Eur Biophys J
DOI 10.1007/s00249-013-0892-2

ORIGINAL PAPER

EBSA
Biophysics in Europe

Study of base pair mutations in proline-rich homeodomain (PRH)–DNA complexes using molecular dynamics

Seifollah Jalili · Leila Karami · Jeremy Schofield

Received: 4 August 2012 / Revised: 11 November 2012 / Accepted: 21 January 2013

Abstract Proline-rich homeodomain (PRH) is a regulatory protein controlling transcription and gene expression processes by binding to the specific sequence of DNA, especially to the sequence 5′-TAATNN-3′. The impact of base pair mutations on the binding between the PRH protein and DNA is investigated using molecular dynamics and free energy simulations to identify DNA sequences that form stable complexes with PRH. Three 20-ns molecular dynamics simulations (PRH–TAATTG, PRH–TAATTA and PRH–TAATGG complexes) in explicit solvent water were performed to investigate three complexes structurally. Structural analysis shows that the native TAATTG sequence forms a complex that is more stable than complexes with base pair mutations. It is also observed that upon mutation, the number and occupancy of the direct and water-mediated hydrogen bonds decrease. Free energy calculations performed with the thermodynamic integration method predict relative binding free energies of 0.64 and 2 kcal/mol for GC to AT and TA to GC mutations, respectively, suggesting that among the three DNA sequences, the PRH–TAATTG complex is more stable than the two mutated complexes. In addition, it is demonstrated that the stability of the PRH–TAATTA complex is greater than that of the PRH–TAATGG complex.

Keywords Proline-rich homeodomain (PRH) · Protein-DNA binding affinity · Hydrogen bond interactions · Molecular dynamics simulations · Free energy calculation · Thermodynamic integration

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

J. Schofield
Chemical Physics Theory Group, Department of Chemistry, University of Toronto, 80 Saint George Street, Toronto, ON M5S 3H6, Canada

Introduction

The binding of proteins to DNA is of fundamental importance in many basic processes of living cells, such as growth, cell division, differentiation and gene expression. The ability of a protein to distinguish particular DNA sequences and bind targeted sequences with high affinity is vital to such processes. The specificity of protein–DNA complexes is governed by a number of factors, including direct and indirect (water-mediated) hydrogen bonds, van der Waals and electrostatic interactions. Molecular modeling and molecular dynamics simulation (MD) methods are useful tools for investigating important physical features that determine the molecular basis of protein–DNA interactions at the atomic level and molecular time scales (Sen and Nilsson 1999; Reyes and Kollman 1999; Tsui et al. 2000).

The proline-rich homeodomain (PRH) 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) and is necessary for controlling cell differentiation and cell proliferation (Guiral et al. 2001). It is involved in many processes, such as the regulation of hematopoiesis in adults (Guo et al. 2003;

Published online: 06 February 2013

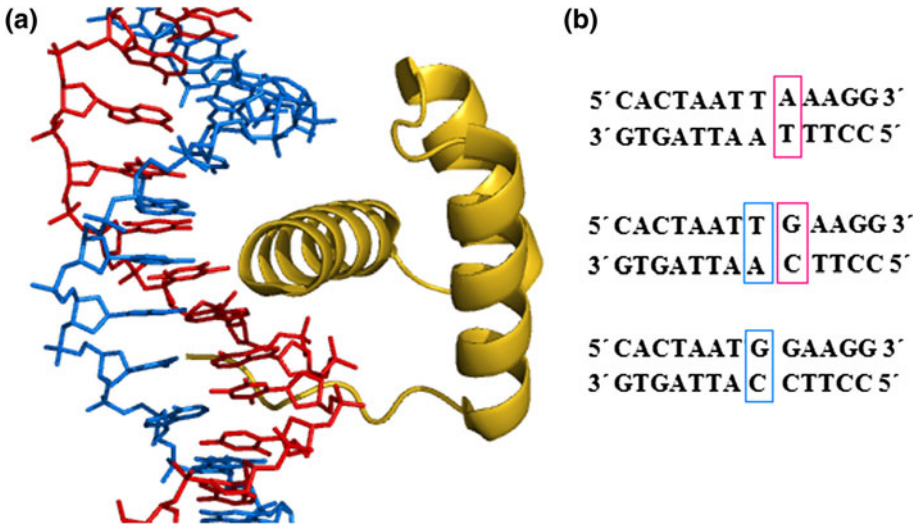

Fig. 1 **a** A schematic representation of the native PRH–DNA complex (PRH and DNA are shown in *ribbon* and *stick* representations, respectively). **b** The sequence of DNA in the native (*middle*), M1 complex (*top*) and M2 complex (*bottom*). The base pairs corresponding to the GC → AT and the TA → GC mutations are put in *pink* and *blue rectangular boxes*, respectively

Jayaraman et al. 2000) and the formation of many vital organs during embryonic development (Swingler et al. 2004; Foley and Mercola 2005). In addition, PRH is a DNA-binding protein with the capability of activating and repressing the transcription of its target genes (Pellizzari et al. 2000; Kasamatsu et al. 2004). The PRH protein contains a proline-rich N-terminal domain, a central homeodomain being essential for sequence-specific DNA binding, and a C-terminal domain (Crompton et al. 1992). The central homeodomain of PRH is a helix-turn-helix type DNA-binding domain with 60 amino acids (amino acids 10–22 = helix-1, amino acids 28–38 = helix-2, amino acids 42–58 = helix-3) (Crompton et al. 1992) that consists of an N-terminal arm and three α-helices that are separated by a short loop and short turn, respectively. The insertion of a third helix (recognition helix) into the major groove and the N-terminal arm into the adjacent minor groove of DNA results in sequence-specific DNA binding (Billeter 1996; Kissinger et al. 1990). Position 50 in the recognition helix, occupied by lysine or glutamine, plays a fundamental role in recognizing specific DNA-binding sites (Hanes and Brent 1991). Regardless of whether the residue at position 50 is glutamine or lysine, the homeodomain interacts with binding sequence 5′-TAATNN-3′ (Laughon 1991). Homeodomains with glutamine at position 50 (Q50), such as PRH, recognize TAATTG, TAATTA or TAATGG DNA sequences (Billeter 1996; Schier and Gehring 1993).

Because of the importance of PRH in biological systems, quantifying the relative affinity of PRH with the aforementioned DNA sequences is an essential step toward understanding the structural and dynamic changes that arise because of these mutations. In this article, the relative affinity of PRH for mutant sequences is computed by creating three models for the PRH–DNA complexes. In a previous study (Jalili and Karami 2012), intermolecular contacts in native PRH–DNA complexes with a TAATTG sequence were investigated. Here we compare the native PRH–DNA complex with complexes created by mutations of the GC and TA base pairs in TAATTG to AT and GC base pairs, respectively, to form TAATTA and TAATGG sequences. The structure of the native PRH–DNA complex and the DNA sequence existing in the native and mutated states is shown in Fig. 1. In this work, calculations were carried out in two phases: first, to investigate conformational and structural changes induced by the base pair mutations, molecular dynamics simulation of native complex and two mutated complexes was performed, and second, the change in the binding free energy relative to the native sequence was investigated for each mutation.

The effect of mutations of a specific residue on a protein or DNA structure has been previously studied experimentally and theoretically (Kwok et al. 2000). Hart and Nilsson (2008) performed MD simulations and free energy calculations on the transcription factor Ndt80 either in complex with the native DNA sequence or with mutant DNA with a switched central base pair, C5–G5′ to G5–C5′. One of their aims was to investigate the conformational changes around the CG5 base pair. Duan and Nilsson (2002) performed MD simulations on several wild-type and mutant homeodomain-DNA complexes to explore the role of residue 50 in homeodomain–DNA interactions and examine the role of different side chains (lysine, glutamine, serine and cysteine) of residue 50 in DNA recognition.

To the best of our knowledge, the effects of particular DNA sequences on the formation, structure and binding free energy of the PRH–DNA complex have not been investigated. Recently, Beierlein et al. (2011) studied the effect of base pair mutations on controller protein (C-protein)–DNA complex using the thermodynamic integration method and calculated the relative binding free energies of a series of mutants of a protein-binding DNA operator

sequence. In a previous study (Jalili and Karami 2012), MD simulations of the native complex (complex between PRH and DNA with TAATTG sequence) were carried out to elucidate the intermolecular contacts in the PRH–DNA complex and the role of water molecules forming water-mediated contacts. In that study, particular attention focused on the dynamic properties of water at the interface of the complex and hydrogen bond patterns of protein–DNA interaction.

