

Binding Polymorphism in the DNA Bound State of the Pdx1 Homeodomain

Volodymyr Babin^{1,2}, Dongli Wang³, Robert B. Rose^{3*}, Celeste Sagui^{1*}

1 Center for High Performance Simulations (CHIPS) and Department of Physics, North Carolina State University, Raleigh, North Carolina, United States of America, **2** Department of Chemistry and Biochemistry, UC San Diego, La Jolla, California, United States of America, **3** Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina, United States of America

Abstract

The subtle effects of DNA-protein recognition are illustrated in the homeodomain fold. This is one of several small DNA binding motifs that, in spite of limited DNA binding specificity, adopts crucial, specific roles when incorporated in a transcription factor. The homeodomain is composed of a 3-helix domain and a mobile N-terminal arm. Helix 3 (the recognition helix) interacts with the DNA bases through the major groove, while the N-terminal arm becomes ordered upon binding a specific sequence through the minor groove. Although many structural studies have characterized the DNA binding properties of homeodomains, the factors behind the binding specificity are still difficult to elucidate. A crystal structure of the Pdx1 homeodomain bound to DNA (PDB 2H1K) obtained previously in our lab shows two complexes with differences in the conformation of the N-terminal arm, major groove contacts, and backbone contacts, raising new questions about the DNA recognition process by homeodomains. Here, we carry out fully atomistic Molecular Dynamics simulations both in crystal and aqueous environments in order to elucidate the nature of the difference in binding contacts. The crystal simulations reproduce the X-ray experimental structures well. In the absence of crystal packing constraints, the differences between the two complexes increase during the solution simulations. Thus, the conformational differences are not an artifact of crystal packing. In solution, the homeodomain with a disordered N-terminal arm repositions to a partially specific orientation. Both the crystal and aqueous simulations support the existence of different stable binding conformers identified in the original crystallographic data with different degrees of specificity. We propose that protein-protein and protein-DNA interactions favor a subset of the possible conformations. This flexibility in DNA binding may facilitate multiple functions for the same transcription factor.

Citation: Babin V, Wang D, Rose RB, Sagui C (2013) Binding Polymorphism in the DNA Bound State of the Pdx1 Homeodomain. *PLoS Comput Biol* 9(8): e1003160. doi:10.1371/journal.pcbi.1003160

Editor: Alexander Donald MacKerell, University of Maryland, Baltimore, United States of America

Received: March 5, 2013; **Accepted:** June 13, 2013; **Published:** August 8, 2013

Copyright: © 2013 Babin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by NSF grant MCB-1021883 and NSF grant MCB-0643830. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: bob_rose@ncsu.edu (RBR); sagui@ncsu.edu (CS)

Introduction

Specific DNA binding plays a key role in the protein-DNA recognition process necessary for the regulation of gene expression. Binding determinants are complex, including direct amino acid-base contacts, indirect water-mediated contacts, and local geometry of the DNA sequence [1]. Although it is possible to identify certain trends in the recognition process, such as the formation of point contacts between certain base pairs and certain amino acids, at present there are no unequivocal correspondences between bases and amino acids. Complicating matters, many regions of transcription factors are disordered in solution and fold only upon binding to their specific targets [2]. The current study proposes an additional level of complexity, suggesting that the bound state may also consist of an ensemble of stable conformations instead of a single low energy conformation.

The homeodomain fold provides an interesting example of the subtle effects of DNA-protein recognition. The homeodomain is one of several small DNA binding motifs with limited DNA binding specificity, yet incorporated in an estimated 235 transcription factors it adopts specific and essential developmental roles [3,4,5]. The homeodomain is composed of a 3-helix domain

and a mobile N-terminal arm. Helix 2 and 3 form a helix-turn-helix type motif that is ordered in solution. Helix 3, also known as the recognition helix, interacts with the DNA bases through the major groove. The N-terminal arm, on the other hand, becomes ordered upon binding a specific DNA sequence through the minor groove [6,7,8].

Homeodomain factors participate in a wide range of functions. The current study focuses on Pdx1 (Pancreatic and duodenal homeobox 1), a ParaHox transcription factor evolutionarily related to the Hox subfamily of homeodomains. Hox homeodomains regulate body plan development from *Drosophila* to humans [9,10,11]. Genome-wide binding studies of Hoxa2 and Pdx1 indicate that they may regulate thousands of genes [12,13]. Pdx1 regulates differentiation of the duodenum and stomach, and is a master regulator of pancreas development [14,15,16,17]. In the mature pancreas Pdx1 is expressed in beta- and delta-cells that secrete the endocrine hormones insulin and somatostatin, respectively. Mutations in Pdx1 cause a form of familial diabetes, maturity-onset diabetes of the young type 4 (MODY-4) [18,19].

How homeodomain factors achieve functional diversity as well as exquisite specificity remains a subject of debate. Many studies have correlated DNA binding affinity of homeodomain factors

Author Summary

All organisms require the capability to control gene expression. In eukaryotes, transcription factors play an important role in gene regulation by recognizing specific DNA control regions associated with each gene. The DNA binding domains of transcription factors belong to evolutionarily conserved families with different protein folds. An example is the homeodomain family. Although this DNA binding domain has been studied for a long time, the properties that determine DNA binding specificity are still not clear. We previously showed in a crystal structure that the homeodomain of a transcription factor Pdx1 (Pancreatic and duodenal homeobox 1) binds DNA in 2 different conformations. In this paper, we used Molecular Dynamics simulations to show that both of these conformations are stable in solution. This is surprising since it is often assumed that proteins recognize DNA by finding a single lowest energy state. This study shows that transcription factors may bind DNA in an ensemble of conformations. This scenario may facilitate their finding the correct binding site among the 3 billion basepairs of DNA in the human genome. It may also provide flexibility in the DNA sequence that homeodomains can recognize to promote gene transcription.

