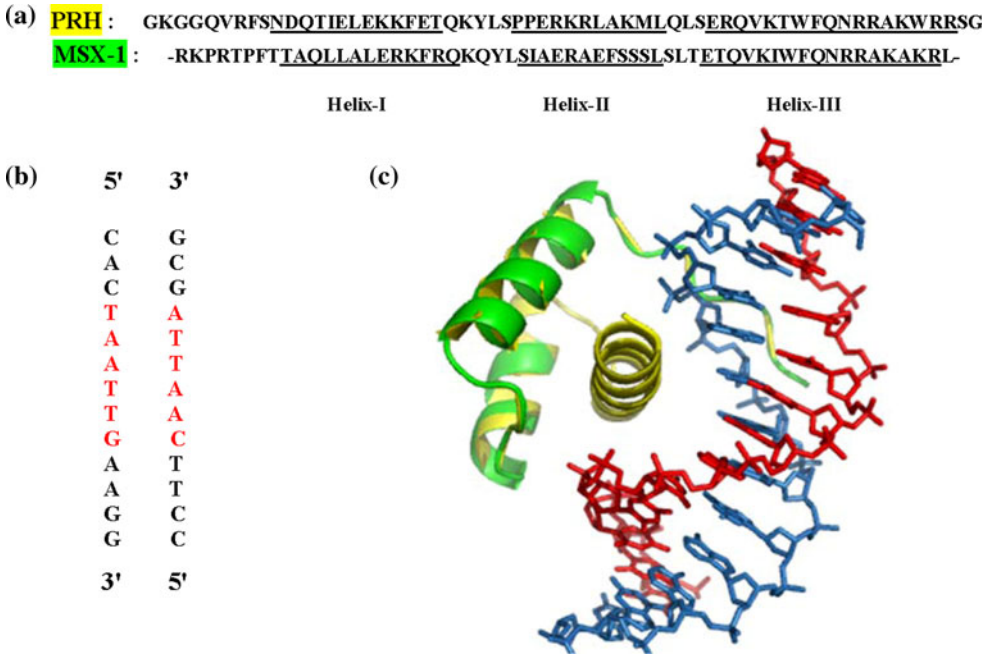

Fig. 1 A molecular model of the PRH–DNA complex. **a** The amino acid sequences of the PRH homeodomain and Msx-1 homeodomain were aligned according to their three-dimensional structures. Those residues corresponded to helix-1, 2 and 3, indicated by underlining, respectively; **b** The DNA sequence present in Msx-1–DNA complex used in PRH–DNA complex. Core sequence of DNA (TAATTG) is coloured *red*; **c** Structure of the two aligned homeodomain proteins PRH (*yellow*) and Msx-1 (*green*) bound to DNA

The time evolution of RMSD from the initial structures was calculated for three production run simulations (free protein, free DNA and protein–DNA complex). RMSDs of the protein Cα and DNA backbone atoms are plotted, respectively, in Fig. 2a (for free protein and protein in complex) and Fig. 2b (for free DNA and DNA in complex). Both free protein and protein in complex became stable after 7 ns, suggesting an equilibrated system. The RMSD value of the free protein fluctuated around 0.371 nm in free homeodomain and 0.203 nm in homeodomain in the protein–DNA complex. As shown in Fig. 2a, RMSD values for protein were reduced upon binding. DNA binding and consequent restriction of molecular motions caused decrease of RMSD value of protein in complex with respect to free protein. The RMSD values of free DNA (black line in Fig. 2b) remained stable around 0.347 nm during the entire simulation time, whereas the RMSD of the DNA in complex increased for nearly 5 ns, and fluctuated around 0.371 nm during the final 15 ns (grey line in Fig. 2b). The RMSD values for DNA were reduced upon binding, like protein. The reduction in RMSD for DNA was lower than that for protein. Due to restriction of molecular motion resulting from DNA binding, the DNA in the complex showed lower RMSD and greater stability than in free status.

Since distance deviations from the starting structure may not necessarily reflect the mobility of structural elements, another parameter, RMSF, was used to obtain information on flexibility. To identify flexible regions in the molecule and to simplify comparisons, RMSFs of the protein Cα and DNA backbone atoms are illustrated in Fig. 3. Figure 3a shows RMSF for free protein and protein in complex, and Fig. 3b shows RMSF for free DNA and DNA in complex. In case of protein, excluding the residues of N- and C-terminal arm (residues 1–9 and 57–60), the RMSF for the two simulations does not exceed 0.15 and 0.85 nm,

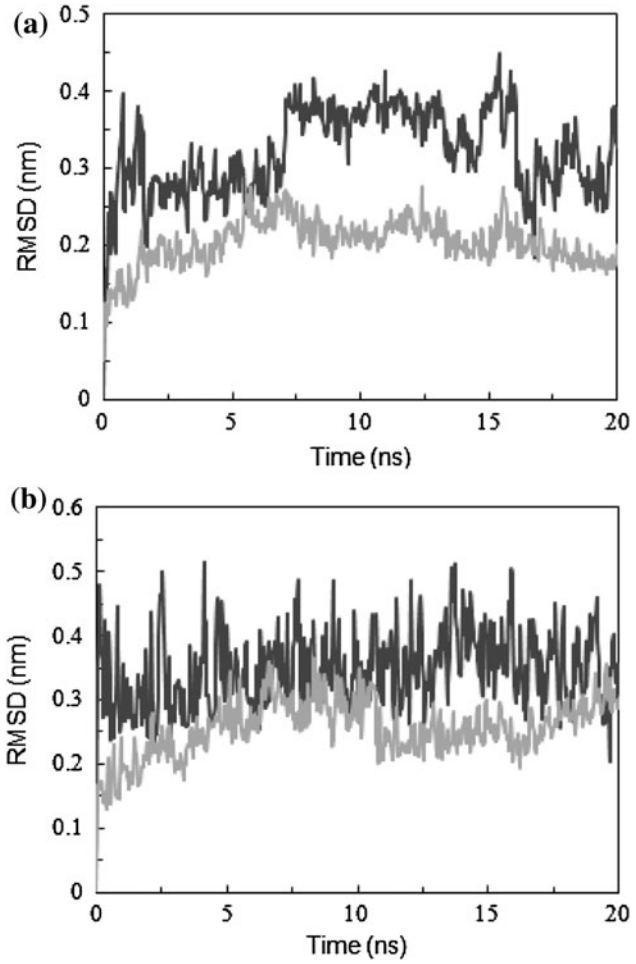

Fig. 2 Time evolution of RMSD measured from the corresponding starting structure: **a** *black line*, free protein; *grey line*, protein in complex (Cα atoms); **b** *black line*, free DNA; *grey line*, DNA in complex (DNA backbone)