In this work, explicit water MD simulations of the native and two mutated complexes were carried out to compare the conformational and structural differences of these complexes. In addition, the relative affinity of PRH protein to each of the DNA sequences was examined by calculating the binding free energy using the thermodynamic integration method.

Computational details

Molecular dynamic simulation

The initial coordinates of the native PRH–DNA complex were constructed by superimposing the structure of the Msx-1 homeodomain-DNA complex (PDB id, 1IG7) (Hovde et al. 2001) on a PRH structure (PDB id, 2E1O) based on 3D alignment and structural similarity at the recognition helix of PRH and Msx-1 homeodomain using the Strap program (Gille 2006). Two other complexes containing mutated DNA were obtained using a nucleotide base mutation script provided on the multiscale Modeling Tools for Structural Biology (MMTSB) website (http://blue11.bch.msu.edu/mmtsb/Main_Page). In this study, residues 4–60 of the homeodomain section of PRH were considered.

For convenience, the native PRH–DNA complex (PRH–TAATTG complex) and the two mutated complexes (PRH–TAATTA and PRH–TAATGG complexes) are henceforth called the native, M1 and M2 complexes, respectively. All simulations were carried out using the GROMACS 4.5.1 simulation package (van der Spoel et al. 2005) based on the AMBER 03 force field (Duan et al. 2003). Three 20-ns MD simulations were carried out for each of the native, M1 and M2 complexes. In all complexes, the zwitterionic state ($NH3^+$ and COO^-, respectively) was used for both the N- and C-terminal arms of the PRH protein. Each initial structure was placed in a periodic cubic box with a minimal distance of 10 Å from the solute to the walls of the simulation box. The boxes were then filled with appropriate amounts of TIP3P water molecules (Jorgensen et al. 1983). To achieve an electrically neutral system, individual water molecules were substituted by added ions (15 Na^+ for each three complexes). Conditions of constant pressure (1 bar) and temperature (310 K) were obtained by application of the Parrinello-Rahman barostat (Parrinello and Rahman 1981) and Nosé–Hoover thermostat (Nosé 1984; Hoover 1985). The time constants used for coupling the barostat and thermostat to the degrees of freedom of the system were 1.0 and 0.5 ps, respectively. The value of the isothermal compressibility was set to 4.5×10^{-5} bar^{-1}. The initial velocities were taken from a Maxwell–Boltzmann distribution at a temperature of 310 K. The cutoff distance for van-der-Waals interactions was set to 10 Å. The particle mesh Ewald (PME) (Essman et al. 1995) method was used to treat the long-range electrostatic interactions, while the SHAKE algorithm (Ryckaert et al. 1977) was used to constrain all bonds involving hydrogen atoms. The leapfrog algorithm (Hockney 1970) with a time step of 2 fs was utilized to integrate the equations of motion. The following protocol was used for all MD simulations: First, 1,000-step steepest descent minimization with position restraints of 1,000 kJ mol^{-1} nm^{-2} on the water molecules was performed, followed by the same minimization condition but with position restraints on the PRH–DNA complex. To relax the system, two 500-step unrestrained minimizations using the steepest descent and conjugate gradient methods were carried out. Then, an equilibration of 1 ns was performed under constant volume and temperature conditions. Finally, a production run was carried out for 20 ns under conditions of constant pressure and temperature. The atomic coordinates were saved every 4 ps for analysis.

Free energy calculations

The development of algorithms for the calculation of the free energy of biological systems using time series extracted from molecular dynamics trajectories has been a major area of research in computational biophysical chemistry and drug discovery (Kollman 1993; Deng and Roux 2009; Beveridge and Dicapua 1989; Sneddon et al. 1989; Gilson et al. 1997; Simonson et al. 2002; Boresch et al. 2003) in the past few years. The free energy perturbation (FEP) (Zwanzig 1954) and thermodynamic integration (TI) (Kirkwood 1935) methods are the most important and practical computational techniques for calculation of binding or solvation free energies. In the present study, the TI method was used to obtain the relative binding free energy of the two created mutations (GC → AT and TA → GC). In this approach, a coupling constant lambda (λ) is introduced, which can take values from 0 (for native state) to 1 (for mutated state). The free energy difference between native and mutated states can be evaluated as

$$\Delta G = \int_0^1 \left\langle \frac{\partial G(\lambda)}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \quad (1)$$

where $G(\lambda)$ is the potential energy of the system as a function of the coupling constant, and $\left\langle \frac{\partial G(\lambda)}{\partial \lambda} \right\rangle_\lambda$ is an ensemble average at a given λ.

To estimate the relative binding free energies ($\Delta\Delta G_{\text{bind}}$) of PRH to the native and mutated forms of DNA, the thermodynamic cycle shown in Fig. 2 was applied. The total difference in binding affinities of the PRH to the two DNAs, $\Delta\Delta G_{\text{bind}}$, is equal to ΔG_3–ΔG_4. Since Gibbs free energy is a state function, relative binding free energy of the system can be described as

$$\Delta\Delta G_{\text{bind}} = \Delta G_3 - \Delta G_4 = \Delta G_1 - \Delta G_2, \quad (2)$$

As the dual-topology (Gao et al. 1989) approach was used in this study, the perturbations between structurally diverse species (purine and pyrimidine bases) were feasible. In the case of atoms that are in only the native or mutated forms of DNA, each mutation was broken up into three steps for consistency (Tutorial 9, AMBER web site 2009) for those atoms that vanish or appear between forms being compared: In the first step of the process, the atomic partial charge on the atoms was removed (discharging step, ΔG_{dischg}). In the second step, the van-der-Waals transformation was performed using a softcore (Steinbrecher et al. 2007) potential (vdW-transformation step, ΔG_{vdw}), and finally the partial charges on new atoms were placed (charging step, ΔG_{chg}) in the third step. This three-step procedure was done for each mutation (GC → AT and TA → GC) in both the free state to yield ΔG_2 [free (discharging, vdW-transformation and charging)] and in the bound state to compute ΔG_1 [bound (discharging, vdW-transformation and charging)]. The MD simulations in each step of the three-step procedure were done independently for each lambda value. For each lambda value, the minimization, equilibration and production run were carried out. An initial structure for the DNA is required in these simulations. Here, the native DNA sequence was made utilizing the HyperChem modeling program (HyperChem Release 7, 2002). To form mutated free DNA and PRH–DNA complex structures, the nucleotide

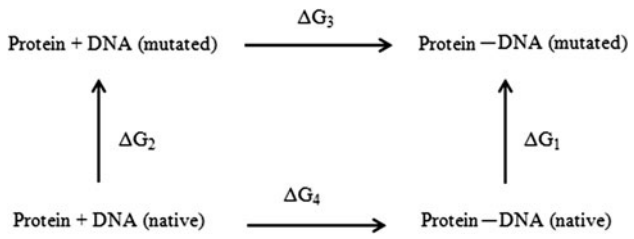

Fig. 2 Thermodynamic cycle for computing the relative binding free energy in the PRH–DNA complex

base mutation script provided on the MMTSB website was applied.

In the free energy calculations, both water-solvated free and bound states were considered. To solvate the system, the starting structure of a free or bound state was placed in a periodic cubic box containing TIP3P water molecules (Jorgensen et al. 1983). Sodium counter ions were added to the system to maintain electroneutrality. All free energy simulations were performed using sander.MPI the in AMBER 10 suite of programs (Case et al. 2008) with the ff99bsc0 (modified amber99) force field (Pérez et al. 2007). In this study, the lambda parameter was set to range from 0.1 to 0.9 in increments of 0.1. In each simulation, a cutoff distance of 10 Å was applied to smoothly truncate van-der-Waals interactions. The particle mesh Ewald (PME) (Essman et al. 1995) method was used to treat the long-range electrostatic interactions, and the SHAKE algorithm (Ryckaert et al. 1977) was used to constrain all bonds involving hydrogen atoms. The following protocol was used for the free energy simulations: First, 2,000-step steepest descent minimization with position restraints of 25 kcal mol^{-1} Å^{-2} on the water molecules was performed, followed by another 2,000 unconstrained minimization steps, switching from the steepest descent to the conjugate gradient method after 1,000 steps. Afterward, a 300-ps equilibration was performed with a Langevin-type thermostat with temperature 310 K and a collision frequency 5 ps^{-1}. In addition, a 200-ps equilibration with isotropic pressure scaling at 1 atm and a pressure relaxation time of 2 ps was carried out. Finally, a production run of 1,500 ps was performed in the NPT ensemble. In both the equilibration procedure and the production run, a time step of 2 fs was utilized. Energy outputs were saved every 0.5 ps. From the total simulation time of 2,000 ps, only data from the production run (500–2,000 ps) were chosen for analysis. Numerical integration of $\mathrm{d}G/\mathrm{d}\lambda$ was done using a standard trapezoid rule. The $\mathrm{d}G/\mathrm{d}\lambda$ values for $\lambda = 0$ and 1 were obtained by linearly extrapolation from the closest two values (Tutorial 9, AMBER web site 2009).