with *in vivo* activity [20,21,22,23,24]. DNA binding affinity of Pdx1 monomers accounts for differences in transcriptional activity, at least in cell culture [23]. Hox binding sites include the TAAT core consensus sequence, cooperative binding sites with the TALE homeodomain factors, and sites with no recognizable binding motif [12,13]. Interactions with the TALE factors PBC and Meis alters DNA binding specificity of the Hox homeodomains [25,26,27,28]. Pdx1 cooperates with Pbx1 and Prep1 on the somatostatin promoter, [29,30], Pbx1 and Mrg1 in pancreatic acinar cells [31], and the basic-helix-loop-helix factor E47/NeuroD on the insulin promoter [32,33,34]. Disordered sequences outside of the homeodomain can influence DNA binding specificity suggesting an auto-inhibitory mechanism [35]. Additionally phosphorylation or sumoylation, important for nuclear localization, may affect activity [36,37,38]. Diversity is also achieved through ‘activity regulation’, by recruiting different coactivators and corepressors to non-conserved regions outside of the homeodomain [22,39].

Many structural studies have characterized the DNA binding properties of homeodomains. All Hox factors contact DNA through similar residues (Figure 1A). Residues Ile 47, Gln 50, Asn 51 and Met 54 from the recognition helix insert in the major groove contacting DNA bases directly or through water bridges [40]. Position 50 is particularly important for specificity for some homeodomains, for example Lys 50 in Bicoid [41], but less so for Hox factors as demonstrated by a Gln 50 to Ala mutation [42]. The conservation of the major groove residues suggests they are insufficient to distinguish binding specificity among Hox factors.

The N-terminal arm sequence is less well conserved than the recognition helix, but typically includes positively charged Lys or Arg residues [43,44]. The arm sequence contributes to DNA binding specificity as demonstrated by chimeric homeodomains with swapped N-terminal residues [24,45,46,47]. Even so the N-terminal arm is often disordered in crystal structures of homeodomain monomers bound to DNA. Coarse-grained Molecular Dynamics simulations indicate that the disordered N-terminal arm facilitates searching the DNA for binding sites through electrostatic attraction by a sliding mechanism or transferring between DNA strands by a “fly catching” mechanism [44,48,49,50].

Recently a crystal structure of the Pdx1 homeodomain/DNA complex was obtained in our lab with a consensus DNA binding sequence C₋₁T₁A₂A₃T₄G₅A₆G₇ [51]. The structure contained two complexes with differences in the conformation of the N-terminal arm, major groove contacts, and backbone contacts, raising new questions about the DNA recognition process by homeodomains (Figure 1 B,C) [51]. At the time we attributed the differences in the two conformations to differences in DNA bending as a result of crystal packing [51]. We proposed an induced fit model in which DNA contacts by residues from helix 3 in the major groove of one conformation stabilized the N-terminal arm in the minor groove.

In this work we apply classical Molecular Dynamics (MD) with a fully atomistic representation of the complex and solvent to simulate both the crystal and solution behavior of both conformations of the Pdx1 homeodomain/DNA complex [51]. In the last decade MD simulations have become an invaluable tool to complement structural information obtained experimentally [52,53,54,55]. MD simulations of the Pdx1/DNA complexes show that differences in DNA contacts persist between the two conformations even in solution due to distinct positioning of the homeodomain relative to the DNA. Conformation A represents a less specific complex than Conformation B. The simulations suggest that one source of diversity of homeodomain function derives from different bound states with different degrees of DNA specificity. The existence of these “isomeric” bound conformations has not been reported before. We propose that multiple bound isomers are an important feature of the homeodomain/DNA binding processes, adding another layer of complexity to what is known about binding specificity.

Materials and Methods

Simulation details

Simulations were carried out for: (i) the crystal unit cell; (ii) aqueous solution; and (iii) the DNA and Pdx1 molecules separately. Initial geometries for the simulations were derived from both Pdx1/DNA complexes in the asymmetric unit of the crystal structure (pdbid 2H1K) (www.rcsb.org) [51,56]. Residues missing in the crystal structure (model A: residues 1–3, 60–61; model B: residues 58–61) were placed in low energy conformations by superimposing short pre-equilibrated peptide fragments onto the experimental structure. For the solution simulations all waters from the crystal structure were removed and replaced with solvent water molecules surrounding the protein/DNA complex.

For the crystal simulation the unit cell of the crystal structure was generated from the asymmetric unit by applying the P212121 symmetry operators. Non-crystallographic water molecules were added by sampling them from a box of water equilibrated at constant pressure and temperature conditions and placed “on top” of the unit cell. Specifically, molecules were picked at random from the water box and “copied” into the unit cell provided no sterically forbidden configuration results. The number of the water molecules was varied until the system’s density remained unchanged in trial MD runs under normal conditions. The density settled at 1.25 g/cm³ with 6901 non-crystallographic waters.