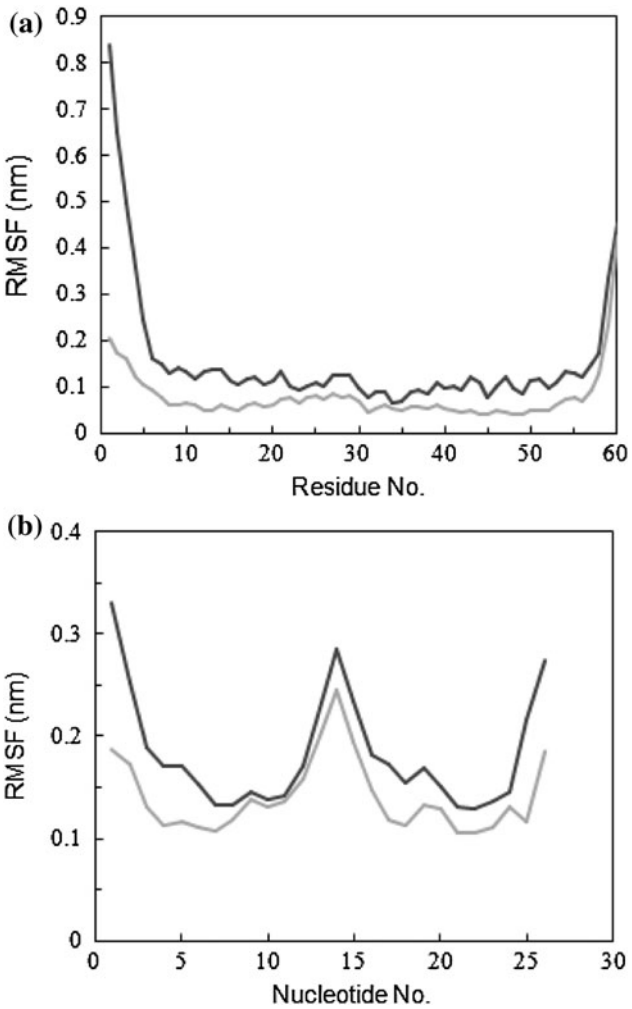

Fig. 3 Time evolution of the RMSF around the average MD structure: **a** *black line*, free protein; *grey line*, protein in complex (Cα atoms); **b** *black line*, free DNA; *grey line*, DNA in complex (DNA backbone)

respectively. For both free protein and protein in complex, the least fluctuating segments are third α-helix (residues 42–54), implying that helix-3 has low flexibility, being a rigid region in the two simulations. Since there are similar trends in free protein and protein in complex, it can be deduced that DNA binding does not greatly change the structural flexibility of homeodomain protein. Reduction of RMSF value of helix-3 upon DNA binding confirms existence of interaction between recognition helix and DNA. These explanations are in agreement with experimentally observed results: N- and C-terminal arm become ordered upon binding to DNA (Fraenkel et al. 1998; Qian et al. 1989). In the case of DNA, two ends of each strand not involved in protein binding show larger fluctuation for the two simulations (nucleotides 1–3 and 10–13 for α strand and 14–17 and 24–26 for β strand). As shown in Fig. 3b, protein binding site (major groove; nucleotides T4-G9 for sense strand and C18-A23 for antisense strand) are rigid. Theses nucleotides have RMSFs ranging from 0.108 to 0.139 nm and from 0.105 to 0.133 nm, respectively. Since major groove makes contacts with the protein, these results are also reasonable. Owing to restriction of molecular motions resulting from DNA binding, the protein and DNA in the complex show lower RMSF values than in free status.

Analysis of secondary structure was done with the DSSP (Kabsch and Sander 1983) program. Secondary structures of free protein and protein in complex as a function of time are depicted in Fig. 4. To distinguish between the secondary structures types, different colours are used. The overall secondary structure pattern of free and bound homeodomain is maintained during the 20-ns MD simulation, although there was slight change at some points as a function of time. These changes are often seen in N- and C-terminals and loop between helix-1 and helix-2. With a glance at Fig. 4, it can be understood that the major secondary structure of PRH in both trajectories is as α-helix and that residues 10–22, 28–38 and 42–58 keep their α-helicity throughout the 20-ns simulation. As shown in Fig. 4, in the free homeodomain, the first nine residues (N-terminal) are nearly unstructured and switch dynamically between bend, turn and coil conformations, while in bound homeodomain, N-terminal arm appears as coils and bends. This greater ordering of the N-terminal arm of the homeodomain upon DNA binding is in good accordance with experimental studies (Fraenkel et al. 1998; Qian et al. 1989) as well as computational studies of the family of homeodomain–DNA complexes (Zhao et al. 2006). In both free and bound homeodomain, residues 28–38 (loop segment) fluctuate between bend and coil conformations, but in bound homeodomain, the proportion of bend conformation increased with respect to free homeodomain. In free homeodomain, C-terminal segment (residues 58–60) has coil conformation, whereas in bound homeodomain at the first 3 ns, C-terminal segment is in turn conformation, whereas for the rest of the time, C-terminal segment adopts coil conformation.