Results and discussion

Structural analysis

To compare the stability of the native and mutated complexes, root mean square deviations (RMSD) of the protein Cα and DNA backbone atoms relative to the starting structures are shown in the Fig. 3. The PRH in the native complex becomes stable after 2.5 ns, and the corresponding RMSD value fluctuates around 0.13 nm (blue line in Fig. 3a). PRH in the mutated complexes equilibrate much more slowly than the native complex does. First, the

RMSD of the PRH in the M1 complex fluctuates around 0.1 nm. After 10 ns, the RMSD increases sharply to nearly 0.17 nm and then remains stable to the end of the simulation (red line in Fig. 3a). The RMSD of the PRH in M2 complexes increases for nearly 5 ns and fluctuates around 0.2 nm for the rest of the simulation time (black line in Fig. 3a). The DNA in the native complex is stable in the sense that the RMSD values are small, whereas DNA in the M1 and M2 complexes deviates slightly from the reference structure. The RMSD value of the DNA fluctuates around 0.15 nm in a native complex and 0.2 and 0.3 nm in the M1 and M2 complexes, respectively (Fig. 3b). The trends of the RMSD for PRH and DNA in complex indicate that the native complex is the most stable structure.

In addition to RMSD values, we did compute (1) SD of the RMSD values, (2) statistical uncertainties of the RMSD values by time-block analysis (block averaging method) and (3) the proportion of the RMSD of the energy relative to the average of the energy values, too (data was not shown). Each of these three values indicates the convergence of the simulation clearly and correctly because both the standard deviation and statistical uncertainties of the RMSD values are of the order of magnitude of 10^{-2} (these values are one order less than the magnitude of the RMSD values), and the proportion of the RMSD of the energy relative to the average of the energy values is of the order of magnitude of 10^{-3}.

To obtain information on the flexibility of structural elements, root mean square fluctuations (RMSF) of the Cα atoms in the protein and the DNA backbone atoms were examined and are shown in Fig. 4. In the case of the PRH protein, the residues of the N- and C-terminal arms (residues 4–9 and 57–60) show the greatest fluctuations in all three complexes. Among these three complexes, PRH in the M2 complex has the greatest flexibility, while the native complex has the least flexibility (Fig. 4a). In the case of the DNA component of the complexes, the two ends of each strand not involved in protein binding show larger fluctuations (nucleotides 1–3 and 10–13 for α strand and 14–17 and 24–26 for β strand). Similar to the proteins in complex, the DNA in the native complex has the smallest RMSF value (Fig. 4b). From the trend obtained in RMSF values per residue, it can be deduced that the rigidity of the DNA in the native complex is greater than that in the mutated complexes.

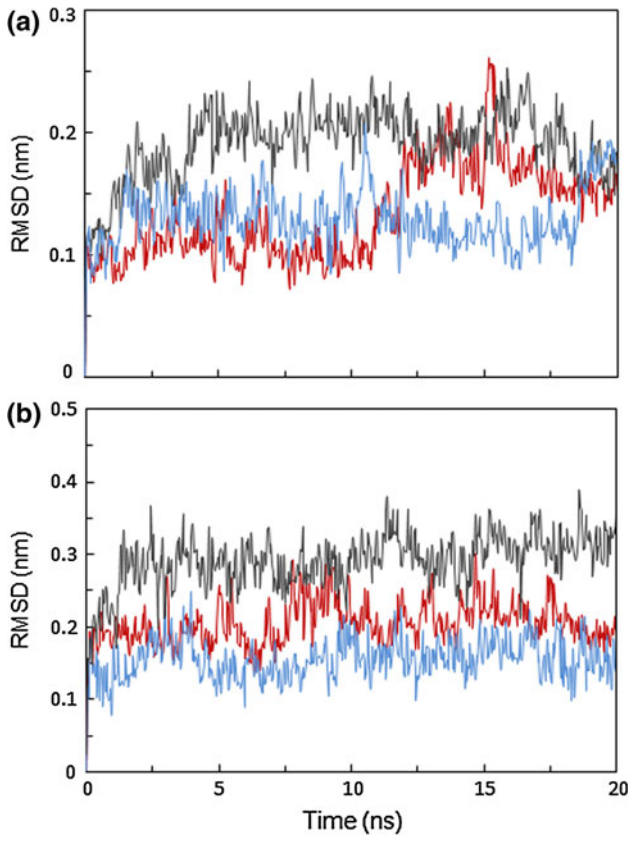

Fig. 3 Time evolution of the root mean square deviation (RMSD) measured from the corresponding starting structure: **a** protein in complex (Cα atoms); **b** DNA in complex (DNA backbone). The native, M1 and M2 complexes are in *blue*, *red* and *black*, respectively

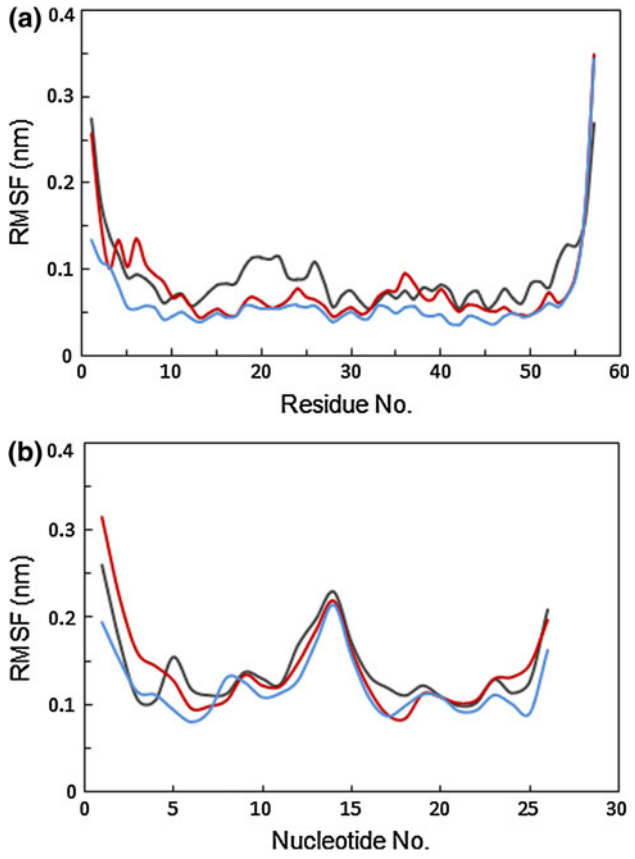

Fig. 4 Root mean square fluctuation (RMSF) around the average MD structure: **a** protein in complex (Cα atoms); **b** DNA in complex (DNA backbone). The native, M1 and M2 complexes are in *blue*, *red* and *black*, respectively

Analysis of secondary structure was done using the DSSP (Kabsch and Sander 1983) program. For this purpose, the average number of residues adopting α-Helix, 3_{10}-Helix, β-Sheet, Bridge, Turn, Bend and Random Coil secondary structural element was calculated for each of the three during 20-ns simulation time and plotted in Fig. 5a. In both the native and mutated complexes, most residues participate in α-Helix conformations, and the number of residues with the α-Helix conformation is the same for all three complexes (42 of 57 residues). The Turn conformation, which often connects two α-Helixes together, is the second most frequently observed secondary structural element, while very few of the residues are involved in β-Sheet and bridge structures. Since the α-Helix is the most common secondary structural element, the probability of participation in an α-Helix conformation for each residue of the PRH in the complexes was calculated and is illustrated in Fig. 5b. For convenience, the probability related to the native complex was multiplied by −1. It is clearly seen from Fig. 5b that the TA → GC mutation changes the probability of participation in an α-Helix conformation more than the GC → AT mutation, especially in the region of the first α-Helix. The observation that the major secondary structural element observed in the PRH protein in each of the three trajectories is an α-helix is in agreement with experimental and computational studies (Fraenkel et al. 1998; Qian et al. 1989) of the family of homeodomain–DNA complexes.