Simulations were performed using the AMBER 10 package along with some “in house” codes [57]. The Pdx1/DNA complex was modeled using the ff99SB [58] force field for protein, parmbsc0 [59] for DNA, TIP3P [60] for water molecules and the AMBER stock 1999 version of the Cornell force field [61] for Na⁺ and Cl⁻ ions. Missing hydrogen atoms were added with the LEaP module of AMBER 10; histidine residues were assumed to be

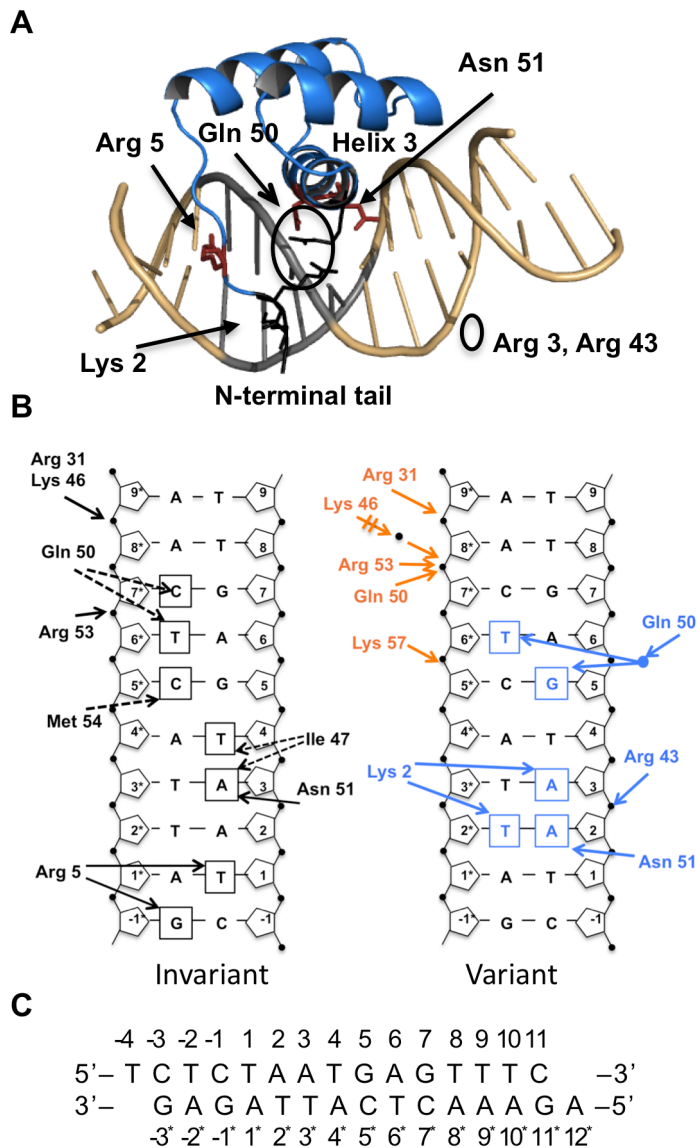


Figure 1. Pdx1 homeodomain/DNA interactions from the crystal structure. A) Structure of the Pdx1 homeodomain/DNA complex. Pdx1 (blue ribbon) binds the TAAT core DNA sequence (grey). The N-terminal tail binds in the minor groove, and the recognition helix, helix 3, binds in the major groove. Key residues contacting the DNA are shown as stick figures (red): Arg 5 in the minor groove, and Asn 51 in the major groove. Gln 50 contacts the phosphate backbone or the DNA bases through a water-mediated contact. Arg 3 and Arg 43 (black line representation, circled) help stabilize the N-terminal arm, and Lys 2, in the minor groove when helix 3 is properly positioned in the major groove. B) Hydrogen bond contacts with the DNA differ between Conformation A and B in the Pdx1 homeodomain/DNA crystal structure (PDB 2H1K) (www.rcsb.org) [56]. In both conformations (left) Arg 5 contacts Thy 1 and Gua -1^* through the minor groove, and Asn 51 contacts Ade 3 through the major groove. The difference in the DNA contacts between Conformations A (orange) and B (blue) is shown on the right. Conformation B makes additional base contacts by Asn 51, by Lys 2 from the ordered N-terminal arm, and a water-mediated contact by Gln 50. Conformation A forms additional phosphate contacts. Arrows represent hydrogen bonds. C) DNA sequence and numbering in the crystal structure. The TAAT core sequence is in bold. doi:10.1371/journal.pcbi.1003160.g001

neutral with a proton at the $\epsilon 2$ position. For the solution simulations, at least a 15 Å thick layer of TIP3P water was added around the solute with the LEaP module, and the system was neutralized using either Na^+ (Pdx1/DNA complex) or Cl^- (protein alone) ions. The structures were thoroughly equilibrated with the SANDER module of AMBER 10 before the production (data gathering) step. During the initial equilibration the heavy DNA and protein atoms were restrained to their initial positions by a harmonic potential. In every case we first performed a few conjugate gradient minimization steps to relax the hydrogens, followed by a 1 nanosecond long constant pressure run at ambient

conditions ($T = 298 \text{ K}$, $P = 1 \text{ atm}$). The last frame of the restrained NPT runs were the starting point of the *unrestrained* NPT production runs reported in this paper.

The production simulations were carried out using the PMEMD module of AMBER 10. The electrostatic interactions were evaluated by the PME method [62,63] using a 9 Å cutoff for the short-range terms. The same cutoff was used for the van der Waals terms with a continuous correction for the long-range terms. The lengths of all bonds that involve hydrogen atoms were fixed via the SHAKE algorithm with the tolerance set to 10^{-6} Å. Langevin dynamics [64] with collision frequency $\gamma = 1 \text{ ps}^{-1}$