Analysis of intermolecular contacts

In general, interactions between protein and DNA consist of hydrogen bonds, hydrophobic, electrostatic and water-

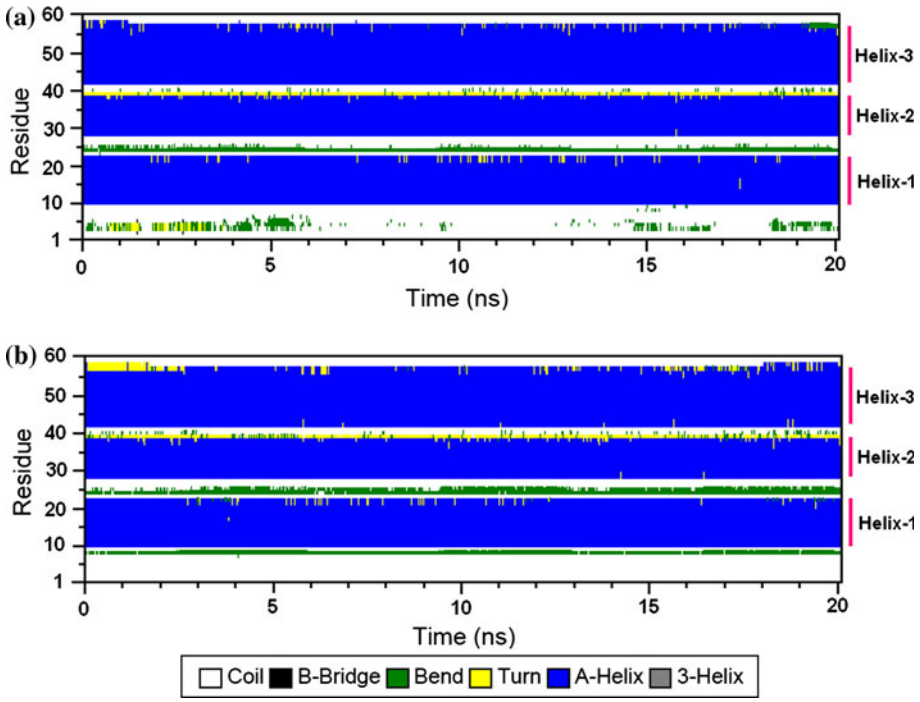

Fig. 4 Secondary structures as function of time for 20-ns simulation, at 310 K: **a** free homeodomain; **b** bound homeodomain

mediated interactions. Of these, we investigated direct and water-mediated hydrogen bonds in detail during the course of the trajectories. For analysis, we define a geometric criterion for a hydrogen bond D–$H\cdots A$, where A is acceptor and D is donor atom, and H is a hydrogen atom: an acceptor–donor pair is considered to form a hydrogen bond if the donor–acceptor distance (d_{DA}) is $\leq$3.5 Å and the donor–hydrogen–acceptor angle (α_{DHA}) is $\geq$135°. Interfacial water molecules were identified by applying a cutoff of 2.5 Å for the distance of water molecule to protein and DNA atoms simultaneously. Therefore, protein–DNA is considered to form a water-mediated contact when at least one water molecule simultaneously bridges through a hydrogen bond both the protein and DNA. In identifying water-mediated hydrogen bonds, we used the same distance and angle cutoffs. In biological systems, water molecules may have structural or functional roles (Wüthrich 1993). They may act as a hydrogen-bond linker, an effective element in recognition and a participant in the stability and specificity of DNA binding. The fact that water molecules are frequently observed experimentally at the interface between biomolecules such as protein–DNA interface suggests that water is necessary for biomolecular recognition. Water is a highly multipurpose component at the interface of biomolecular systems (Ladbury 1996; Tame et al. 1996). Water molecules are able to act as both a hydrogen-bond donor and acceptor. In this status, water molecules impose few steric constraints on bond formation, as taking part in multiple hydrogen bonds is possible for water molecules.

The trp repressor–DNA complex was the first for which the role of water in promoting macromolecular associations was suggested (Otwinowski et al. 1988). In this complex, there are no direct base-specific protein–DNA contacts and water-mediated polar contacts are responsible for the specificity. Since then, a number of studies have identified water molecules at protein–DNA interfaces (Billeter et al. 1996; Qian et al. 1993). In PRH–DNA complex, unlike trp repressor–DNA complex, there are both direct and water-mediated hydrogen bonds. These direct hydrogen bonds are of two types: amino acid–DNA base interactions (specific) and amino acid–DNA backbone interactions (non-specific).

Interfacial water molecules

We studied the time evolution of a number of water molecules during the entire simulation time. In this investigation, those water molecules appearing in interface for a total of at least 1 ns of the last 5 ns were selected. In Fig. 5, the time evolution of the mentioned water molecules during the entire simulation time is investigated. Based on the above-mentioned criterion, 17 water molecules were considered, of which 12 were long-life water molecules

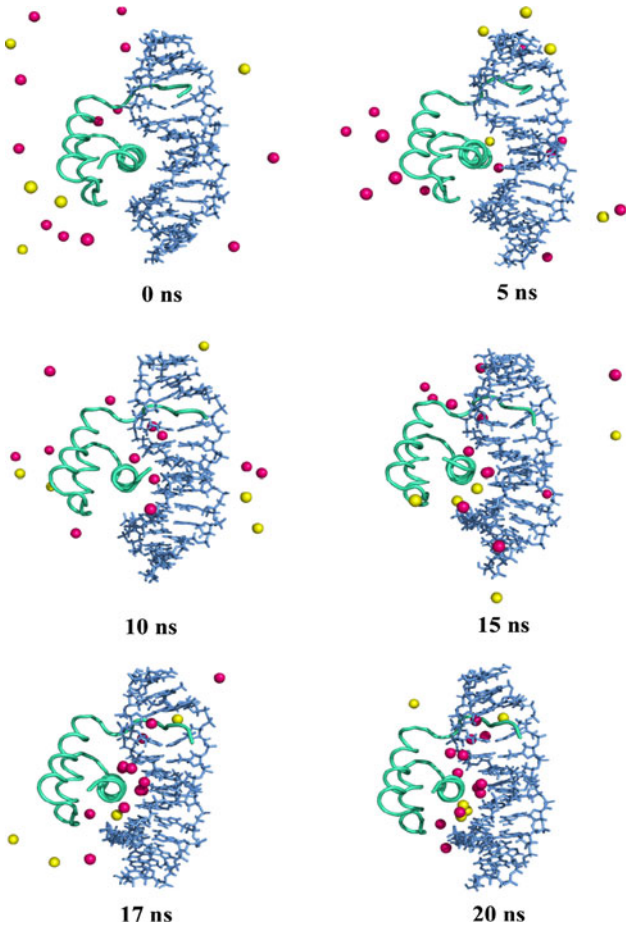

Fig. 5 Snapshots taken at 0, 5, 10, 15, 17 and 20 ns to illustrate the trajectory of some water molecules. Water molecules which are simultaneously closer than 2.5 Å to at least one protein atom and one DNA atom are shown as *balls*. Long-life water molecules and fast-exchanging water molecules are coloured in *pink* and *yellow*, respectively (see text for definitions). Protein (*green*) and DNA (*blue*) are shown in *cartoon* and *stick* representation, respectively