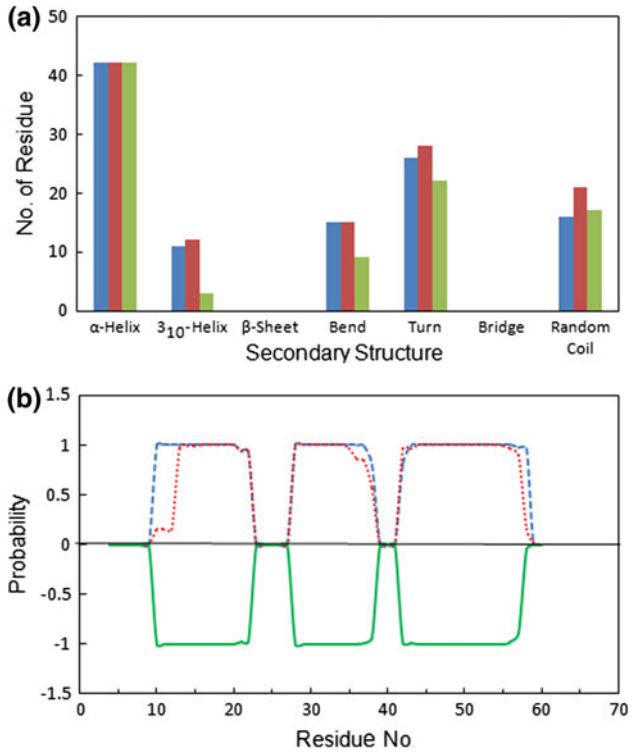

Fig. 5 Secondary structure for 20-ns simulation. **a** Number of residues adopting α-Helix, 3_{10}-Helix, β-Sheet, Bridge, Turn, Bend and Random Coil secondary structures. **b** The probability of α-Helix conformation for each residue of the PRH in complex. The PRH in the native, M1 and M2 complexes are in *green*, *blue* and *red*, respectively

Hydrogen bond analysis

To study the influence of mutations on direct and water-mediated hydrogen bonds between PRH and DNA, the percentage of hydrogen bond occupancy was calculated between donor and acceptor atoms, and the results are listed in Tables 1, 2 and 3. In the hydrogen bond analysis, a hydrogen bond was defined by a certain criterion [donor–acceptor distance (d_{DA}) ≤ 3.5 Å and the donor-hydrogen-acceptor angle (α_{DHA}) ≥ 135°]. A water-mediated hydrogen bond is defined to exist if a water molecule simultaneously bridges both the protein and DNA through two hydrogen bonds. In identifying the water-mediated hydrogen bonds, we have used the same distance and angle cutoffs. The stability of hydrogen bonds was measured by the percentage of their presence during the simulation. In these tables, contacts populated over 30 % in the 10–20-ns-long trajectory are listed. Upon mutation, the overall pattern of hydrogen bond interactions is maintained in the mutated complexes during the 20-ns MD simulation. Snapshots of the three complexes after simulation shown in Fig. 6 demonstrate that in each of the three complexes, the recognition helix was inserted into the major groove, while the N-terminal arm was inserted into the adjacent minor groove of the DNA. On the other hand, despite some changes in the hydrogen-bonding network, especially in the mutation region, there are no significant structural changes between native and mutated complexes. In a recent study (Jalili and Karami 2012), by analyzing the direct and water-mediated hydrogen bonds, contacts such as Gln50-C18, Gln50-A19, Gln50-water-T7 and Asn51-A6 were found to be important in PRH–DNA recognition in the native complex. Based on our new results in the present study, among the aforementioned interactions, there is only the Gln50-A19 interaction in the M1 complex and only the Gln50-C18 interaction in the M2 complex. Another important interaction between Asn51 and A6 exists in each of the three complexes but with lower occupancy in the mutated complexes (84, 73 and 57 % for the native, M1 and M2 complexes, respectively). These important hydrogen bond interactions in native and mutated complexes in final frame of the trajectory are depicted in Fig. 7. The comparison of the three complexes (as illustrated in Tables 1, 2 and 3) reveals that the number of hydrogen bonds (direct or water-mediated) that are formed between PRH and DNA in the native and the M1 complexes is nearly the same (with slightly more bonding in the native complex than in the M1 complex), but this number decreases in the M2 complex. The same trend was also observed in the hydrogen bond occupancy. For

Table 1 PRH–DNA direct (top) and water-mediated (bottom) hydrogen bonds of the native complex

Protein residue		DNA base		Occupancy (%)
Arg7	NH1	A6	O1P	40
Arg7	N	A5	O1P	45
Phe8	N	A5	O1P	41
Arg31	NH1	T17	O2P	49
Arg31	NH2	T17	O2P	35
Lys32	NZ	T16	O2P	38
Arg43	NH2	T7	O2P	47
Gln50	OE1	C18	N4	87
Gln50	OE1	A19	N6	85
Asn51	HD21	A6	N7	84
Asn51	ND2	T4	O5′	67
Arg53	NH2	C18	O1P	51
Arg57	NH2	A19	O2P	41
Arg58	NE	C3	O2P	73
Ag58	NH2	C3	O5′	56
Arg58	NH2	C3	O2P	49
Ser59	OG	C3	O2P	67
Protein residue···Water···DNA base				Occupancy (%)
Gly4 (N)···OW–HW···C3 (O3′)				90–73
Arg7 (NE)···OW–HW···T4 (O3′)				63–96
Trp48 (N)···OW–HW···A6 (O2P)				72–84
Gln50 (NE2)···OW–HW···T7 (O4)				86–82
Asn51 (OD1)···HW–OW···A20 (H61)				64–89

Only contacts populated over 30 % in the 10–20 ns of the trajectory are listed. For the criterion, see the methods

Table 2 PRH–DNA direct (top) and water-mediated (bottom) hydrogen bonds of the M1 complex

Protein residue		DNA base		Occupancy (%)
Gly4	N	G24	O1P	33
Arg7	N	A6	O1P	51
Tyr25	OH	T17	O2P	93
Arg31	NH1	T17	O2P	98
Arg43	NH2	T7	O2P	33
Gln44	NE2	A6	O1P	45
Gln50	NE2	A19	N7	65
Gln50	OE1	A19	N6	62
Asn51	ND2	A6	N7	68
Asn51	OD1	A6	N6	73
Arg53	NH1	T18	O2P	95
Arg53	NH2	T18	O2P	79
Lys55	NZ	T4	O2P	31
Arg57	NH1	A19	O2P	72
Protein residue···Water···DNA base				Occupancy (%)
Val6 (N)···OW–HW···C3 (O1P)				48–57
Gln44 (O)···HW–OW·····A6 (N6)				46–63
Gln50 (NE2)···OW–HW···A20 (N3)				59–42
Asn51 (OD1)···HW–OW···A20 (N6)				47–52

Only contacts populated over 30 % in the 10–20 ns of the trajectory are listed. For the criterion, see the methods

Table 3 PRH–DNA direct (top) and water-mediated (bottom) hydrogen bonds of the M2 complex

Protein residue		DNA base		Occupancy (%)
Gly4	N	T4	O2	45
Phe8	N	A5	O1P	75
Arg43	NH2	T7	O2P	51
Arg43	NE	T7	O2P	46
Gln44	NE2	A6	O1P	82
Gln50	OE1	C18	N4	40
Asn51	ND2	A6	N7	57
Arg53	NE	T17	O2P	52
Arg53	NH2	T17	O1P	48
Arg57	NE	C19	O2P	47
Arg57	NH1	C18	O2P	69
Protein residue···Water···DNA base				Occupancy (%)
Gly4 (N)···OW–HW···T4 (O3′)				31–39
Gln50 (NE2)···OW–HW···T7 (O4)				30–36
Asn51 (ND2)···OW–HW···A5 (O2P)				34–40

Only contacts populated over 30 % in the 10–20 ns of the trajectory are listed. For the criterion, see the methods

example, in the native complex, the Gln50-A19 and Gln50-C18 interactions, the most important interactions in PRH–DNA recognition, have occupancy of 85 and 87 %, respectively, whereas the occupancy of the Gln50-A19 interaction in the M1 complex and the Gln50-C18 interaction in the M2 complex are 65 and 40 %, respectively. The data obtained from the hydrogen bond analysis indicate that the reduction of the stability in the M2 complex, which is mainly the result of a reduction in the stability of hydrogen bonds, is greater than that in the M1 complex. These data are consistent with the binding free energy analysis discussed in the next section.