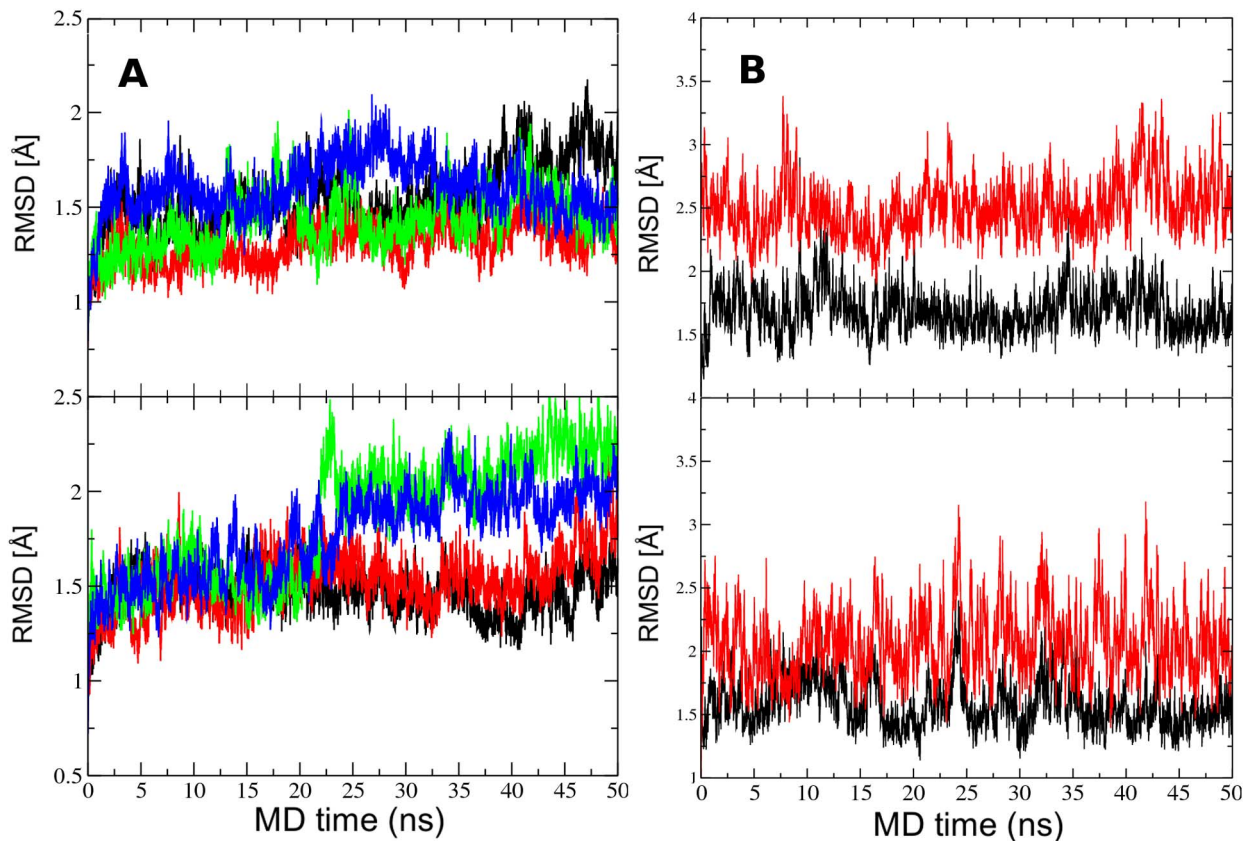


Figure 2. Agreement of simulations with the crystal structure. A) Stability of the crystal simulation. Mass weighted RMSD relative to the crystal structure (PDB 2H1K) (www.rcsb.org) for the eight Pdx1/DNA complexes comprising the unit cell of the 2H1K crystal during unrestrained molecular dynamics in the crystal environment. Conformation A is shown in the top panel and conformation B in the bottom panel. Different colors correspond to the different asymmetric units as specified in Figure S1. B) Stability of the solution simulation. Mass weighted root mean square deviation of the Pdx1/DNA complex computed with respect to the crystal Conformation A (black) and crystal conformation B (red) starting from Conformation A (top panel) and starting from Conformation B (bottom panel). Both simulations were closer to the crystal Conformation A. The two overhanging DNA bases in the crystal structure were excluded from the simulation.
doi:10.1371/journal.pcbi.1003160.g002

(ps = picosecond) was used to maintain the temperature at 298 K with a different random number generator seed set for every run. The Berendsen algorithm [65] with relaxation time $\tau_P = 1$ ps was used to maintain the pressure at 1 atm. The total time for each simulation was 50 ns with a time step of 2 femtoseconds and coordinates saved for analysis every 10 ps (5000 steps).

Analysis

The PTRAJ module of AMBER 10 was used for basic analysis (centering and imaging of the trajectories, computations of RMS deviations, etc.), 3DNA (v. 1.5) [66] for the calculation of DNA structural parameters, and simple in-house programs for the identification and counting of the intermolecular contacts. The latter were defined as follows [67]: a hydrogen bond was assumed if the distance between the donor hydrogen and the acceptor oxygen or nitrogen was 2.8 Å or less, and the angle formed by the donor, hydrogen and acceptor atoms exceeded 145°; a hydrophobic contact was defined as a pair of sulfur/carbon atoms separated by less than 4.5 Å; a water contact was identified if the oxygen of a water molecule was within 3 Å of a nitrogen or an oxygen atom. A simultaneous water contact from two different macromolecules to the same water molecule is referred to as “water bridge”.

Figures of protein structures were generated with Pymol [68] and labeled in Powerpoint (Microsoft Office). The molecular graphics image of the unit cell was produced using the UCSF

program Chimera [69]. The figures displaying distances through the simulation are displayed as a running average of 100 ps (10 trajectory frames).

