Role of DNA conformation & energetic insights in Msx-1-DNA recognition as revealed by molecular dynamics studies on specific and nonspecific complexes

Sangita Kachhap and Balvinder Singh*

Bioinformatics Centre, CSIR-Institute of Microbial Technology, Sector 39A, Chandigarh, India

Communicated by Ramaswamy H. Sarma

(Received 18 June 2014; accepted 3 December 2014)

In most of homeodomain–DNA complexes, glutamine or lysine is present at 50th position and interacts with 5th and 6th nucleotide of core recognition region. Molecular dynamics simulations of Msx-1–DNA complex (Q50-TG) and its variant complexes, that is specific (Q50K-CC), nonspecific (Q50-CC) having mutation in DNA and (Q50K-TG) in protein, have been carried out. Analysis of protein–DNA interactions and structure of DNA in specific and nonspecific complexes show that amino acid residues use sequence-dependent shape of DNA to interact. The binding free energies of all four complexes were analysed to define role of amino acid residue at 50th position in terms of binding strength considering the variation in DNA on stability of protein–DNA complexes. The order of stability of protein–DNA complexes shows that specific complexes are more stable than nonspecific ones. Decomposition analysis shows that N-terminal amino acid residues have been found to contribute maximally in binding free energy of protein–DNA complexes. Among specific protein–DNA complexes, K50 contributes more as compared to Q50 towards binding free energy in respective complexes. The sequence dependence of local conformation of DNA enables Q50/Q50K to make hydrogen bond with nucleotide(s) of DNA. The changes in amino acid sequence of protein are accommodated and stabilized around TAAT core region of DNA having variation in nucleotides.

Keywords: homeodomain; binding energy; protein-DNA interactions; roll angle; HTH motif

Introduction

Atomic level study of protein–DNA recognition will reveal binding of proteins and provide vital information to understand specificity between the partners. The specificity of interaction has been proposed to result from a subset of interactions that are sequence specific for a particular protein–DNA complex (Rohs et al., 2010). Molecular structures of the homeodomains solved by X-ray diffraction as well as NMR have also been found to be very much conserved for a number of protein–DNA complexes (Hard, 1999). DNA offers major as well as minor grooves for binding to helices or N-/C-terminal residues or both with N-/C-terminal residues taking up structure at times only in the presence of DNA (Hard, 1999). Phosphate groups or bases of DNA provide sites of interactions for side chains of amino acid residues of proteins in the complexes. In addition, the water-mediated interactions between protein and DNA may also sometimes provide further stability to these interactions. Thus, direct or water-mediated interactions will determine the specificity of the interactions between protein and DNA (Gorfe, Caflisch, & Jelesarov, 2004; Baird-Titus et al., 2006).

Homeobox genes are class of genes that code for regulatory proteins involved in transcription (Cillo, Failla, Cantile, & Boncinelli, 1999; Gehring, 1986; McGinnis & Krumlauf, 1992). These proteins contain 60 amino acid residues that confer it sequence specificity for binding to DNA. Homeodomains were initially identified as master control genes of drosophila. Later, these have been identified in all animal as well as plant and fungal species (Lewis, 1978; Manak, Mathies, & Scott, 1994) and found to be involved in developmental processes and tissue differentiation (Gehring, 1987). A few deformities of craniofacial structures have been found to be associated with defects in the homeoproteins. Homeodomains are predominantly used for understanding protein–DNA interactions mainly because of their small size and binding to specific DNA sequence in the absence of rest of protein (Baird-Titus et al., 2006). Amino acid sequence of homeodomain is highly conserved among this class which includes three helices – helix I, helix II and helix III where latter two helices constitute Helix–Turn–Helix (HTH) motif. Homeodomain binds to DNA via helix III (in major groove) and N-terminal arm (in minor groove). DNA sequence recognized by homeodomain is 5′-TAATNNN-3′ where N is the nucleotide that varies in recognition site of different homeodomains. Since nucleotides of 5th and 6th base pair interact with 50th residue of homeodomain, it has been considered as the residue to be involved in
determining specificity of interaction (Hanes & Brent, 1989; Percival-Smith, Muller, Affolter, & Gehring, 1990; Treisman, Gonczy, Vashishtha, Harris, & Desplan, 1989). Amino acid residue at 50th position is slightly variable among homeodomains. Glutamine occurs at 50th position in 80% of the homeodomains such as Engrailed, Msx-1, Even-skipped and Antennapedia where sequences TAATTA, TAATGG and TAATGG are involved in binding, respectively (Fraenkel, Rouald, Chambers, & Pabo, 1998; Hirsch & Aggarwal, 1995; Hovde, Abate-Shen, & Geiger, 2001). A few of homeodomains have lysine and serine at 50th position as in bicoid and Pax3 homeodomains (Baird-Titus et al., 2006; Birrane, Soni, & Ladias, 2009), respectively, and the preferred DNA sequences are TAATCC and TAATYG where Y is any pyrimidine for lysine (pax2 and bicoid) and serine (Pax3), respectively. However, the number of homeodomains having glutamine at 50th position is large and it recognizes different nucleotides at 5th and 6th positions in different complexes (Fraenkel et al., 1998; Hovde et al., 2001).

Three-dimensional structures have been solved for many of the homeodomain–DNA complexes at reasonably good resolutions (Joshi et al., 2007; LaRonde-LeBlanc & Wolberger, 2003; Longo, Guanga, & Rose, 2007; Pradhan et al., 2012). These complex structures provide a map of interactions between homeodomain and DNA. However, molecular dynamics (MD) simulations will aid the understanding of the mechanistic details of their binding from an ensemble of structures, thus generated. Also, free energy calculations using molecular mechanics Poisson Boltzmann Surface Area approach (MM/PBSA) reveal interacting forces and stability of protein–DNA complexes. This kind of thermodynamics analysis of ensemble of structures generated from MD simulations has provided understanding of cooperativity and specificity of Zinc finger protein–DNA interactions (Lee, Kim, & Seok, 2010).