Free energy analysis

Free energy computations were performed using the TI method to explore the effects of the base pair mutations (GC → AT and TA → GC) on the relative binding affinities of the PRH for three DNA sequences. To ensure having well-equilibrated systems, free energy gradients $dG/d\lambda$ versus time were plotted for the $\lambda = 0.1, 0.5$ and 0.9 for both of simulations (Fig. 8 for GC → AT and Fig. 9 for TA → GC mutation). As can be seen from these figures, stable structures for both mutations were obtained after 500 ps of equilibration in both free and bound states. From

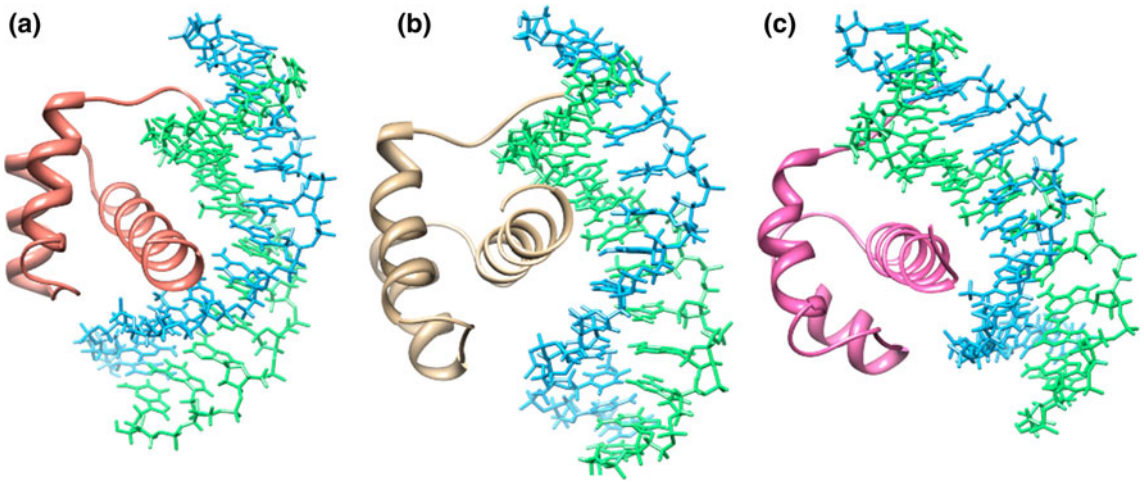

Fig. 6 Snapshots of the **a** native, **b** M1 and **c** M2 complexes related to the final frame of simulation (PRH and DNA are shown in *ribbon* and *stick* representations, respectively)

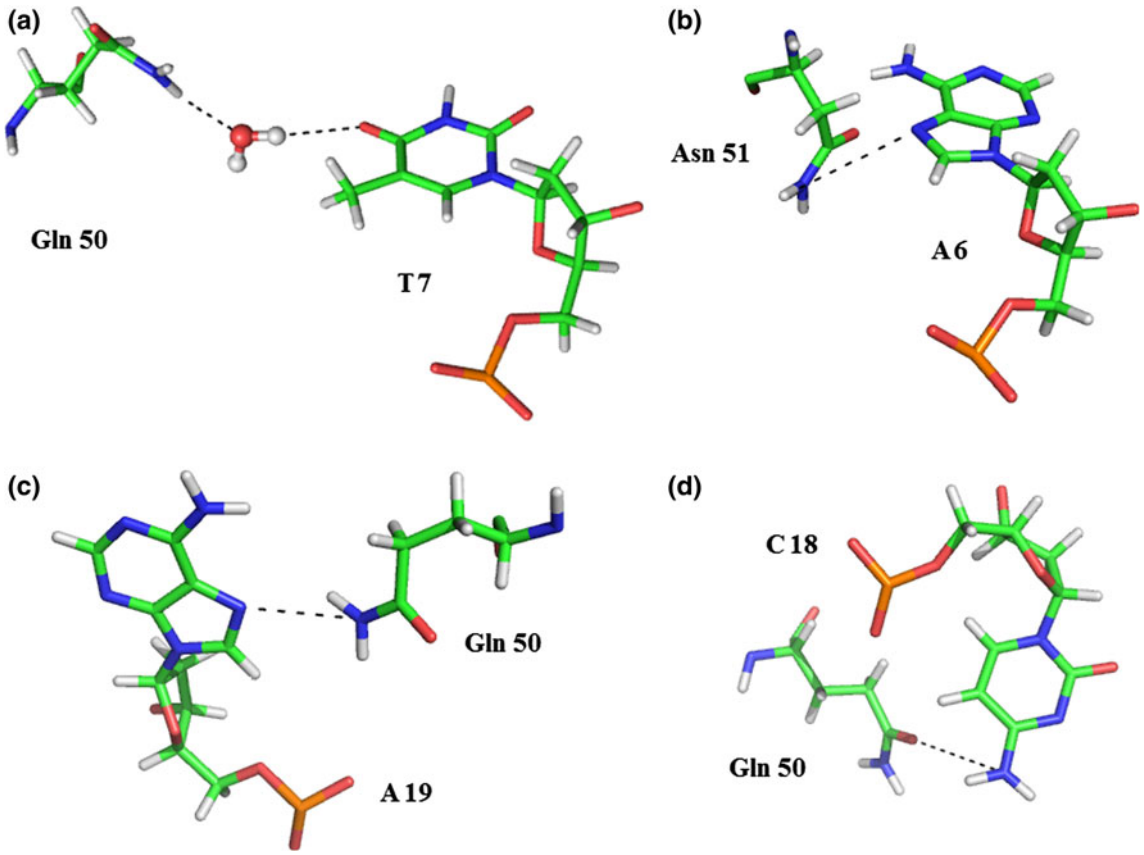

Fig. 7 Direct and indirect hydrogen bonds in the native and the mutated complexes, **a** Gln50-water-T7 and **b** Asn51-A6 in native complex, **c** Gln50-A19 in M1 complex and **d** Gln50-C18 in M2 complex

these figures, it is also obvious that, after the equilibration, 1,500 ps of simulation is sufficient to calculate the binding free energy because the cumulative average of the derivative $dG/d\lambda$ (shown in red, black and violet lines for $\lambda = 0.1$, 0.5 and 0.9, respectively) has converged to acceptable tolerances for all simulations. Another purpose of plotting $dG/d\lambda$ versus time is to observe trends in how $dG/d\lambda$ changes with increasing the lambda value (from 0.1 to 0.9) in each simulation. To this end, lambda values 0.1, 0.5 and 0.9 (in the beginning, middle and end of perturbation) were selected. In all cases (as shown in Figs. 8, 9), the free energy gradients $dG/d\lambda$ decrease with increasing