Results

Conformation-specific DNA contacts in the crystal structure

The two conformations of the Pdx1/DNA complex in the crystal structure contained invariant contacts found in both conformations A and B, and variable contacts specific to each conformation (Figure 1B) [51]. Two residues formed direct hydrogen bonds with DNA bases in both conformations: Asn 51 with Ade 3 (CTAA₃T) in the major groove, and Arg 5 with Thy 1 (C₋₁T₁AAT) and Gua -1* (opposite Cyt -1) in the minor groove (Figure 1A). Conformation B was more specific than Conformation A. In Conformation B, Gln 50 formed a water-mediated contact with Gua 5 and Thy 6* (TAATG₅A₆), and Asn 51 contacted Ade 2 in addition to Ade 3. The N-terminal arm was also more ordered in conformation B, with Lys 2 hydrogen bonded with the bases Ade 3 and Thy 2* in the minor groove.

Crystal simulation

The flexibility of the interactions in the two conformations was investigated by MD of the crystallographic unit cell. During the

simulation the four copies of model A and model B in the unit cell (Figure S1) showed some variation in mobility and conformation, probably due to fluctuating differences in their local instant environment. This agrees with the crystallographic information where the same fluctuations are likely responsible for the high B-factors.

The experimental molecular geometries were well preserved during the simulation (Figure 2A). The instantaneous mass-weighted root-mean-square deviations (RMSDs) with respect to the crystal structure were 2.5 Å or less. The average structure over all times and over the four replicas of each molecule was computed and its RMSD with respect to each of the crystal conformations was calculated. Note that the RMSD of the *average structure* is not the same as the average of the instantaneous RMSDs. (The average structure is of course a better approximation of the experimental structure than the various instantaneous structures). The RMSD of the *average structures* with respect to the original 2H1K coordinates were 0.86 Å and 1.10 Å for conformations A and B, respectively. We interpret this as validation of the model, the force field and the simulation protocol.

Trajectories of the Pdx1/DNA complex in the crystal

In general the differences between conformations A and B were less pronounced after the crystal simulation. Arg 5 is the one residue to hydrogen bond consistently with the same bases in all 8 models in the unit cell, to Thy 1 and Gua -1* through the minor groove, as it does in the crystal structures. The major groove contacts are more variable. The hydrogen bond by Asn 51 with Ade 3 (CTA₂A₃T) is lost consistently in Conformation A, while it is more stable in Conformation B (Figure S2A,B). Both Asn 51 and Gln 50 contact the phosphate backbone of the DNA in the major groove of Conformation A, with Ade 2 and Cyt 7*, respectively (Figure S2C). Only the Gln 50- Cyt 7* contact is accessible to Conformation B. These backbone contacts are characteristic of the partially specific Conformation A after the solution simulation.

In the crystal structure five phosphate contacts are unique to Conformation A by residues from helix 2 and 3: Arg 31, Lys 46, Gln 50, Arg 53 and Lys 57 (Figure 1B) [51]. During the simulation all of these contacts are also formed in conformation B except Arg 31 and Lys 46 with the phosphate backbone of Ade 8* (Figure S2D,E). Arg 31-Ade 8* is unique to Conformation A in solution too.

While Arg 5 is consistently ordered in all homeodomain/DNA complexes, the residues N-terminal of Arg 5 are often disordered [7,70,71,72]. In the crystal structure, these residues are ordered in Conformation B and disordered in Conformation A. During the simulation Lys 2 remains predominantly in the minor groove in Conformation B, hydrogen bonded with Thy 2* O2P (Figure 3A) but not Ade 3. In model B4 residues 1–4 of the N-terminal arm escape from the minor groove after about 20 ns and remain mobile. In Conformation A Lys 2 never enters the minor groove during the simulation. Interestingly, Arg 3 does enter the minor groove to contact Thy 2* for about 20 ns in model A2, and in model A4 (at 30–50 ns) and at the end of the simulation in model A3. (Figure 3B). Mobility of the N-terminal arm appears to be required for Arg 3 to enter the minor groove since the arm executes large motions in models A2 and A4, while these motions are restricted in models A1 and A3 by phosphate backbone contacts by Lys 2 or the acetylated N-terminus.

From the crystal structure we proposed that ordering of Lys 2 in the minor groove is stabilized by a network of contacts between Arg 43 and His 44 from helix 3 in the major groove and Arg 3 in the minor groove (Figure 1A) [51]. These interactions were maintained in conformation B during the crystal simulation: both