In most of the homeodomain–DNA structures, it has been found that Q50 does not interact directly with DNA but via water-mediated hydrogen bond (Fraenkel et al., 1998; Hovde et al., 2001). MD simulations have provided an account on the importance of interfacial water molecules in the protein–DNA complexes in addition to overall understanding of the homeodomain–DNA binding. MD studies on Antennapedia (Antp) homeodomain reveal existence of water at the protein–DNA interface (Billeter, Guntert, Lugibihl, & Wuthrich, 1996). However, in Q50K-CC-specific complex, hydrophobic interactions stabilize the complex in the absence of water (Gutmanas & Billeter, 2004). X-ray diffraction structure data of Q50K engrafted nonspecific complex show that K50 makes hydrogen bonds with guanine at positions 5th and 6th of DNA (Tucker-Kellogg et al., 1997). MD simulation study shows that Q→K mutation at 50th position switches specificity from 5'-TAATGG-3' to 5'-TAATCC-3'. Q50K mutation has been proposed to allow TAATCC sequence of DNA to move with respect to protein perpendicular to DNA axis as a result K50 makes hydrogen bond with base pairs at positions 5th and 6th in Antp homeodomain (Gutmanas & Billeter, 2004). MD simulation of Engrailed homeodomain discusses the role of Q50 in specificity of protein–DNA interaction as it does not form direct interaction with DNA base but makes water-mediated hydrogen bond with DNA base (Zhao, Huang, & Sun, 2006).

The present study investigates the role of conserved amino acid residues at 50th position and sequence dependent structure of DNA in protein–DNA recognition of Msx-1–DNA complex. Here we also show that binding free energy of N-terminal amino acid residues contributes more to the stability of protein–DNA complex as compared to the residue at 50th position.

**Materials and methods**

**Protein–DNA complexes**

The structure of Msx-1–DNA complex that has been solved by X-ray diffraction was taken from Protein Data Bank (PDBID – 1IG7) (Hovde et al., 2001) for MD simulations. The hanging lone bases from both the sides were removed leaving 13 base pairs (5'-CAC-TAATTGAAGG-3') in the DNA sequence. Msx-1 homeodomain consisted of 2–59 amino acid residues. The mutation (Q50K) in the protein has been made using PyMOL in Msx-1–DNA complex. Lysine replaced Q50 with its side chain in extended conformation (Figure S1). The stereo-chemical quality of mutant protein has been checked by ProSA (Sippl, 1993; Wiederstein & Sippl, 2007) and Verify3d (Eisenberg, Luthy, & Bowie, 1997) servers. The z-score for mutant calculated by ProSA is −4.29 which is in the range of such scores for NMR and X-ray structures. The pattern of line diagrams/values of raw data has been similar for wild-type and mutant proteins in Q50-TG and Q50K-TG complexes, respectively, in the Verify3d server. The bases in DNA have been replaced to create TG→CC base pairs in Msx-1–DNA complex using MMTSB toolset (Feig, Karanicolas, & Brooks III, 2004). To create double replacement in specific complex, Q50K-CC, protein mutant has been used, employing the MMTSB toolset. Thus, four protein–DNA complexes, Q50-TG & Q50K-CC also known as specific complexes, nonspecific complexes where TG replaced CC (Q50-CC), Q50K mutant complex (Q50K-TG), have been prepared as described above.

**MD simulation**

Leap module of AMBER11 suite of programs (Zhang, Hou, Schafmeister, Ross & Case) has been used in setting the initial structure for MD simulations. All the four complexes (Q50-TG, Q50-CC, Q50K-TG and Q50K-CC) have been subjected to identical minimization and MD
procedures. MD simulations of these protein–DNA complexes were performed in AMBER11 software package (Case et al., 2005; 2010; Pearlman et al., 1995) using force field ff99SB and a modified force field of DNA parmsc0 (Perez et al., 2007). MD simulation was carried out after solvating protein–DNA complex with TIP3P water (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) with buffer size 8 Å. The resulting sizes of octahedral box were 68 × 68 × 68 Å for all the four complexes. These systems comprising of different protein–DNA complexes have been neutralized with appropriate number of sodium ions. The solvated system was minimized in two steps to relax solvent and to prevent any steric clashes between solute and solvent. First minimization of solvent and ions was performed for 5000 (1000 cycles of steepest descent followed by 4000 cycles of conjugate gradient) cycles with positional restraint of 10 kcal/mol/Å^2 on protein and DNA. Then final minimization of protein–DNA complexes was performed for 6000 (2500 cycles of steepest descent followed by 3500 cycles of conjugate gradient) cycles without any positional restraint on protein and DNA. Solvated minimized protein–DNA complexes were heated from 10 to 300 K for 300 ps at NVT ensemble with 5 kcal/mol/Å^2 positional restraint on backbone atoms of protein and those of DNA involved in base pairing via hydrogen bond to maintain double-stranded helical structure of DNA. Positional restraint was removed gradually in two steps, 150 ps of each from 3 to 1 kcal /mol/Å^2 after heating. Unrestrained constant pressure equilibration dynamics of 400 ps was carried out at 300 K. Finally, production run of 30 ns was carried out at NVT ensemble. For each simulation, SHAKE (Miyamoto & Kollman, 1992; Ryckaert, Cicotti, & Berendsen, 1997) algorithm was used to constrain bond involving hydrogen atom and nonbonded interaction distance cut-off was set to 10 Å. Newtonian equation of motion was integrated for every 2 fs during MD simulations. Particle mesh Ewald summation was used at a cut-off distance of 10 Å with grid spacing kept at 1 Å with cubic spline interpolation (Darden, York, & Pedersen, 1993). Thirty nanoseconds long MD simulation has been carried out during production run at constant volume and temperature of 300 K for all the protein–DNA complexes. Though NPT is close to the physical set-up of most experiments, but time-dependent behaviour is a suspect. Further, most of the properties of proteins are insensitive to change between NVT and NPT during production run. The temperature was kept constant throughout the simulation with use of Langevin control using low-value collision frequency. SHAKE was used, and the coordinates were saved after every 2 ps. Other parameters such as nonbonded cut-off and electrostatics have been kept same as in equilibration run.