(depicted as pink balls in Fig. 5). It is clearly seen from Fig. 5 that these 17 water molecules are not in the interface at $t = 0$ and are placed randomly. Over time, especially after 15 ns, water molecules gradually neared protein residues or DNA bases and located around helix-3 and N-terminal; we can call them interfacial water molecules. At the end of simulation ($t = 20$), only some of these water molecules are in the interface. Water molecules lying around helix-3 and N-terminal result in the formation of water-mediated hydrogen bonds between protein and DNA, as discussed in more detail below. As mentioned above, the criterion for an interfacial water molecule is to be simultaneously closer than 2.5 Å to at least one atom on the protein and one atom on the DNA. In general, water molecules can be classified into two groups: those belonging to bulk solvent and those that interact with the surface of solutes, the latter being generally in exchange with bulk solvent (Otting et al. 1991; Denisov and Halle

1995). Despite the presence of water molecules in interface of protein and DNA, interestingly, we observed that some water molecules were relatively mobile and exchanged rapidly with bulk solvent. These fast-exchanging water molecules had short residence time (shown as yellow balls in Fig. 5). Using Fig. 5, in addition to hydration of protein–DNA interface, the orientation and location of the protein with respect to the DNA can be investigated during MD simulation. As in the starting structure of PRH–DNA complex, helix-3 and N-terminal lie in the major groove and adjacent minor groove of the DNA, respectively; during the simulation, N-terminal and helix-3 retain their location also. Figure 5 confirms this. Analysis of direct and water-mediated hydrogen bonds was done for the entire simulation time (20 ns), but since water molecules mostly lie in the interface between protein and DNA (specially, helix-3 and N-terminal) in the time window of 15–20 ns, we list only contacts populated >40% in the last 5 ns of the trajectory.

Direct protein–DNA hydrogen bonds

Table 1 summarizes direct hydrogen bonds extracted from the trajectory with their populations. In this table, contacts populated over 40% in the last 5 ns of the trajectory are listed. The stability of hydrogen bonds was measured by the percentage of their presence during the 15–20 ns trajectory of PRH–DNA simulation. As shown in Table 1, secondary structure elements of PRH, N-terminal (Gly1, Gly4, Val6, Arg7 and Phe8) and helix-3 (Gln44, Lys46, Gln50, Asn51, Arg53, Arg57 and Arg58) residues are involved in direct hydrogen bonding between PRH and DNA more than the rest. Protein–DNA direct hydrogen bonds include both specific interactions (between PRH and DNA base), being the most important determinants of the molecular recognition process in protein–DNA complexes, and non-specific interactions (between PRH and DNA sugar/phosphate backbone), being significant for the overall stability of the complex. Fifteen protein residues and 11 DNA bases participate in both direct and water-mediated hydrogen-bond interactions. As a complement to Table 1, some contacts are depicted as pictures or as time evaluation of bond distance in detail.

Asn51 participates in multiple contacts to A6 with high stability. Two hydrogen bonds are formed between the Asn51 OD1 atom and the H62 atom of A6 and also between Asn51 HD21 atom and the N7 atom of A6. Both hydrogen bonds mentioned are well consistent with experimental and computational studies (Zhao et al. 2006; Gruschus et al. 1997).

Arg57 also participates in multiple contacts to C18, but with lower occupancy than Asn51 (81%, 64% and 52% versus 98% and 99%). Arg57 forms three direct hydrogen bonds to C18 by donating two hydrogen of the NH1 and a hydrogen of NH2 to O1P, O5′ and O2P atoms of C18, respectively. The time courses of the latter two contacts during the 20-ns MD simulation are shown in Fig. 6. These two hydrogen bonds show motions with amplitudes of several angstroms in the first half of the trajectory. Over time, the distance between donor and acceptor involved in these two contacts gradually decreases and reaches hydrogen-bond scale. Studies by Neidle and Goodwin (1994) confirm the observed interaction between Arg57 and phosphate group of C18.

Lys46 is involved in multiple interactions with T16 and T17. Figure 7 illustrates these interactions. The HZ3 atom of Lys46 forms a direct hydrogen bond with O3′ atom of T16; at the same time, another proton of the NZ group of

Table 1 Protein–DNA direct hydrogen bonds observed in MD simulation

Protein residue		DNA base		Occupancy (%)
Gly1	O	A23	H61 (b)	99
Gly4	O	G24	H22 (b)	90
Gly4	H	G24	N2 (b)	67
Val6	H	G26	O1P (p)	46
Arg7	H	A5[a]	O1P (p)	45
Arg7	H	A5[a]	N3 (b)	46
Arg7	HH22	A5[a]	O4′ (s)	55
Phe8	H	A5[a]	O5′ (p)	43
Phe8	H	A5[a]	O1P (p)	95
Arg31	HH11	T16	O2P (p)	43
Arg31	HH12	T16	O1P (p)	42
Gln44	HE21	A6[a]	O1P (p)	96
Lys46	HZ3	T16	O3′ (p)	83
Lys46	HZ2	T17	O2P (p)	75
Gln50	OE1	C18	H41 (b)	98
Gln50	HE22	A19	N6 (b)	89
Asn51	HD21	A6[a]	N7 (b)	99
Asn51	OD1	A6[a]	H62 (b)	98
Arg53	HE	C18	O2P (p)	98
Arg53	HH22	C18	O1P (p)	90
Arg53	HE	C18	O1P (p)	45
Arg57	HH11	C18	O1P (p)	81
Arg57	HH12	C18	O5′ (p)	64
Arg57	HH21	C18	O2P (p)	52
Arg58	HH22	A20	O1P (p)	51
Arg58	HE	A20	O2P (p)	48
Arg58	HH21	A19	O3′ (p)	47