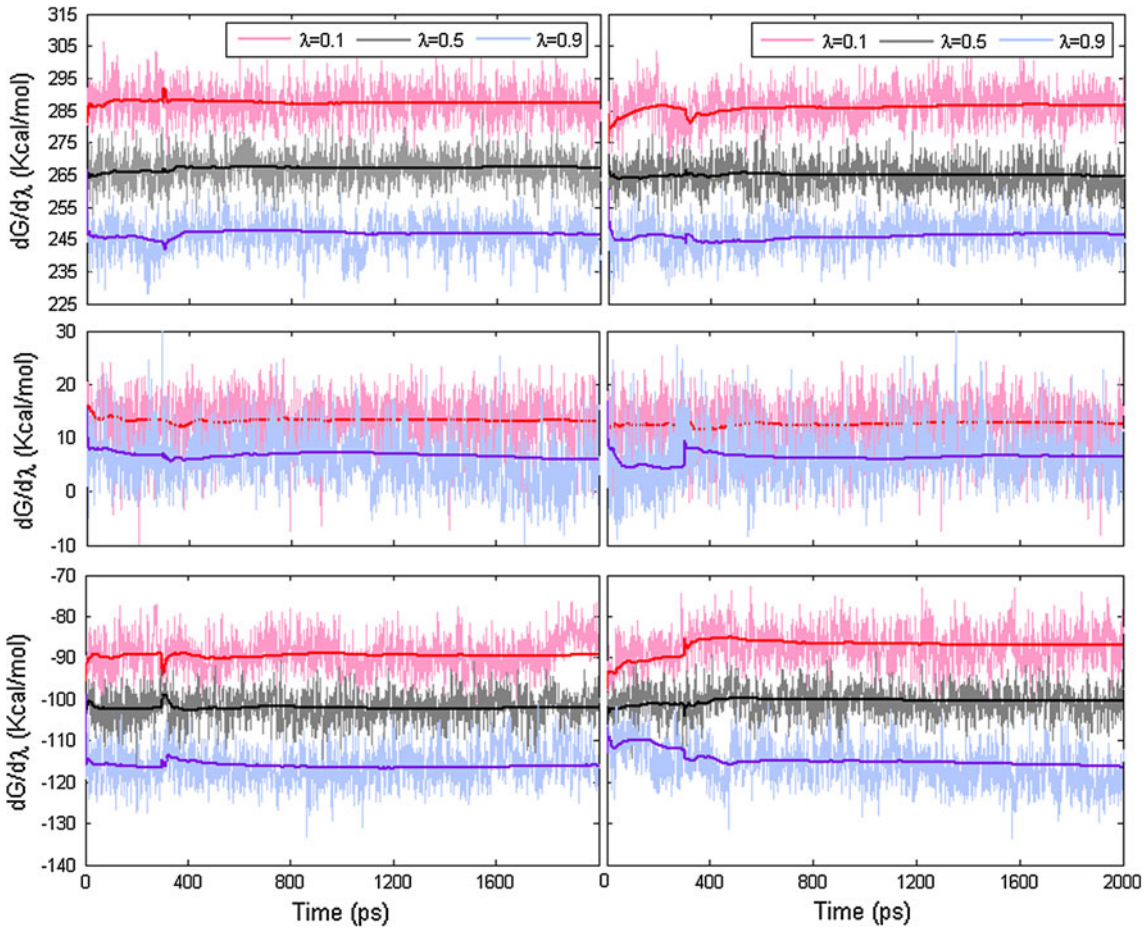

Fig. 8 Free energy gradients dG/dλ for discharging (*top*), vdW-transformation (*middle*) and charging steps (*bottom*) of the GC → AT mutation as a function of time for $\lambda = 0.1$, 0.5 and 0.9 for the discharging and charging steps and for $\lambda = 0.1$ and 0.9 in the vdW-transformation step (*left* and *right plots* are related to the free and bound states, respectively). The cumulative averages are shown in *red*, *black* and *violet lines* for $\lambda = 0.1$, 0.5 and 0.9, respectively

lambda value. Because the plot for $\lambda = 0.5$ in the vdW transformation step of both mutations (in Fig. 8) is close to the plot for $\lambda = 0.1$ and 0.9, for having better comparison and obtaining more obvious plots, only the free energy derivatives dG/dλ for $\lambda = 0.1$ and 0.9 are shown. For similar reasons, plots of the free energy derivatives are not shown for the vdW-transformation step of both mutations in Fig. 9.

The free energy derivatives dG/dλ are shown as a function of the coupling constants in Fig. 10 for the GC → AT mutation and in Fig. 11 for the TA → GC mutation. In each plot, the root mean square fluctuations (SD) of the free energy derivative dG/dλ are denoted by vertical lines (error bars). These data were used to numerically integrate the free energy derivatives to determine the relative binding affinities. In the case of the GC → AT mutation (see Fig. 10), the discharging and charging steps in the free and bound states show smoother behavior (with a standard deviation of approximately 5.5 and 4.6 kcal/mol for the discharging and charging steps, respectively) compared to the vdW-transformation step (with a standard deviation of approximately 6.1 kcal/mol), particularly for the bound state. In Fig. 11 for the case of the TA → GC mutation, the free energy derivative decreases smoothly with λ in the discharging and charging steps for both the free and bound states (with SD of approximately 5.9 and 6.4 kcal/mol, respectively), whereas the free energy derivative in the vdW-transformation step is not smooth for both the free and bound states (with SD of approximately 10.3 kcal/mol). Overall, it can be deduced that the fluctuations of the dG/dλ in the case of mutations from purines to pyrimidines and vice versa (such as TA → GC mutation) are greater than in the case of mutations from purines to purines or from pyrimidines to pyrimidines (such as the GC → AT mutation). This result is in good agreement with data obtained from a theoretical study performed by Beierlein et al. (2011).

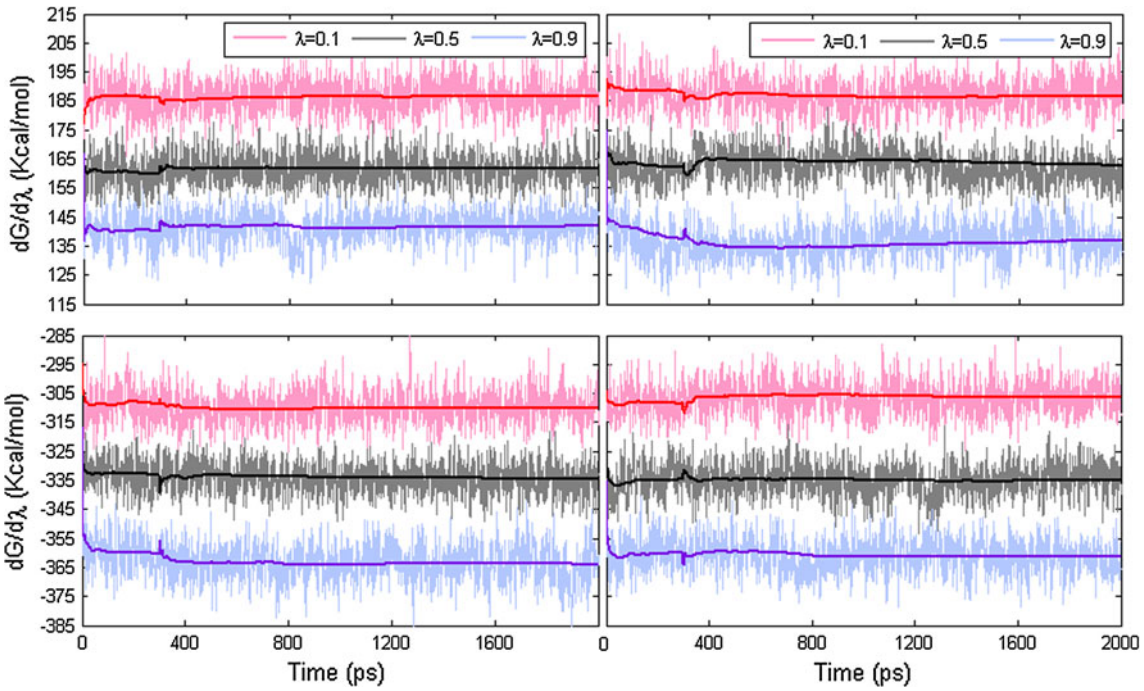

Fig. 9 Free energy gradients dG/dλ for the discharging (*top*) and charging steps (*bottom*) of the TA → GC mutation, as a function of time for $\lambda = 0.1$, 0.5 and 0.9 (*left* and *right plots* are related to the free and bound states, respectively). The cumulative averages are shown in *red*, *black* and *violet lines* for $\lambda = 0.1$, 0.5 and 0.9, respectively

To understand the detailed molecular mechanism of binding, the different components of the free energy were analyzed. The predicted relative binding affinities and the free energy components are summarized in Table 4. To get a better view about which of the three steps, the discharging, vdW-transformation or charging step, has the most effect on the binding free energy, the free energy change associated with the three individual steps (ΔG_{dischg}, ΔG_{vdw} and ΔG_{chg}) in both the free and bound state were carefully compared. Initially, the simple perturbation of the GC → AT and then more complex perturbation of the TA → GC were analyzed.