The trajectory analysis was carried out by ptraj module of AMBER11. Hydrogen bond and hydrophobic interactions were analysed for all protein–DNA complexes. Donor–acceptor distance and donor–hydrogen–acceptor angle cut-off used for hydrogen bond analysis were less than 3 Å and 145°, respectively. Hydrophobic contacts were calculated with carbon–carbon distance less than or equal to 4.5 Å. DNA helical parameters were analysed with Curves+ (Lavery, Moakher, Maddocks, Petkeviciute, & Zakrzewska, 2009).

MMPBSA analysis

MM PBSA module (Srinivasan, Cheatham III, Cieplak, Kollman, & Case, 1998) of AMBER11 was used for calculation of binding free energy. Binding free energies of protein–DNA complexes specific as well as nonspecific have been calculated as ΔG(binding) = G(complex) - G(protein) - G(DNA), where G(complex), G(protein) and G(DNA) are free energy of complex, protein and DNA molecules, respectively.

Single-trajectory approach was employed for binding free energy calculation, and gas phase energy was calculated by Sander. Single-trajectory method is computationally efficient and fast, while triple-trajectory method is used when significant conformation changes occur on binding (Treesuwan et al., 2009; Wan, Coveney, & Flower, 2005). PB equation was solved by Adaptive Poisson–Boltzmann Solver (APBS) program through the iAPBS interface with AMBER11. Internal and external dielectric constants were set to 1.0 and 80.0, respectively. Entropy calculation was done in gas phase using nmode program in AMBER11. To calculate enthalpy, every 10 snapshots of last 15 ns trajectory were used for each of the four protein–DNA complexes; thus, total 750 snapshots of 15 ns (7500 snapshots) trajectory were used for calculation. For entropy calculation, a total of 250 snapshots were extracted from last 15 ns (7500 snapshots) trajectory at an interval of 30 ps.

Results

MD of the specific protein–DNA complexes (Q50-TG) (Hovde et al., 2001) and Q50K-CC (Baird-Titus et al., 2006) and nonspecific protein–DNA complexes – having TG replaced by CC (Q50-CC) and Q50K mutant protein complex (Q50K-TG) – have been carried out, and trajectories have been analysed for different types of interactions and stabilizing forces between atoms of protein and DNA forming these complexes.

MD trajectories of protein–DNA complexes

The root mean square deviation (RMSD) of all four protein–DNA complexes during MD simulations has been computed for backbone atoms (N, Ca and C of protein and P, O5′, C5′, C4′, O4′, C3′, O3′ of DNA) for each
frame of MD trajectories with respect to minimized structure. All homeodomain amino acid residues and a region of DNA (5′-TAATTGAA-3′) interacting with homeodomain have been considered for RMSD calculations. Three base pairs on 5′ side and two base pairs on 3′ side of DNA, if included, contribute to overall higher values of RMSD, therefore, have not been considered as these do not interact with any of amino acid residues of homeodomain. RMSD plots of all four protein–DNA complexes are given in Figure 1.

In specific complex Q50-TG, average RMSD of protein was 0.92 ± 0.15 Å and of DNA was 1.32 ± 0.25 Å (Figure 1(A)) while in TG→CC substituted nonspecific complex Q50-CC, average RMSD of protein and DNA increased a little up to 1.16 ± 0.20 Å and 1.51 ± 0.20 Å, respectively (Figure 1(B)). The trajectory of Q50K-TG has been observed to have an average RMSD of 1.13 ± 0.17 Å and 1.23 ± 0.15 Å for protein and DNA, respectively (Figure 1(C)). TG to CC substitution and Q50K mutation in protein–DNA complex, Q50K-CC, shows average RMSD of protein and DNA similar to Q50-TG, that is 0.91 ± 0.16 Å and 1.30 ± 0.22 Å, for protein and DNA, respectively (Figure 1(D)).

Further, root mean square fluctuations (RMSFs) of amino acids of proteins and nucleotides of the protein–DNA complexes have been calculated. In specific complex (Q50-TG), RMSFs of terminal nucleotides and C-terminal end residue of protein have reached up to 2.6 and 2.8 Å, respectively, while that of remaining residues is below 2 Å (Figure S2(A)). This implies that conformation of amino acid residues of protein and DNA nucleotides remains stable throughout the trajectory of MD simulations. In nonspecific complex Q50-CC, RMSF of terminal nucleotide of each strand of DNA as well as
that of respective protein reached up to 3 Å while for rest of the amino acid residues or nucleotides, it remained below 2 Å similar to Q50-TG, a specific complex (Figure S2(B)). This indicates that mutation in DNA has little effect on the flexibility of amino acid residues of protein as well as that of nucleotide of DNA. In protein nonspecific complex Q50K-TG, RMSF of terminal amino acid residues of Msx-1 has increased to 4 Å while for remaining amino acid residues as well as for nucleotides of DNA, it remained less than 2.3 Å (Figure S2(C)). This also indicates that mutation in protein does not affect flexibility or conformation of rest of the amino acid residues of protein as well as nucleotides of DNA. In specific protein–DNA complex, Q50K-CC, amino acid residues of C-terminal residues have RMSF of 4 Å while rest of the amino acid residues in protein has RMSF of less than or equal to 2.3 Å similar to Q50K-TG complex (Figure S2(D)).

Protein–DNA contacts

Hydrogen bond interaction

In Msx-1–DNA-specific complex, homeodomain binds to major groove of DNA through amino acid residues of helix III (K46, Q50, N51, R53, K55 and R57), R31 of helix II and Y25 of loop between helix I and II. Homeodomain also binds to minor groove of DNA via amino acid residues in N-terminal arm (R2, K3, R5 and T6). The interactions between protein residues and DNA nucleotides have been analysed during MD simulations and are presented in Figure 2. The interactions that are observed in crystal structure of Msx-1–DNA complex have also been observed during MD simulations of specific complex, Q50-TG.