Only contacts populated >40% in the last 5 ns of the trajectory are listed. Residues are listed from N-terminal to helix-3. DNA base names are followed in parentheses by the letters b, p and s, indicating that the protein residue contacts base atoms, phosphate oxygens or sugar atoms, respectively

[a] DNA base belonging to α strand of DNA (1–13)

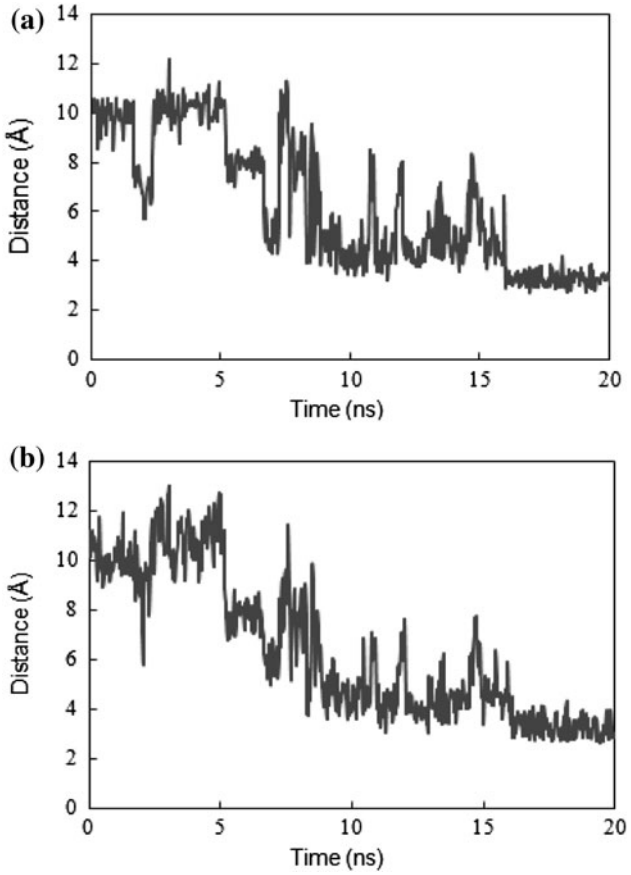

Fig. 6 Protein–DNA bond distances along the MD trajectory of the PRH and DNA complex. Distance measurements between **a** atom HH12 of Arg57 and atom O5′ of C18, and **b** atom HH21 of Arg57 and atom O2P of C18

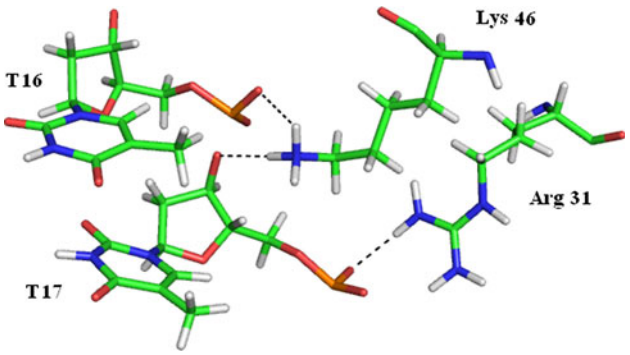

Fig. 7 Direct interactions between Lys46, Arg31 and T16, T17 on the DNA. Hydrogen bonds are depicted as *broken black lines*

Lys46 forms a new hydrogen bond with the phosphate oxygen atom (O2P) of T17. T16 base, in addition to participating in hydrogen bond with Lys46, forms two direct hydrogen bonds with HH11 and HH12 atoms of Arg31.

As examples of long-lasting base-specific contacts, we point out interactions such as Gly1–A23 and Gly4–G24 with 99% and 90% population, respectively. Arg53 participates in three non-specific contacts with C18 such that two of them persist for about 90% of the simulation time, but the third is not as stable as the rest. The observed interactions between Arg53 and phosphate group of C18 are confirmed by studies of Neidle and Goodwin (1994). Moreover, the H atom of Phe8 contacts with O5′ atom of A5, this interaction being weakened after approximately 8 ns. In the remaining part of the trajectory, the H atom of Phe8 contacts with O1P atom of A5.

Water-mediated protein–DNA hydrogen bonds

Since GROMACS cannot directly compute water-mediated hydrogen bonds, the following description is used: First, the g_hbond tool of GROMACS was applied to analyze hydrogen bonds between protein and water molecules; In the next step, hydrogen bonds between DNA and water molecules were analyzed; Finally, water molecules common to both analyses are considered as taking part in water-mediated hydrogen bonds. A perl script was used to obtain the occupancy (%) of hydrogen bonds. Table 2 provides complementary data for the role of water molecules in contact with protein and DNA simultaneously (water-mediated hydrogen bonds) during the trajectory. For water molecules which mediate hydrogen bonds between the protein and DNA, again a 40% population cutoff was used. Similar to direct hydrogen bonds, water-mediated hydrogen bonds were mostly formed in N-terminal and helix-3. Water-mediated hydrogen bonds are believed to be important for the specificity of biomolecular recognition and overall complex stability (Janin 1999; Schwabe 1997; Jones et al. 1999).