The statistical uncertainties of dG/dλ values obtained from free energy simulation were estimated using block averaging method (Hess 2002) (Table 4). Block averaging is useful to analyze correlated data, such as the results obtained from a MD simulation. In this method, each trajectory was divided into blocks, and the average dG/dλ of each block was calculated. The errors in dG/dλ value were then estimated as 1 SD of the block averages. In GC → AT mutation, the discharging step of free and bound states (0.29 and 0.64 kcal/mol, respectively) have the smallest and largest error. Corresponding values in TA → GC mutation are 0.23 kcal/mol for the charging step of the free state and 0.73 kcal/mol for the vdW-transformation step of the bound state. These uncertainties are quite acceptable for the analysis of the computed results. This suggests that the MD simulations used are sufficiently well equilibrated to calculate free energies of mutations.

For the GC → AT mutation, only those atoms that are different in purine and pyrimidine bases (and are replaceable to transform purine to pyrimidine and vice versa) were affected. The replaceable atoms of guanine and cytosine bases were discharged, and the replaceable atoms of adenine and thymine bases were charged in the discharging and charging steps, respectively. In the vdW-transformation step, a vdW softcore potential was used for theses atoms. For DNA in the free state (solvated DNA in water), ΔG_{dischg} and ΔG_{chg} are 266.81 and −102.39 kcal/mol, respectively, while ΔG_{vdw} gives 9.81 kcal/mol. From Table 4, it can be understood that for DNA in the bound state (solvated PRH–DNA complex), the free energy related to each of the three steps is not considerably changed; ΔG_{dischg} decreases by 0.67 to 266.14 kcal/mol, ΔG_{vdw} increases by 0.49 to 10.30 kcal/mol, and ΔG_{chg} increases by 0.82 to −101.57 kcal/mol. The overall relative binding free energy of the GC → AT mutation ($\Delta\Delta G$) is equal to 0.64 kcal/mol. Although this value is small, the contribution of ΔG_{chg} is nonetheless greater than that of ΔG_{dischg} and ΔG_{vdw}, respectively.

In the TA → GC mutation, all atoms in thymine and adenine bases were discharged in the discharging step, and all atoms in guanine and cytosine bases were charged in the charging step owing to the transformation of the purine to pyrimidine bases and vice versa. These atoms were treated with the vdW softcore potential in the vdW-transformation step.

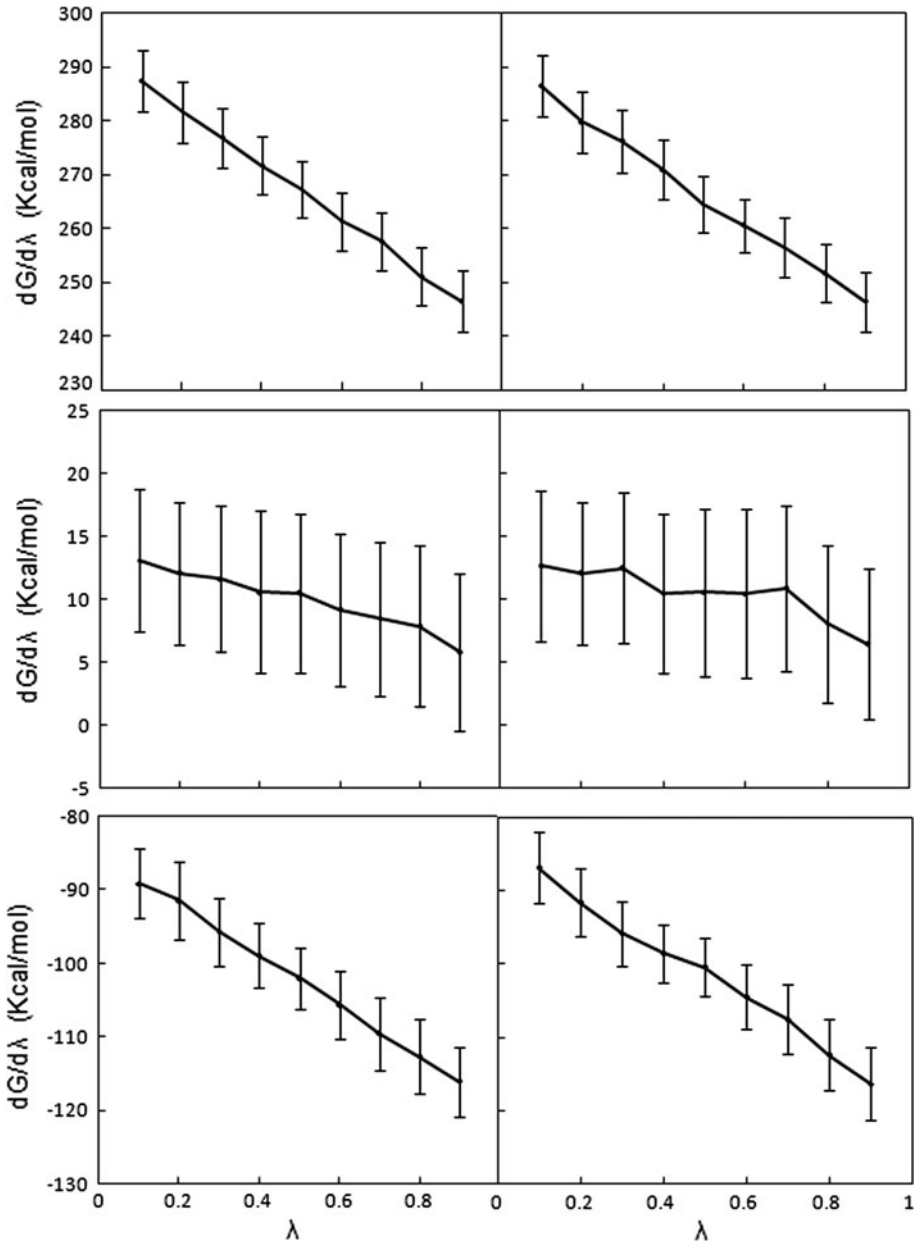

Fig. 10 Free energy gradients $dG/d\lambda$ for discharging (*top*), vdW-transformation (*middle*) and charging steps (*bottom*) of the GC → AT mutation as a function of coupling constant (*left* and *right plots* are related to the free and bound states, respectively). The *error bars* show the SD

For DNA in the free state, ΔG_{dischg} and ΔG_{chg} are 163.08 and −335.67 kcal/mol, respectively. ΔG_{vdw} is equal to −0.84 kcal/mol. From Table 4, it can be found that the thermodynamic effect of removing the charge is very similar when done in the free and bound state. The change in the free energy in the bound and the free states for the charging step (1.8 kcal/mol) is greater than that of the vdW-transformation (1.12 kcal/mol) and discharging (0.91 kcal/mol) steps. In addition, the difference of the free energies in the bound and free states for each of the three steps is larger than the difference observed in the case of the GC → AT mutation. This result can be attributed to the fact that changes created in the TA → GC perturbation is greater than those in the GC → AT perturbation. Interestingly, the main contribution to the difference in binding free energies ($\Delta\Delta G$ of 2 kcal/mol) results from the charging step. The results obtained from the free energy analysis are in good accordance with those of hydrogen bond analysis. In both cases, native complex is more stable than M1 and M2 complexes, and the stability of the M1 complex is more than that of the M2 complex.

Conclusions

In the current study, the structural and energetic effects of base pair mutations on the binding of the PRH protein with specific DNA sequences were investigated. The combination of MD simulation with free energy calculation provides