The interactions of Q50, with nucleotides of DNA have been monitored closely as glutamine is conserved at position 50 and it has been previously implicated in recognition of specific DNA sequence. In specific complex Q50-TG, the side chain of Q50 makes a hydrogen bond with nitrogen of Cyt22 at distance of less than 3 Å between hydrogen bond donor and acceptor (Figure 3(A)). Water molecules have been found to be resident during MD trajectories at the interface of protein and DNA. Thirteen different positions of water molecules that have been proposed by Hovde et al. (2001) to bridge the interaction between protein and DNA are observed as shown in the last structure of MD trajectory of Q50-TG (Figure 4). There is exchange of water molecules between Q/K50 and 11th and 12th nucleotides at protein–DNA interface and bulk of the solvent during the trajectories of specific and nonspecific protein–DNA complexes. Hydrogen atoms of a water molecule have been observed to point towards Thy11 & Gua12 and oxygen of water towards Thy50 for approximately 1.5 ns for Q50-TG complex in the trajectory. The time of occurrence of water molecules bridging interactions between the point of mutation, Q/K50 and corresponding nucleotides of DNA during MD trajectories of protein–DNA complexes up to distance of 3.5 Å is presented in Table 1. In Q50-CC nonspecific complex, Q50 interacts with oxygen of phosphate of Gua22 instead of base as observed in Q50-TG complex (Figure 3(B)). This interaction of Q50 with phosphate group of Gua22 in Q50-CC also facilitates the interaction of side chain of N51 with phosphate oxygen of Ade8 instead of base of Ade9 as in Q50-TG complex. Also, a water molecule was observed between Q50 and Thy21 for a residence time of 11 ns during simulations of Q50-CC complex.

Figure 2. (A) Occupancy of hydrogen bonds in protein–DNA complexes – – Q50-TG, – Q50-CC, – Q50K-TG and – Q50K-CC during MD simulations. Lower axis shows interacting nucleotide amino acid residue combination. (B) Snapshot from MD trajectory of Msx-1–DNA complex depicting interactions (dotted lines) between amino acid residues and nucleotides of DNA.
This position has been occupied by four different transient water molecules with maximum residence time of 11% for each of the water molecules. The interaction involving N51 and phosphate oxygen of Ade8 has also been observed in Q50K-TG complex. K50 in Q50K-TG makes single hydrogen bond with oxygen of Thy10 because Cyt22 does not have hydrogen bond acceptor to form hydrogen bond with K50 (Figure 3(C)).

In addition, two molecules of water bridge interactions between K50 and Ade23 in nonspecific complex, Q50K-TG replacing one another for a total of 7 ns (Table 1). K50 is involved in two hydrogen bonds, that is with Gua23 and Thy10 (Figure 3(D)) in Q50K-CC complex. Due to complementary changes in protein as well as DNA, there is restoration of interaction of amino acid at 50th position that is K50 with nucleotides at 11th

---

Figure 3. Changes in interactions of amino acid residues of Msx-1-Y25, Q/K50 and N51 with DNA in protein–DNA complexes. (A) Q50-TG, (B) Q50-CC, (C) Q50K-TG and (D) Q50K-CC.
and 12th positions of DNA similar to as in specific complex, Q50-TG at corresponding positions of DNA. A water molecule occupies a position between K50 and Thy10 for a total of 18 ns of trajectory of specific complex, Q50-TG, but not simultaneously. However, this interaction is lost with one of the oxygen atoms of phosphate in Q50K-TG and Q50K-CC complexes while in Q50-CC complex, there is a loss of interaction of Y25 with DNA as Q50 interacts with phosphate group of Gua22 instead of base of Cyt22 as in specific complex of Q50-TG. This also allows side chain of R53 to have bifurcated interactions with oxygen atoms of phosphate group of Gua22. Also, side chain of R53 forms hydrogen bond with phosphate oxygen of Gua22 /Cyt22 for longer time during trajectory in Q50-TG, Q50K-TG and Q50K-CC as compared to that in Q50-CC (Figure 2(A)). In Q50-TG, Y25 interacts with one of oxygen of phosphate of Cyt22 for 53% of time while R53 interacts with other oxygen of phosphate of Cyt22. Side chain of K57 is involved in hydrogen bond interactions with both oxygen atoms of phosphate group of Gua23 due to the replacement of bases in Q50-CC complex.

Hydrophobic interaction
The hydrophobic interactions involving R2, F8 and I47 with Thy10 & Ade27, Ade8 and Thy10, respectively as observed in X-ray structure of protein–DNA complex have also been observed during MD simulations of specific complex, Q50-TG as well as other nonspecific complexes. In addition, the amino acid residues K3 and T6 show interactions with Ade9 and Ade8, respectively, during MD simulations of Q50-TG complex. In nonspecific complex Q50-CC, there is increase in distance between I47 and base of Ade9 resulting in loss of interaction as a result of increase in distance between Q50 and Cyt22 and decrease of distance between Q50 and phosphate

Table 1. Time for which water molecules have been observed at the interface of Q/K50 and corresponding nucleotides in specific (Q50-TG & Q50K-CC) and nonspecific (Q50K-TG & Q50K-CC) protein–DNA complexes.

| Protein–DNA contacts | Protein–DNA complexes (ps) |
|----------------------|---------------------------|
|                      | Q50-TG  | Q50-CC | Q50K-TG | Q50K-CC |
| Total number of snapshots |        |        |         |         |
| Q/K50-Thy11          | 1500    |       | 7000    |         |
| Q50-Gua12            | 1500    |       |         |         |
| Q50-Thy21            |         | 11,000 |         |         |
| K50-Ade23            |         |       | 7000    |         |
| K50-Thy10            |         |       |         | 18,000  |
group of Gua22 in nonspecific complex as compared to Q50-TG complex. On similar lines, there is increase in distance between I47 and Ade9 & Thy10 in Q50K-TG complex resulting in their loss of interactions as the amino acids K50 and N51 interact with Thy10 and Ade8, instead of Cyt22 and Ade9 in simulations of Q50-TG complex, respectively. In Q50K-CC specific complex having changes in DNA as well as in protein, the interactions between I47 and Ade9 & Thy10 are retained along with the interactions of N51 with Ade9 and K50 with Thy10 & Gua23. Thus, there is restoration of interaction of I47 with Ade9 and Thy10 in this specific complex, Q50K-CC as compared to interactions of nonspecific complex, Q50K-TG, where only amino acid of protein has been replaced.