Gln50 accepts a hydrogen bond from N4 of C18 (98% population) and donates a hydrogen bond to N6 of A19

Table 2 Water-mediated hydrogen bonds observed in MD simulation

Protein residue·····Water·····DNA base	Occupancy (%)
Gly4 (H)·····OW–HW·····C3 (O3′)	91–78
Gly4 (H)·····OW–HW·····C4 (O4′)	91–91
Gln5 (H)·····OW–HW·····T4 (O)	32–45
Arg7 (HE)·····OW–HW·····T4 (O3′)	60–95
Arg7 (HE)·····OW–HW·····A5 (O4′)	45–96
Trp48 (H)·····OW–HW·····A6 (O2P)	90–99
Gln50 (HE21)·····OW–HW·····T7 (O4)	85–90
Asn51 (OD1)·····HW–OW·····A20 (H61)	63–92
Arg58 (HH22)·····OW–HW·····A20 (N7)	43–30

Only contacts populated >40% in the last 5 ns of the trajectory are listed. Residues are listed from N-terminal to helix-3. Atoms involving in hydrogen bond are in parentheses after protein residue and DNA base name

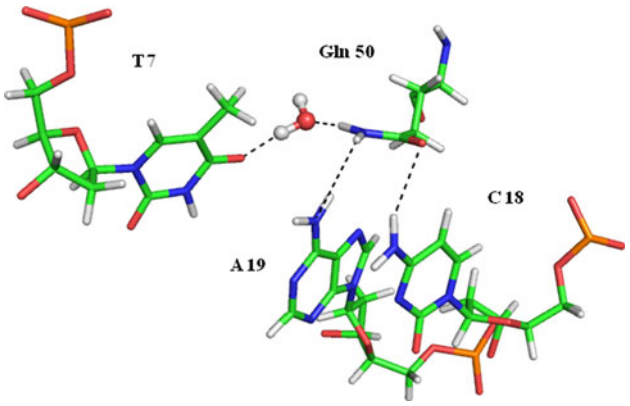

Fig. 8 Direct contacts between C18, A19 and Gln50 and water-mediated contacts between Gln50 and T7 nucleotide of the DNA

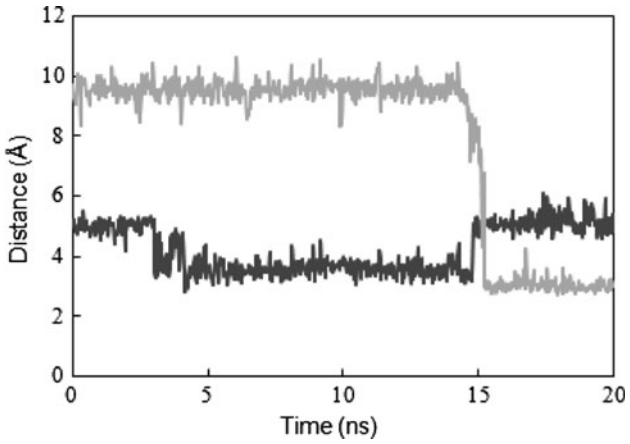

Fig. 9 Variation of intermolecular distance during the 20-ns MD simulation. Distance between Arg7 (HH22) and A5 (O4′) and distance between HE of Arg7 and oxygen atom of the water are shown by *black* and *grey lines*, respectively

(89% population) (Fig. 8). Furthermore, a water molecule fits between the NE2 group of Gln50 and the O4 atom of T7. This water-mediated interaction between Gln50 and T7 was observed in more than three-quarters of the snapshots examined. The switch in interaction of Gln50 from C18 to A19 in the course of the simulation identifies the dynamic character of this very important interaction. The observed network of direct and water-mediated interactions illustrates the role of Gln50 in protein–DNA binding affinity and specificity. The direct base-specific hydrogen bond between Gln50 and C18 and water-mediated hydrogen bond between Gln50 and T7 are consistent with literature (Zhao et al. 2006; Neidle and Goodwin 1994; Flader et al. 2003).

Arg7 and A5 were involved in both direct and water-mediated hydrogen bonds (Fig. 9). In the first 4 ns of trajectory, there was no connection between them. The direct hydrogen bond connecting H22 atom of Arg7 with O4′ atom of A5 was present in 55% of the trajectory (in the 4–15 ns interval). Arg7 lost its contacts with A5 from approximately 15 ns to the end of the simulation, as the distance between donor and acceptor increased (black line in Fig. 9). Closer inspection showed that this could be related to presence of water molecule between them. Investigation of the distance between HE atom of Arg7 and oxygen atom of the selected water molecule as a function of time clarifies this. As shown in Fig. 9, this water molecule lies between Arg7 and A5 at about 15 ns, as direct hydrogen bond is broken and this contact is replaced by water-mediated hydrogen bond. Selected water molecule forms water-mediated hydrogen bond by donating a hydrogen atom to O4′ of A5 and accepting HE atom of Arg7.

MD simulation identifies that a hydrogen bond between the H atom of Arg7 and N3 atom of A5 is observed in the initial 10 ns. This direct base-specific hydrogen bond is replaced by a hydrogen bond with the O1P atom of A5 (Fig. 10a). In addition to two direct hydrogen bonds mentioned already, water-mediated hydrogen bond between Arg7 and A5 is also shown in this figure.

Another residue involved in both direct and water-mediated hydrogen bonds is Arg58 (Fig. 10b). This residue

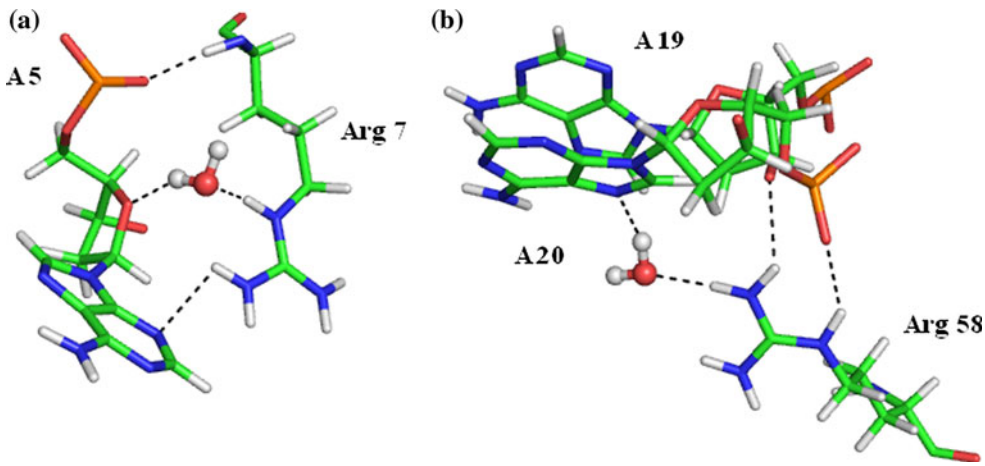

Fig. 10 Direct and water-mediated interactions: **a** Arg7–A5; **b** Arg58–A19 and Arg58–A20