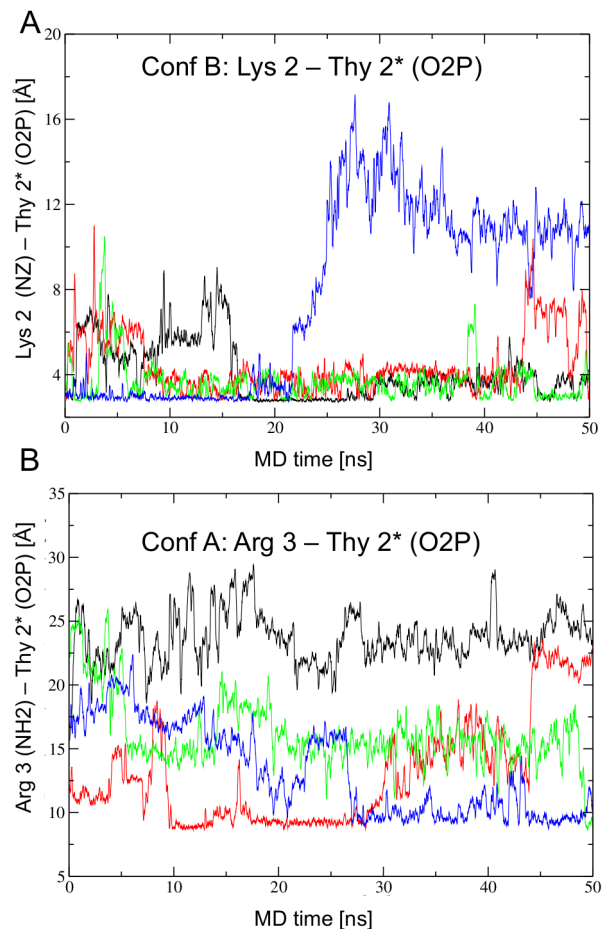


Figure 3. N-terminal arm contacts in the crystal simulation. A) In Conformation B, the N-terminal arm is mostly ordered with Lys 2 hydrogen bonding with the base of Thy 2* O2P. In model B4 (blue) the N-terminal arm escapes from the minor groove. B) The N-terminal arm in Conformation A starts the simulation disordered. Lys 2 never enters the minor groove, but Arg 3 enters the minor groove in model A2 (red), A4 (blue), and it seems to do so at the end of the simulation in A3 (green). The colors represent one of the four asymmetric units as depicted in Figure S1: A1, B1 black; A2, B2 red; A3, B3 green; A4, B4 blue.
doi:10.1371/journal.pcbi.1003160.g003

Arg 3 and Arg 43 hydrogen bond with the phosphate backbone, with Thy 4 and Ade 3, respectively (Figure S3 A,B). The proximity of the guanidinium groups of Arg 3 and Arg 43 suggest pi-pi stacking. His 44 stabilizes the conformation of Arg 43 (Figure S3C). In model B4, after the N-terminal arm escapes the minor groove, the Arg 3-Thy 4 O2P and Arg 43-His 44 contacts are broken, consistent with their role in stabilizing the N-terminal arm in the minor groove.

In Conformation A the N-terminal 3 residues and the side chain of Arg 43 are disordered in the crystal structure. Arg 43 never associates stably with His 44 (Figure S3D) or Ade3 O2P, but forms a stable hydrogen bond (~60% of the time) with Thy 4 O2P in model A2 and A3 (Figure S3E). In model A2, the Arg 43-Thy 4 O2P backbone contact correlates with insertion of Arg 3 in the minor groove to contact the base of Thy 2* (Figure 3B).

In summary, the simulation reduces somewhat the differences between Conformations A and B found in the crystal structure, particularly in the major groove. Three phosphate contacts are specific to Conformation A: by Asn 51 with Ade 2, and by Arg 31

and Lys 46 with Ade 8*. The contact by Arg 43 from the major groove with the phosphate backbone correlates with stabilizing the N-terminal arm. In Conformation B the N-terminal arm is mostly ordered with Lys 2 binding in the minor groove. In Conformation A the N-terminal arm is mostly disordered. Arg 3 enters the minor groove in models A2 and A4, suggesting a second position for the N-terminal arm not present in the crystal structure. We attributed the different contacts between the two conformations in the crystal structure to differences in DNA bending due to crystal packing [51]. The DNA bending of the crystal structure is maintained during the crystal simulation.

Pdx1/DNA complex in aqueous solution

Simulations of the Pdx1/DNA complex in aqueous solution were initiated from both conformations reported in the 2H1K PDB (www.rcsb.org) [56] structure, and trajectories recorded for 50 ns. Mass-weighted RMSDs were calculated relative to the crystal Conformation A or the crystal Conformation B (Figure 2B). After an initial relaxation time, the simulations of both Conformation A and B resembled Conformation A more than B, indicating that Conformation A in the crystal structure, with the less bent DNA, is closer to the solution conformation. The average of the *instantaneous* RMSD values for complexes A and B relative to the experimental structure in Conformation A were 1.69 Å and 1.58 Å (black lines in top and bottom panels of Figure 2B), respectively; and relative to Conformation B were 2.51 Å and 2.08 Å, respectively (red lines in top and bottom panels Figure 2B).

Specific DNA contacts in the major groove by conformation B

The flexibility of the DNA during the simulation of both conformations resulted in an average straight helical axis with large fluctuations in the bending angle (not shown). We were therefore surprised that differences between the conformations persisted throughout the simulation. In conformation A, both Gln 50 and Asn 51 stably contacted the phosphate backbone in the major groove, with Cyt 7* and Ade 2, respectively (Figure 4). In contrast Asn 51 in conformation B did not contact the phosphate backbone of Ade 2 (Figure 4B) but formed a direct hydrogen bond with Ade 3 N7 (Figure 5 A,B). Periodically Asn 51 OD1 formed a second specific hydrogen bond with Ade 3 N6 (not shown). Gln 50 was too far from the DNA for a direct contact with the DNA bases, but during the simulation two water molecules sometimes (~20% of the time) bridged between Gln 50 and Asn 51 and the bases of Thy 4, Gua 5 and Thy 6*. Water 1 (W1 in Figure 5A) also bridged between Gln 50 and Asn 51. These direct DNA contacts indicate that helix 3 continues to form more specific major groove contacts in Conformation B than in A.