**DNA helical parameter analysis**

The nucleotide sequence of DNA is different for specific and nonspecific complexes. The variations in helical parameters of DNA have been monitored while being in complex with proteins. The base pair roll angles of MD simulated DNA alone and in specific Q50-TG & non specific protein–DNA complexes were compared with that of DNA in crystal structure (Figure 5). In Q50-TG, TpG base pair (bp) step has high roll angle than that of X-ray structure of this specific complex by 5° (Figure 5(A)). In addition, the roll angle of DNA at the 8th position in Q50-TG complex is 15.8° as compared to 7.9° for CpC bp step as in Q50-CC (Figure 5(B)) protein–DNA complex. This difference between the roll angles of DNA has also been observed at positions—TpG and CpC (Figure 5(C)) in respective DNAs that is TG DNA, have high roll angle for TpG bp step as compared to CpC bp step in CC DNA. Thus, roll angle of DNA is sequence dependent for 8th bp step in these protein–DNA complexes. Due to high roll angle of TpG bp step in Q50-TG, hydrogen bond donor of Cyt22, N4 is available to form hydrogen bond with Oε1 of Q50. However, in Q50-CC as roll angle of CpC bp step decreases compared to TpG bp step in Q50-TG, hydrogen bond acceptor atom, N7 of Gua22 is not available to interact with Nε2 of Q50. Instead, it makes hydrogen bond with phosphate rather than base. Thus, the change in roll angle of DNA resulting from change in the sequence of DNA is responsible for loss of interactions in protein–DNA complex.

**Analysis of free energy contribution to the stability of the protein–DNA complexes**

The binding free energies along with the energy terms for specific and nonspecific Q50-TG, Q50-CC, Q50K-TG and Q50K-CC protein–DNA complexes are given in Table 2. Comparison of binding free energy of all the protein–DNA complexes shows that specific protein–DNA complex, Q50-TG, is the most stable while another specific complex, Q50K-CC is 5.5 kcal/mol less stable than the most stable complex. The order of stability of protein–DNA complexes is Q50-TG > Q50K-CC > Q50K-TG > Q50-CC. The protein–DNA complex, Q50-CC, is 17 kcal/mol less stable than the most stable specific protein–DNA complex. The specific complex, Q50K-CC, is more stable (by more than 4 kcal/mol) than corresponding nonspecific complex, Q50K-TG. Nonpolar interactions are comparable for specific protein–DNA complexes, Q50-TG and Q50K-CC complex. However, polar as well as entropic contributions are the maximum for Q50K-CC complex in comparison with others. In all the protein–DNA complexes, free energy contribution from nonpolar interactions is considerably higher as compared to those by polar interactions. In other words, ΔG_{non} contributes ~75% to binding energy while ΔG_{polar} contributes ~25% for specific as well as nonspecific complexes. The contribution of polar interactions in binding energy is almost equal in all the protein–DNA complexes. Among components of nonpolar energy, van der Waals interactions have more contribution to enthalpy of complex than nonpolar interactions due to solvation energy. Some of positively charged residues such as R2 and R5 contribute high van der Waals energy towards the stabilization of protein–DNA complex in addition to electrostatic energy for all the protein–DNA complexes. In case of polar energy component, contribution of electrostatic energy favours complex formation as charge-charge interactions are involved in protein–DNA complex formation.

** Decomposition of free energy for different protein–DNA complexes**

Binding free energy of protein–DNA complex was further decomposed to that of individual amino acid residues to know their contributions. The amino acid residues that contribute majorly in the stabilization of protein–DNA complexes are given in Figure 6. In specific complex Q50-TG, positively charged residues (like R2, K3, R5 and R31) are involved in interactions with DNA and also have major contribution to binding free energy. In addition to these major contributing residues, R53, K55 and R58 contribute to binding free energy of this specific protein–DNA complex. N-terminal arm residues including R2, K3, R5, T6 and F8 contribute maximally towards binding free energy of the complexes as compared to other amino acid residues in all the protein–DNA complexes. The contribution of these amino acid residues more or less remains same in nonspecific complex of Q50-CC. The decrease in the contribution of binding free energy of R31 and its corresponding nucleotide, Thy20 in nonspecific complex, Q50-CC results in
A decrease in binding free energy of complex. The contributions of R53, K57 and R58 in binding free energy increase considerably due to their interactions with Gua22, Gua23 & Ade24 and Ade24, respectively. Binding energy contributions of corresponding nucleotide also stabilize the complex. The contribution of Y25 and Gua22 to binding free energy has decreased in this protein–DNA complex due to loss of interactions between the respective groups of protein and DNA.

Amino acid residue, K50, makes least contribution of 1.1 kcal/mol in the binding free energy of the nonspecific complex, Q50K-TG. Amino acid residue, K46, and corresponding nucleotide, Thy20, have increased its contribution to the binding free energy of this nonspecific complex as compared to specific complex, Q50-TG. In addition to these residues, there is also decrease in binding energy contribution of S27 & I28 with latter amino acid residue having lost the interaction with DNA in

Figure 5. DNA roll angle (A–C) Roll angle of standard B-DNA (■); DNA in the crystal structure (●); DNA in MD simulated Q50-TG (▲) and Q50-CC (▼) complexes and MD simulated TG (◄) & CC (►) DNAs. Vertical lines represent ± SD of fluctuation in (°).
Q50K-TG nonspecific complex. Amino acid residue, I28, loses interactions with Thy20 due to interactions of phosphate oxygen atoms of Thy20 with K46. Though some of the interactions in protein–DNA complex involving TG→CC replacement break, however the interaction of R53 with phosphate group of Gua22 has become stabilized, thus explaining the increased contribution of this amino acid residue in the stabilization of the complex, Q50-CC. The contribution of Y25 has been found to be nearly same in specific complexes, Q50-TG and Q50K-CC, as it is involved in interactions with Cyt22 and Gua22, respectively. Some of the terminal residues such as K3 (N-terminal) and R58 (C-terminal) contribute less to binding free energy of Q50K-TG complex. Due to Q→K mutation, K50 interacts with Thy10 and interaction of N51 with phosphate of Ade8 rather than base of Ade9 leads to decrease in binding energy contribution, of Ade9 in Q50K-TG nonspecific complex. The breaking of interactions of K50 with Cyt22 and interactions with Thy10 results in decrease and increase in their binding energy contributions respectively, in this nonspecific complex. Further, the weakening of K3 interaction with Ade9 in Q50K-CC reflected in decreased binding free energy contribution of Ade9. In addition, the interactions of K50 with Thy10 as well as Gua23 lead to increase in binding free energy contribution. Amino acid residue-wise energy decomposition analysis shows that contribution of Q50 is less as compared to the positively charged residues towards binding free energy of specific protein–DNA complex of Q50-TG. Positively charged residues that interact with DNA have contribution of more than or equal to 4 kcal/mol while those of Q50 in specific and nonspecific complexes have total energy contribution of less than or equal to 2.2 kcal/mol. Lysine residue at 50th position in specific complex of Q50K-CC contributes maximally of 3.5 kcal/mol among all the protein–DNA complexes studied.