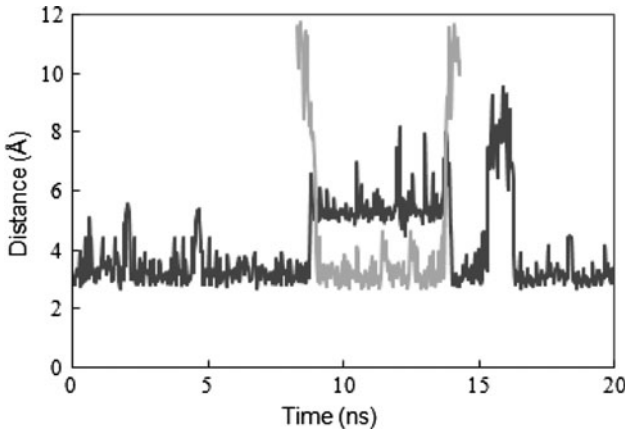

Fig. 11 Penetration of a water molecule between the nitrogen atom of Gly4 and the N2 atom of G24 is shown by the distance between these two atoms (*black line*) and the distance between the nitrogen atom of Gly4 and the oxygen of the penetrating water molecule (*grey lines*)

is in contact with A19 and A20, respectively (HH21 and HE of Arg58 with O3′ and O2P). Water-mediated hydrogen bond between Arg58 and A20 with a modest population is observed during simulation time. It seems that the reason for this low population is that water molecules mediating Arg58–A20 contact exchange with bulk solvent.

Time evaluation of the distance between H atom of Gly4 and N2 atom of G24, shown in Fig. 11, suggests water-mediated contacts in this interaction. The grey line in this figure represents the distances between oxygen atom of selected water molecule and H atom of Gly4. Until about 8.7 ns, there is a stable direct hydrogen bond between Gly4 and G24. In the 8.7–13.5 ns interval, the distance between donor and acceptor increased to 6 Å. This coincides with a decrement in distance between water oxygen and Gly4. In other words, this water molecule penetrates at about 9 ns and leaves this position at 13.3 ns. The Gly4–G24 interaction reappears in the 13.5–15.3 ns interval. At 15.3 ns, another water molecule forms a hydrogen bond bridge between Gly4 and G24, as the distance between Gly4 and G24 increased to 9 Å again. On the exit of this second water from the area at 16.5 ns, direct hydrogen bond between Gly4 and G24 is formed again and remains there for the rest of the simulation. In general, owing to the movements of the water molecules, contacts such as Gly4–G24 constantly break and form in short intervals.

Two additional bridging water molecules are observed in simulation. One mediates interactions between the H atom of Trp48 and the O2P atom of A6. A second water molecule in the vicinity mediates OD1 atom of Asn51 and the H61 atom of A20. The latter contact is in good agreement with data obtained from theoretical study of the Engrailed homeodomain–DNA complex (Zhao et al. 2006).

Error estimation relating to the distance between one donor atom and one acceptor atom participating in hydrogen bonding was calculated by block averaging methods (Hess 2002). In this method, each trajectory was divided into blocks, and the average distance of each block was calculated. The errors in distance were then estimated as one standard deviation of the block averages. For distances in Fig. 9, these errors range from 0.39 Å for the black line to 0.9 Å for the grey line. The estimated errors for donor–acceptor distances in two hydrogen bonds in Asn51–A6 are 0.023 and 0.011 Å. As expected, the block size increases until the standard error estimate becomes constant and approaches the true error.

Conclusions

Molecular dynamics simulations of the proline-rich homeodomain (PRH)–DNA complex have been carried out for 20 ns of total simulation time to investigate both the dynamical properties of water in the interface of the complex and hydrogen-bond patterns of the protein–DNA interaction.

At the beginning of the simulation, water molecules were placed randomly. During simulation, water molecules gradually neared protein residues or DNA bases and located around helix-3 (recognition helix) and N-terminal. The simulation shows the important role of water molecules in the formation of the binding interface. By analyzing direct and water-mediated hydrogen bonds, we found contacts that are important in PRH-DNA recognition and binding, such as Gln50–C18, Gln50–water–T7 and Asn51 and A6. The observed interactions are in good agreement with literature. Besides the analysis of hydrogen bonds between protein and DNA, structural properties such as RMSD, RMSF and secondary protein structure were evaluated. DNA binding and consequent restriction of molecular motions resulted in lower fluctuation of both the protein and DNA in the complex than in the free status, but the stabilization due to complex formation was larger for the protein than for the DNA. Both termini of the protein become ordered upon binding to DNA. The overall secondary structure pattern of free and bound homeodomain is maintained during the 20-ns MD simulation.

References

Billeter M (1996) Homeodomain-type DNA recognition. Prog Biophys Mol Biol 66:211–225

Billeter M, Güntert P, Luginbühl P, Wüthrich K (1996) Hydration and DNA recognition by homeodomains. Cell 85:1057–1065

Crompton MR, Bartlett TJ, MacGregor AD, Manfioletti G, Buratti E, Giancotti V, Goodwin GH (1992) Identification of a novel vertebrate homeobox gene expressed in haematopoietic cells. Nucleic Acids Res 20:5661–5667

Denisov VP, Halle B (1995) Protein hydration dynamics in aqueous solution: a comparison of bovine pancreatic trypsin inhibitor and ubiquitin by oxygen-17 spin relaxation dispersion. J Mol Biol 245:682–697

Duan J, Nilsson L (2002) The role of residue 50 and hydration water molecules in homeodomain DNA recognition. Eur Biophys J 31:306–316

Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24:1999–2012

Essman U, Perela L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593

Flader W, Wellenzohn B, Winger RH, Hallbrucker A, Mayer E, Liedl KR (2003) Stepwise induced fit in the pico- to nanosecond time scale governs the complexation of the even-skipped transcriptional repressor homeodomain to DNA. Biopolymers 68:139–149

Foley AC, Mercola M (2005) Heart induction by Wnt antagonists depends on the homeodomain transcription factor Hex. Gene Dev 19:387–396