Ordering of the N-terminal arm

The contacts by Arg 5 with Gua -1* and Thy 1 through the minor groove are conserved in the trajectories of both conformations (Figure 6). This remains the only direct hydrogen bond with a DNA base in Conformation A. The N-terminal residues 1–3 are initially ordered in conformation B, and Lys 2 continues to form a hydrogen bond with Thy 2* O2 for 35 ns of the simulation (Figure S4A). After that the N-terminus of Pdx1 moves outside of the minor groove, but Arg 3 and Arg 43 remain in contact with the DNA phosphate backbone, stabilizing the N-terminal arm (Figure S4 B–D). It is therefore plausible that Lys 2 would return to the minor groove in a longer simulation. In contrast to the crystal simulation, Arg 43 contacts Thy 4 O2P instead of Ade 3 O2P when the N-terminal arm is ordered in Conformation B (Figure S4

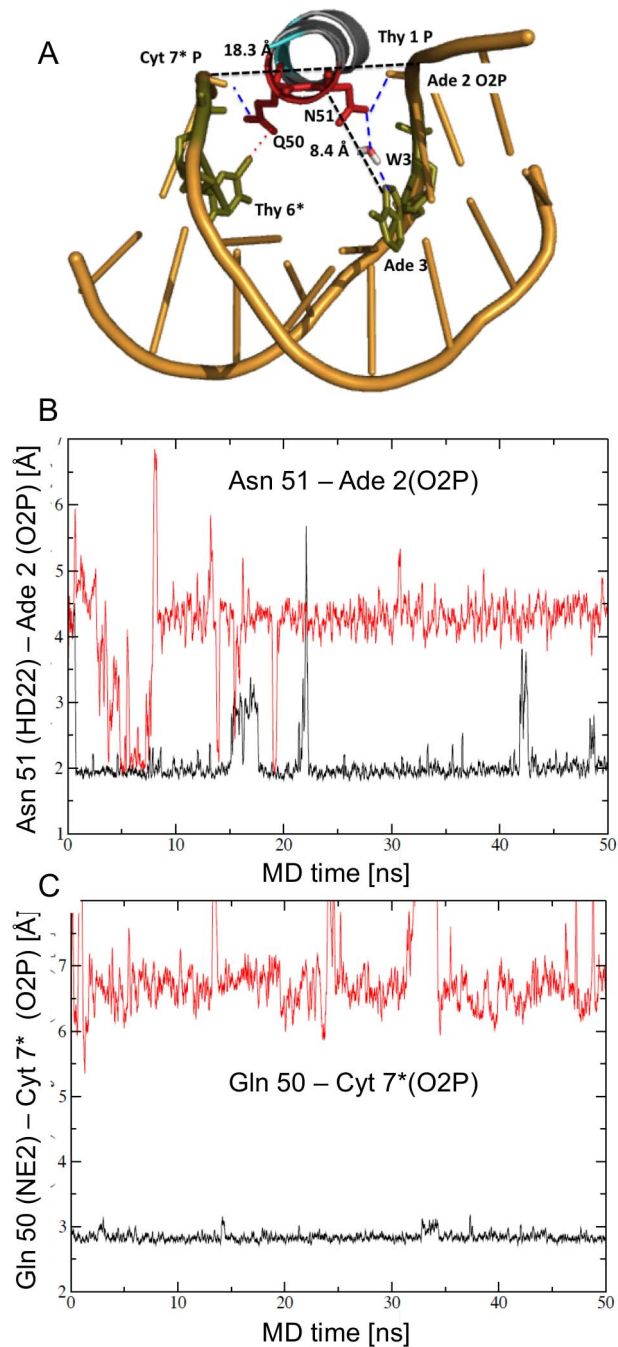


Figure 4. Contacts by helix 3 in the major groove of Conformation A during the solution simulation. A) Ribbon diagram looking into helix 3 and the major groove. Gln 50 and Asn 51 (labeled as Q50 and N51) contact the phosphate backbone only, with Cyt 7* and Ade 2, respectively (blue dashed lines). Gln 50 is within van der Waals contact of Thy 6* C7 (green, connect by a red dotted line). About 7% of the time a water molecule (W3) mediates contact between Asn 51 and Ade 3 (green). The position of helix 3 in the major groove is measured by the distance between Asn 51 C-alpha to Ade 3 N7 (8.4 Å), and the width of the major groove: Thy 1 P – Cyt 7* P (18.3 Å). The structure represents interactions at 30 ns of the simulation. B) Asn 51 contacts the backbone of Ade 2 O2P only in Conformation A (black), not in Conformation B (red). C) Gln 50 contacts the phosphate backbone of Cyt 7* O2P only in Conformation A (black), not in Conformation B (red). doi:10.1371/journal.pcbi.1003160.g004