**Table 2.** Binding free energy (kcal/mol) along with its components. Specific complexes: Q50-TG & Q50K-CC and nonspecific complexes: Q50-CC & Q50K-TG.

| Energy terms | Q50-TG | Q50-CC | Q50K-TG | Q50K-CC |
|--------------|--------|--------|---------|---------|
| $\Delta E_{el}$ | $-5346.5$ | $-5304.9$ | $-5901.3$ | $-5963.6$ |
| $\Delta E_{vdw}$ | $-95.7$ | $-89.9$ | $-85.2$ | $-95.1$ |
| $\Delta G_{elasolv}$ | $-11.4$ | $-11.3$ | $-10.8$ | $-11.5$ |
| $\Delta G_{elasolv}$ | $5303.6$ | $5270.0$ | $5855.8$ | $5917.1$ |
| $\Delta G_{adv}$ | $5292.2$ | $5258.7$ | $5845$ | $5905.8$ |
| $\Delta G_{gas}$ | $-442.2$ | $-5394.8$ | $-5986.5$ | $-6058.7$ |
| $\Delta G_{rep}$ | $-107.1$ | $-101.2$ | $-105.9$ | $-106.6$ |
| $T \Delta S$ | $-42.9$ | $-34.9$ | $-45.5$ | $-46.5$ |
| $\Delta H$ | $-150.1$ | $-136.1$ | $-141.6$ | $-153$ |
| $\Delta G$ | $-108.2$ | $-91$ | $-99.3$ | $-103.7$ |

**Discussion**

Sequence-specific variability of the proteins in correspondence of the sequence of cognate DNA has been exploited by nature in forming specific protein–DNA complex. There are different ways of formation of specific protein–DNA complexes, that is protein recognizes either chemical property of DNA base to bind a sequence (base-readout) or sequence-dependent shape of DNA (shape readout) (Joshi et al., 2007). Shape-readout may involve kinking, bending and changes in major or minor groove width of DNA. MD simulations have been performed on four different specific and nonspecific complexes involving HTH domain and DNA during the current study. The trajectory of molecular simulations on the specific complex, Q50-TG, shows conserved interactions between protein and DNA as compared to those in crystal structure of this complex and those of MD simulations of similar homeodomain–DNA complexes (Gutmanas & Billeter, 2004; Zhao et al., 2006). Hydrogen bond analysis of specific and nonspecific complexes shows that Q50 forms single hydrogen bond with DNA in specific complex, Q50-TG. However, bidentate hydrogen bond has been shown to have high degree of specificity than bifurcated hydrogen bond (Roehs et al., 2010) and single hydrogen bond (Coulocheri, Pigis, Papavassiliou, & Papavassiliou, 2007) while single hydrogen bond rarely confers specificity (Luscombe, Laskowski, & Thornton, 2001). Thus, in Msx-1, Q50 appears to have least role in recognition of specific sequence because of single hydrogen bond formation with DNA. This has also been observed in MD simulations of other complexes involving Antp and Engrailed homeodomains (Gutmanas & Billeter, 2004; Zhao et al., 2006). In non-specific complex Q50-CC, Q50 forms single hydrogen bond with phosphate group of Gua22 of DNA, not with base as observed in specific complex Q50-TG, due to steric hindrance of bulker group of Gua22 and Gua23. This is also accompanied by breaking and changes in interaction of Y25 and N51, respectively, with DNA. Further, the side chain of lysine, another amino acid residue at 50th position, is accommodated and stabilized in Q50K-CC complex as it forms bifurcated hydrogen bond with DNA. The side chains of amino acid residues at this position in other nonspecific complexes do not form strong interactions with bases of DNA at protein–DNA interface. The specific complex, Q50-TG, shows presence of water molecules at the interface of protein–DNA complex in a similar way as observed in the crystal structure of Msx-1–DNA complex (Hovde et al., 2001). A water molecule held by Q50, Thy11 and Gua12 provides stability to the protein–DNA interface in this complex. Protein–DNA complexes have been found to be hydrated during the trajectory of specific and nonspecific complexes. The longest water–mediated interactions...
Figure 6. Comparison of binding free energies of amino acid residues between specific and nonspecific complexes. Q50-TG & Q50K-TG (top), Q50-TG & Q50K-CC (middle) and Q50K-TG & Q50K-CC (bottom). – Q50-TG, – Q50K-TG, – Q50K-CC.
observed in the trajectories of protein–DNA complexes are those between K50 and Thy10 of Q50K-CC complex. The interaction between K50 and Thy10 is also supported by one of the conformations of lysine residue in the engrailed homeodomain–DNA complex (Tucker-Kellogg et al., 1997).