Fraenkel E, Rould MA, Chamber KA, Pabo CO (1998) Engrailed homeodomain–DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other Engrailed structures. J Mol Biol 284:351–361

Gille C (2006) Structural interpretation of mutations and SNPs using STRAP-NT. Protein Sci 15:208–210

Gruschus JM, Tsa DHH, Wang L, Nirenberg M, Ferretti JA (1997) Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. Biochemistry 36:5372–5380

Guiral M, Bess K, Goodwin G, Jayaraman PS (2001) PRH represses transcription in hematopoietic cells by at least two independent mechanisms. J Biol Chem 276:2961–2970

Guo Y, Chan R, Ramsey H, Li W, Xie X, Shelley WC, Martinez-Barbera JP, Bort B, Zaret K, Yoder M (2003) The homeoprotein Hex is required for hemangioblast differentiation. Blood 102:2428–2435

Hanes SD, Brent R (1991) A genetic model for interaction of the homeodomain recognition helix with DNA. Science 251:426–430

Hess B (2002) Determining the shear viscosity of model liquids from molecular dynamics simulations. J Chem Phys 116:209–217

Hockney RW (1970) The potential calculation and some applications. Methods Comput Phys 9:135–211

Hoover WG (1985) Canonical dynamics: equilibrium phase-space distributions. Phys Rev A 31:1695–1697

Hovde S, Abate-Shen C, Geiger JH (2001) Crystal structure of the Msx-1 homeodomain/DNA complex. Biochemistry 40:12013–12021

HyperChem (TM) (2002) Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA

Janin J (1999) Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition. Structure 7:277–279

Jayaraman PS, Frampton J, Goodwin G (2000) The homeodomain protein PRH influences the differentiation of haematopoietic cells. Leukemia Res 24:1023–1031

Jones S, van Heyningen P, Berman HM, Thornton JM (1999) Protein–DNA interactions: a structural analysis. J Mol Biol 287:877–896

Jorgensen WL, Charndrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern-recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

Kasamatsu S, Sato A, Yamamoto T, Keng VW, Yoshida H, Yamazaki Y (2004) Identification of the transactivating region of the homeodomain protein, hex. J Biochem 135:217–223

Kissinger CR, Liu BS, Martin-Blanco E, Kornberg TB, Pabo CO (1990) Crystal structure of an Engrailed homeodomain–DNA complex at 2.8 Å resolution: a framework for understanding homeodomain–DNA interactions. Cell 63:579–590

Ladbury JE (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. Chem Biol 3:973–980

Laughon A (1991) DNA binding specificity of homeodomains. Biochemistry 30:11357–11367

Martinez-Barbera JP, Clements M, Thomas P, Rodriguez T, Meloy D, Kioussis D, Beddington RS (2000) The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. Development 127:2433–2445

Neidle S, Goodwin GH (1994) A homology-based molecular model of the proline-rich homeodomain protein, PRH, from haematopoietic cells. FEBS Lett 345:93–98

Newman CS, Chia F, Krieg PA (1997) The XHex homeobox gene is expressed during development of the vascular endothelium: overexpression leads to an increase in vascular endothelial cell number. Mech Develop 66:83–93

Nosé S (1984) A unified formulation of the constant temperature molecular dynamics methods. J Chem Phys 81:511–519

Otting G, Liepinsh E, Wüthrich K (1991) Protein hydration in aqueous solution. Science 254:974–980

Otwinowski Z, Schevitz RW, Zhang RG, Lawson CL, Joachimiak A, Marmorstein RQ, Luisi BF, Sigler PB (1988) Crystal structure of trp repressor/operator complex at atomic resolution. Nature 335:321–329

Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. J Appl Phys 52:7182–7190

Pellizzari L, D'Elia A, Rustighi A, Manfioletti G, Tell G, Damante G (2000) Expression and function of the homeodomain-containing protein Hex in thyroid cells. Nucleic Acids Res 28:2503–2511

Qian YQ, Billeter M, Otting G, Müller M, Gehring WJ, Wüthrich K (1989) The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell 59:573–580

Qian YQ, Otting G, Wüthrich K (1993) NMR detection of hydration water in the intermolecular interface of a protein-DNA complex. J Am Chem Soc 115:1189–1190

Reyes CM, Kollman PA (1999) Molecular dynamics study of U1A-RNA complexes. RNA 5:235–244

Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-Alkanes. J Comput Phys 23:327–341

Schier AF, Gehring WJ (1993) Functional specificity of the homeodomain protein fushi tarazu: the role of DNA-binding specificity in vivo. Proc Natl Acad Sci USA 90:1450–1454

Schwabe JWR (1997) The role of water in protein–DNA interactions. Curr Opin Struct Biol 7:126–134

Sen S, Nilsson L (1999) Structure, interaction, dynamics and solvent effects on the DNA-EcoRI complex in aqueous solution from molecular dynamics simulation. Biophys J 77:1782–1800

Swingler TE, Bess KL, Yao J, Stifani S, Jayaraman PS (2004) The proline-rich homeodomain protein recruits members of the groucho/transducin-like enhancer of split protein family to co-repress transcription in hematopoietic cells. J Biol Chem 279:34938–34947

Tame JRH, Sleigh SH, Wilkinson AJ, Ladbury JE (1996) The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. Nat Struct Biol 3:998–1001

Tanaka T, Inazu T, Yamada K, Myint Z, Keng VW, Inoue Y, Taniguchi N, Noguchi T (1999) cDNA cloning and expression of rat homeobox gene, Hex, and functional characterization of the protein. Biochem J 339:111–117

Tsui V, Radhakrishnan I, Wright PE, Case DA (2000) NMR and molecular dynamics studies of hydration of a zinc finger-DNA complex. J Mol Biol 302:1101–1117

van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible and free. J Comput Chem 26:1701–1718

Wüthrich K (1993) Hydration of biological macromolecules in solution: surface structure and molecular recognition, in DNA and chromosomes. Cold Spring Harbor Symp Quant Biol 58:149–157

Zhao X, Huang X, Sun C (2006) Molecular dynamics analysis of the Engrailed homeodomain–DNA recognition. J Struct Biol 155:426–437