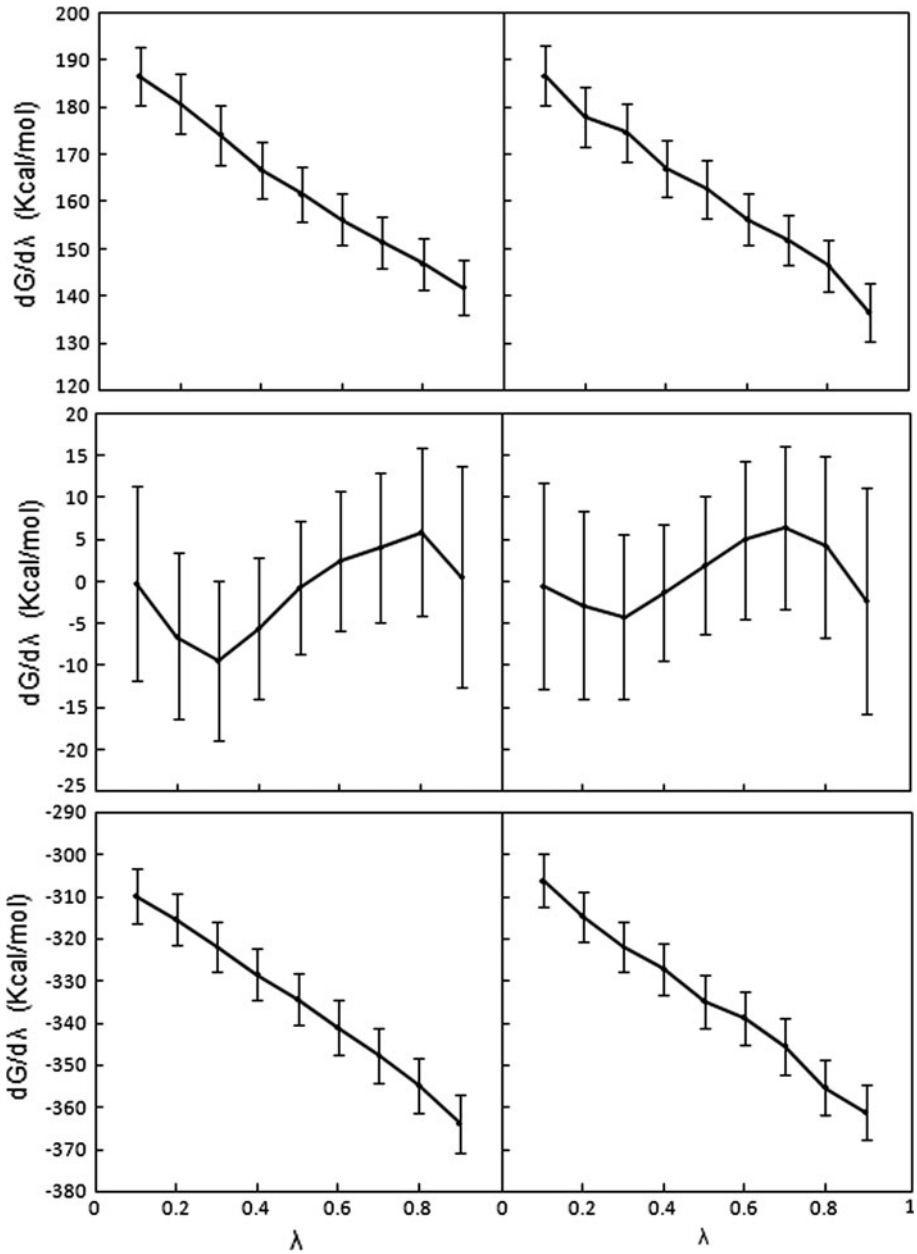

Fig. 11 Free energy gradients $dG/d\lambda$ for the discharging (*top*), vdW-transformation (*middle*) and charging steps (*bottom*) of the TA → GC mutation as a function of the coupling constant (*left* and *right plots* are related to the free and bound states, respectively). The *error bars* show the SD

Table 4 Binding free energy results for GC → AT and TA → GC mutations and their statistical uncertainties (in parentheses)

Free energy changes	Solvated DNA in water[a]	Solvated PRH–DNA complex[a]	Solvated DNA in water[b]	Solvated PRH–DNA complex[b]
ΔG_{dischg}	266.81 (0.29)	266.14 (0.64)	163.08 (0.32)	162.17 (0.57)
ΔG_{vdw}	9.81 (0.49)	10.30 (0.43)	−0.84 (0.68)	0.28 (0.73)
ΔG_{chg}	−102.39 (0.48)	−101.57 (0.55)	−335.67 (0.23)	−333.87 (0.37)
ΔG_{total}	174.23	174.87	−173.43	−171.43
$\Delta\Delta G$	0.64		2	

Free energies in kcal/mol

[a] (GC → AT) mutation

[b] (TA → GC) mutation

details about the molecular mechanism of binding between the protein and the DNA.

In the first part of the study, three MD simulations of 20 ns were performed for the native and two mutated complexes. Structural properties such as the RMSD, RMSF and the secondary structure of protein were evaluated from these trajectories. Upon mutation, analysis of molecular dynamics trajectories showed that the RMSD and RMSF values for both PRH and DNA in the complex were increased, indicating the greater stability of the native complex relative to the mutated complexes, particularly relative to the M2 complex. Despite base pair mutation and changes in DNA sequence, the major secondary structural element of PRH in the native and mutated complexes is an α-helix, and the overall secondary structure pattern of PRH in the complexes is maintained upon the mutations. Hydrogen bond analysis shows, despite maintaining the overall pattern of hydrogen bond interactions, the number of direct and water-mediated hydrogen bonds existing between PRH and DNA in the native complex is greater than that of the two mutated complexes, particularly for the M2 complex. Upon mutation, some hydrogen bonds playing a role in the stability of the complex are lost. The analysis indicates that the occupancy of the hydrogen bonds in the native complex is greater than in the M1 complex, while the M1 complex has greater occupancy than the M2 complex.

In the second part of the study, the relative binding free energy for GC → AT and TA → GC mutations was investigated using a computational approach in which each perturbation is done in three steps. It was found that the predicted relative binding free energies are 0.64 and 2 kcal/mol for GC → AT and TA → GC mutations, respectively. Since the $\Delta\Delta G$ value is positive for both mutations and $\Delta\Delta G$ for the TA → GC mutation is larger than that for the GC → AT mutation, it can be deduced that among the three DNA sequences investigated, the native complex is more stable than the two mutated complexes. However, stability of the M1 complex is greater than that of the M2 complex and is close to that of the native complex. Overall, there is good agreement between the results obtained from the free energy and hydrogen bond analysis that confirm that the native complex is the most stable.

Acknowledgments Computations were performed on the gpc supercomputer at the SciNet (Loken et al. 2010) HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund—Research Excellence; and the University of Toronto.

References

Beierlein FR, Kneale GG, Clark T (2011) Predicting the effects of basepair mutations in DNA-protein complexes by thermodynamic integration. Biophys J 101:1130–1138

Beveridge DL, Dicapua FM (1989) Free-energy via molecular simulation—applications to chemical and biomolecular systems. Annu Rev Biophys Biophys Chem 18:431–492

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

Boresch S, Tettinger F, Leitgeb M, Karplus M (2003) Absolute binding free energies: a quantitative approach for their calculation. J Phys Chem B 107:9535–9551

Case DA, Darden TA, Kollman PA (2008) AMBER 10. University of California, San Francisco, CA

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

Deng YQ, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. J Phys Chem B 113:2234–2246

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

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

Gao J, Kuczera K, Karplus M (1989) Hidden thermodynamics of mutant proteins: a molecular dynamics analysis. Science 244:1069–1072

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

Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J 72:1047–1069

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

Hart K, Nilsson L (2008) Investigation of transcription factor Ndt80 affinity differences for wild type and mutant DNA: a molecular dynamics study. Proteins 73:325–337

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

Jalili S, Karami L (2012) Study of intermolecular contacts in the proline-rich homeodomain (PRH)–DNA complex using molecular dynamics simulations. Eur Biophys J 41:329–340

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

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

Kirkwood JG (1935) Statistical mechanics of fluid mixtures. J Chem Phys 3:300–313

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

Kollman P (1993) Free-energy calculations—applications to chemical and biochemical phenomena. Chem Rev 93:2395–2417

Kwok JBJ, Li Q-X, Hallupp M, Whyte S, Ames D, Beyreuther K, Masters CL, Schofield PR (2000) Novel Leu723Pro amyloid precursor protein mutation increases amyloid β42(43) peptide levels and induces apoptosis. Ann Neurol 47:249–253

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

Loken C et al (2010) SciNet: Lessons learned from building a power-efficient top-20 system and data centre. J Phys Conf Ser 256:012026

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

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

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

Pérez A, Marcha'n I, Orozco M (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. Biophys J 92:3817–3829

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

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

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

Simonson T, Archontis G, Karplus M (2002) Free energy simulations come of age: protein-ligand recognition. Acc Chem Res 35: 430–437

Sneddon SF, Tobias DJ, Brooks CL (1989) Thermodynamics of amide hydrogen-bond formation in polar and apolar solvents. J Mol Biol 209:817–820

Steinbrecher T, Mobley DL, Case DA (2007) Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. J Chem Phys 127:214108–214121

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

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

Tutorial 9, AMBER web site (2009) http://ambermd.org/tutorials/advanced/tutorial9/

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

Zwanzig RW (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. J Chem Phys 22:1420–1426