The changes in the roll angle of DNA have been observed as TG replaces CC in DNA sequence. In specific type complex, Q50-TG, DNA kinks at 8th base pair step (TpG Pyrimidine-purine base pair step) due to high roll angle that allows Q50 to form hydrogen bond with Cyt22 base. Similarly, in nonspecific complex Q50-CC, 9th base pair step (CpA Pyrimidine-purine base pair step) has high roll angle introducing a kink at (CpA) base pair step. Though hydrogen bond acceptor is available to make hydrogen bond with Q50 but due to high roll angle for 9th base pair step, there is increase in distance between these as a result Q50 makes an interaction with phosphate of Gua22 instead of base. TpG base pair step has been observed to have high roll angle as compared to the similar values of roll angles in aristaless–clawless–DNA complex (Miyazono et al., 2010). Pyrimidine-purine base pair step TpA, CpA, (TpG) and CpG are less stabilized through base stacking and are considered as hinge step (Olson, Gorin, Lu, Hock, & Zhurkin, 1998). These changes in DNA sequence and the structure are responsible for the change in the hydrogen bonding pattern between protein and DNA, thus governing the formation of specific protein–DNA complex. These results further suggest that in specific complex Q50-TG, Msx-1 uses indirect recognition mechanism to bind specific DNA sequence. Thus, the changes in DNA sequence accompanied by variability in the structure of DNA appear to accommodate different amino acid residues of the protein around the TAAT core region of the homeodomain.

MM/PBSA analysis of specific and nonspecific complexes reveals that the specific complex Q50-TG is more stable than nonspecific complexes – Q50-CC and Q50K-TG. The Q50-TG protein–DNA complex is largely stabilized by the enthalpic energy contribution as compared to entropic. Mutation in DNA (Q50-CC) affects the stability of protein–DNA complex more as compared to those of other nonspecific complex, that is Q50K-TG as interactions of amino acid residues (Y25, N51 and I47) are disrupted with DNA in addition to those of Q50 in these complexes. Thus, the observations suggest that Msx-1 binds to specific sequence with more binding free energy than nonspecific DNA sequence. This was further supported by the fact that binding free energy of engrailed homeodomain for specific DNA (TAATTA) is much greater than nonspecific DNA (TAATCC) (Ades & Sauer, 1994). Similarly, decrease in stability of nonspecific complex Q50K-TG, where protein has single point mutation compared to specific complex Q50-TG, involves disruption of interaction of I47 with DNA and is supported by specific DNA complex of engrailed homeodomain having less $K_d$ value than Q50K engrailed with nonspecific DNA (Ades & Sauer, 1994). The equilibrium and kinetic DNA binding data further confirm the strong binding of Q50K with specific DNA sequence, that is TAATCC for engrailed homeodomain complex (Tucker-Kellogg et al., 1997). MM/PBSA analysis for Msx-1 homeodomain complex in our study also supports similar protein–DNA binding. The residue wise decomposition of binding energy of protein–DNA complexes reveals that Q50 has less contribution to binding free energy as compared to positively charged residues, at protein–DNA interface. Glutamine has hydrogen bond donor $N_{\text{ε1}}$ and acceptor $O_{\text{ε2}}$ atoms in its side chain that allows it to form protein–DNA complex utilizing different types of interactions – direct such as hydrogen bond, water-mediated and van der Waals interactions with different DNA sequences. However, the lysine residue at this position shows more binding energy in specific complex, Q50K-CC as compared to that of Q50 in Q50-TG due to its greater interaction with DNA. Since amino acid sequences of homeodomains are almost conserved along with their DNA binding site having conserved TAAT core region along with 5th and 6th nucleotides that vary for different homeodomain DNA binding site(s), this appears that glutamine at position 50th of the sequence of protein allows the protein to interact with a varied number of DNA sequences either via hydrogen bonding or via water-mediated contacts. The amino acid residue at this position may facilitate to present the interacting groups of the protein to relatively conserved TAAT core region of the DNA or itself interacts with nucleotides near to TAAT region of DNA. Decomposition analysis of binding energy calculations further reveals that amino acid residues of N-terminal arm (R2, K3, R5, T6 and F8) contribute mainly to the formation of protein–DNA complexes in this study. The greater energy contribution to the stability of complex by these amino acid residues underlines their role in providing specificity to the protein–DNA complex.

Conclusions

MD simulations of specific and nonspecific homeodomain–DNA complexes show the conservation or differences in polar and nonpolar interactions between protein and DNA with regard to their respective mutations. The role of roll angle of DNA has been presented in differentiating direct or indirect mechanism for specific and nonspecific protein–DNA complex formation. The binding energy analysis further reveals a little role in specificity for the conserved amino acid residues at 50th position of protein, however, highlighting the role of N-terminal arm
residues in providing more stability to the protein–DNA complex in addition to TAAT core region.

Supplementary material
The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2014.995709.

Acknowledgements
S.K. acknowledges the research fellowship received from Council of Scientific & Industrial Research (CSIR), New Delhi, India. This work was supported by CSIR network project GENESIS and Department of Biotechnology, New Delhi, India.

References
Ades, S. E., & Sauer, R. T. (1994). Differential DNA-binding specificity of the engrailed homeodomain: The role of residues 50. *Biochemistry*, 33, 9187–9194.
Baird-Titus, J. M., Clark-Baldwin, K., Dave, V., Caprelli, C. A., Ma, J., & Rance, M. (2006). The solution structure of the native K50 Bicoid homeodomain bound to the consensus TAATCC DNA-binding Site. *Journal of Molecular Biology*, 356, 1137–1151.
Billetter, M., Guntert, P., Lugmair, P., & Wuthrich, K. (1996). Hydration and DNA recognition by homeodomains. *Cell*, 85, 1057–1065.
Birrame, G., Soni, A., & Ladias, J. A. (2009). Structural basis for DNA recognition by the human PAX3 homeodomain. *Biochemistry*, 48, 1148–1155.
Case, D. A., Cheatham, T., Daeden, T., Gohlke, H., Luo, R., Merz, K. M., … Woods, R. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26, 1668–1688.
Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., … Merz, K. M. (2010). *AMBER 11*. San Francisco: University of California.
Cillo, C., Faellolla, A., Cantile, M., & Boncinelli, E. (1999). Homeobox genes and cancer. *Experimental Cell Research*, 248, 1–9.
Couloucheri, S. A., Pigis, D. G., Papavassiliou, K. A., & Papavassiliou, A. G. (2007). Hydrogen bonds in protein–DNA complexes: Where geometry meets plasticity. *Biochimie*, 89, 1291–1303.
Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98, 10089–10092.
Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). Verify3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology*, 277, 396–404.
Feig, M., Karanicolas, J., & Brooks III, C. (2004). MMTSB tool set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics & Modelling*, 22, 377–395.
Fraenkel, E., Roul, M. A., Chambers, K. A., & Pabo, C. O. (1998). Engrailed homeodomain–DNA complex at 2.2 A resolution: A detailed view of the interface and comparison with other engrailed structures. *Journal of Molecular Biology*, 284, 351–361.
Gehring, W. J. (1986). On the homeobox and its significance. *BioEssays*, 5, 3–4.
Gehring, W. J. (1987). Homeo boxes in the study of development. *Science*, 236, 1245–1252.
Gorf, A. A., Catlicsch, A., & Jellesarov, I. (2004). The role of flexibility and hydration on the sequence-specific DNA recognition by the Tn916 integrase protein: A molecular dynamics analysis. *Journal of Molecular Recognition*, 17, 120–131.
Gutmanas, A., & Billetter, M. (2004). Specific DNA recognition by the Anp homeodomain: MD simulations of specific and nonspecific complexes. *Proteins: Structure, Function, and Genetics*, 57, 772–782.
Hanes, S. D., & Brent, R. (1989). DNA specificity of the bicom activator protein is determined by homeodomain recognition helix residue 9. *Cell*, 57, 1275–1283.
Hard, T. (1999). NMR studies of protein-nucleic acid complexes: Structures, solvation, dynamics and coupled protein folding. *Quarterly Reviews of Biophysics*, 32, 57–98.
Hirsch, J. A., & Aggarwal, A. K. (1995). Structure of the even-skipped homeodomain complexed to AT-rich DNA: New perspectives on homeodomain specificity. *EMBO Journal*, 14, 6280–6291.
Hovde, S., Abate-Shen, C., & Geiger, J. H. (2001). Crystal structure of the Mx-1 homeodomain/DNA complex. *Biochemistry*, 40, 12013–12021.
Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79, 926–935.
Joshi, R., Passner, J. M., Roeh, R., Jain, R., Sosinsky, A., Crickmore, M. A., … Mann, R. S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell*, 131, 530–543.
LaRonde-LeBlanc, N. A., & Wobberger, C. (2003). Structure of HoxA9 and Pbxb1 bound to DNA: Hexa epipetide and DNA recognition anterior to posterior. *Genes & Development*, 17, 2060–2072.
Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D., & Zakrzewska, K. (2009). Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Research*, 37, 5917–5929.
Lee, J., Kim, J. S., & Seok, C. (2010). Cooperativity and specificity of Cys 2 His 2 zinc finger protein–DNA interactions: A molecular dynamics simulation study. *The Journal of Physical Chemistry B*, 114, 3762–3761.
Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature*, 276, 565–570.
Longo, A., Guanga, G. P., & Rose, R. B. (2007). Structural basis for induced fit mechanisms in DNA recognition by the Pdx1 homeodomain. *Biochemistry*, 46, 2948–2957.
Luscombe, N. M., Laskowski, R. A., & Thornton, J. M. (2001). Amino acid-base interactions: A three-dimensional view of protein–DNA interactions at an atomic level. *Nucleic Acids Research*, 29, 2860–2874.
Manak, J. R., Mathies, L. D., & Scott, M. P. (1994). Regulation of a decapentaplegic midgut enhancer by homeotic proteins. *Development*, 120, 3605–3619.
McGinnis, W., & Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell*, 68, 283–302.
Miyamoto, S., & Kollman, P. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13, 952–962.
Miyazono, K., Zhi, Y., Takamura, Y., Nagata, K., Suigo, K., Kojima, T., & Tanokura, M. (2010). Cooperative DNA-binding and sequence-recognition mechanism of aristalless and clawless. *The EMBO Journal*, 29, 1613–1623.

**Specific vs. nonspecific protein-DNA recognition**

Gehring, W. J. (1987). Homeo boxes in the study of development. *Science*, 236, 1245–1252.
Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proceedings of the National Academy of Sciences, 95*, 11163–11168.

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., … Kollman, P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, 91, 1–41.

Percival-Smith, A., Muller, M., Affolter, M., & Gehring, W. J. (1990). The interaction with DNA of wild-type and mutant fushi tarazu homeodomains. *EMBO Journal, 9*, 3967–3974.

Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C. A., & Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophysical Journal, 92*, 3817–3829.

Pradhan, L., Genis, C., Scone, P., Weinberg, E. O., Kasahara, H., & Nam, H. J. (2012). Crystal structure of the human NKX2.5 homeodomain in complex with DNA target. *Biochemistry, 51*, 6312–6319.

Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., & Mann, R. S. (2010). Origins of specificity in protein–DNA recognition. *Annual Review of Biochemistry, 79*, 233–269.

Ryckaert, J.-P., Cicotti, G., & Berendsen, H. J. C. (1997). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics, 23*, 327–341.

Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Genetics, 17*, 355–362.

Srinivasan, J., Cheatham, T., III, Cieplak, P., Kollman, P., & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramide–DNA helices. *Journal of the American Chemical Society, 120*, 9401–9409.

Treesuwan, W., Wittayanarakul, K., Anthony, N. G., Huchet, G., Alniss, H., Hannonbua, S., … Mackay, S. P. (2009). A detailed binding free energy study of 2:1 ligand–DNA complex formation by experiment and simulation. *Physical Chemistry Chemical Physics, 11*, 10682–10693.

Treisman, J., Gonczy, P., Vashishtha, M., Harris, E., & Desplan, C. (1989). A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell, 59*, 553–562.

Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T., & Pabo, C. O. (1997). Engrailed (Gln50→Lys) homeodomain–DNA complex at 1.9 Å resolution: Structural basis for enhanced affinity and altered specificity. *Structure*, 5, 1047–1054.

Wan, S., Coveney, P. V., & Flower, D. R. (2005). Peptide recognition by the T cell receptor: Comparison of binding free energies from thermodynamic integration, Poisson-Boltzmann and linear interaction energy approximations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 363*, 2037–2053.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research, 35*, W407–W410.

Zhao, X., Huang, X. R., & Sun, C. C. (2006). Molecular dynamics analysis of the engrailed homeodomain-DNA recognition. *Journal of Structural Biology, 155*, 426–437.