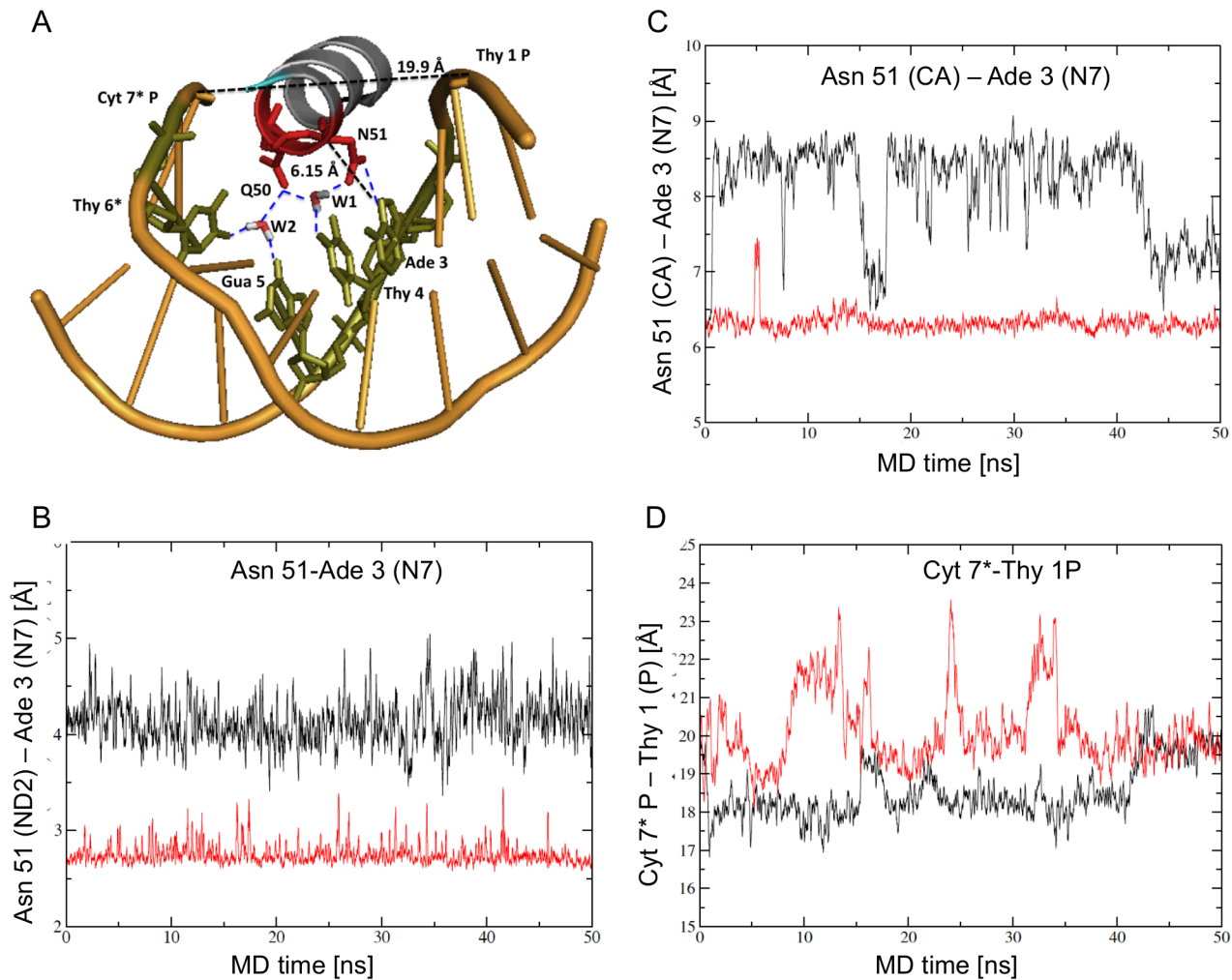


Figure 5. Contacts by helix 3 in the major groove of Conformation B during the solution simulation. A) Ribbon diagram looking into helix 3 and the major groove. Asn 51 contacts Ade 3 directly (blue dotted line). Gln 50 makes no direct contact with the DNA. About 20% of the time, water mediated contacts bridge between Gln 50 and Asn 51 (W1) and Gln 50 and DNA bases (in green) (W2). Distances show helix 3 binds farther in the major groove (Asn 51 C-alpha to Ade 3 N7 distance 6.15 Å), and the major groove is slightly wider than in Conformation A (Thy 1 P – Cyt 7* P distance 19.9 Å). B) Asn 51 forms a direct hydrogen bond with the base Ade 3 N7 only in Conformation B (red), not in Conformation A (black). C) and D) The position of the homeodomain differs in Conformation A and Conformation B during the solution simulation. C) Helix 3 binds closer to the DNA in Conformation B (red) than Conformation A (black), as measured by the distance between Asn 51 and Ade 3. D) The major groove is wider in Conformation B than Conformation A during most of the solution simulation, as measured by the Cyt 7* P-Thy 1 P distance. This is consistent with Pdx1 binding deeper in the DNA major groove in Conformation B. doi:10.1371/journal.pcbi.1003160.g005

B,E, Figure S3A). His 44 does not contact Arg 43 during the solution simulation, unlike in the crystal simulation (Figure S3C).

In conformation A, Lys 2 begins the simulation outside of the minor groove and does not enter during the simulation. As seen in the crystal simulation, Arg 3 in Conformation A enters the minor groove towards the end of the simulation, but it never settles in a single position, contacting the base of Ade 3 only transiently. The residues that stabilize the N-terminal arm, Arg 3 and Arg 43, are more mobile in conformation A than conformation B (Figure S4 B–D). Arg 43 contacts Ade 3 only about a third of the trajectory (Figure S4E).

Specific phosphate contacts

All five hydrogen bonds that were specific to Conformation A in the crystal structure were accessible to Conformation B. The Arg 31-Ade 3* contact is favored in Conformation A. The position of Arg 31 is stabilized through a hydrogen bond with Glu 42 in Conformation A (Figure S4F, Figure 6A,B). The contact between

Tyr 25 and Thy 6* phosphate is restricted to Conformation B in the solution simulation (Figure S4G, Figure 6C,D). This contact was accessible to both conformations in the crystal structure.

Position of the homeodomain with respect to the DNA

Clearly different contacts persist between conformations A and B through the 50 ns of the solution simulation. As mentioned, the DNA is highly flexible during the simulation indicating DNA bending cannot explain the conformational differences. Instead the overall positioning of the homeodomain of Pdx1 relative to the DNA differs for the two conformations, as indicated by the distance between Asn 51 CA and Ade 3 N7 (Conformation A : 8.0 ± 0.7 Å; conformation B: 6.4 ± 0.4 Å) (Figure 5C) and the width of the major groove, measured as the distance between the phosphate of Cyt 7* and Thy 1 (Conf A 18.6 ± 0.8 Å, Conf B 20.1 ± 1.0 Å, defined from the atom centers without subtracting 5.8 Å for the phosphorous van der Waals radius) (Figure 5D).

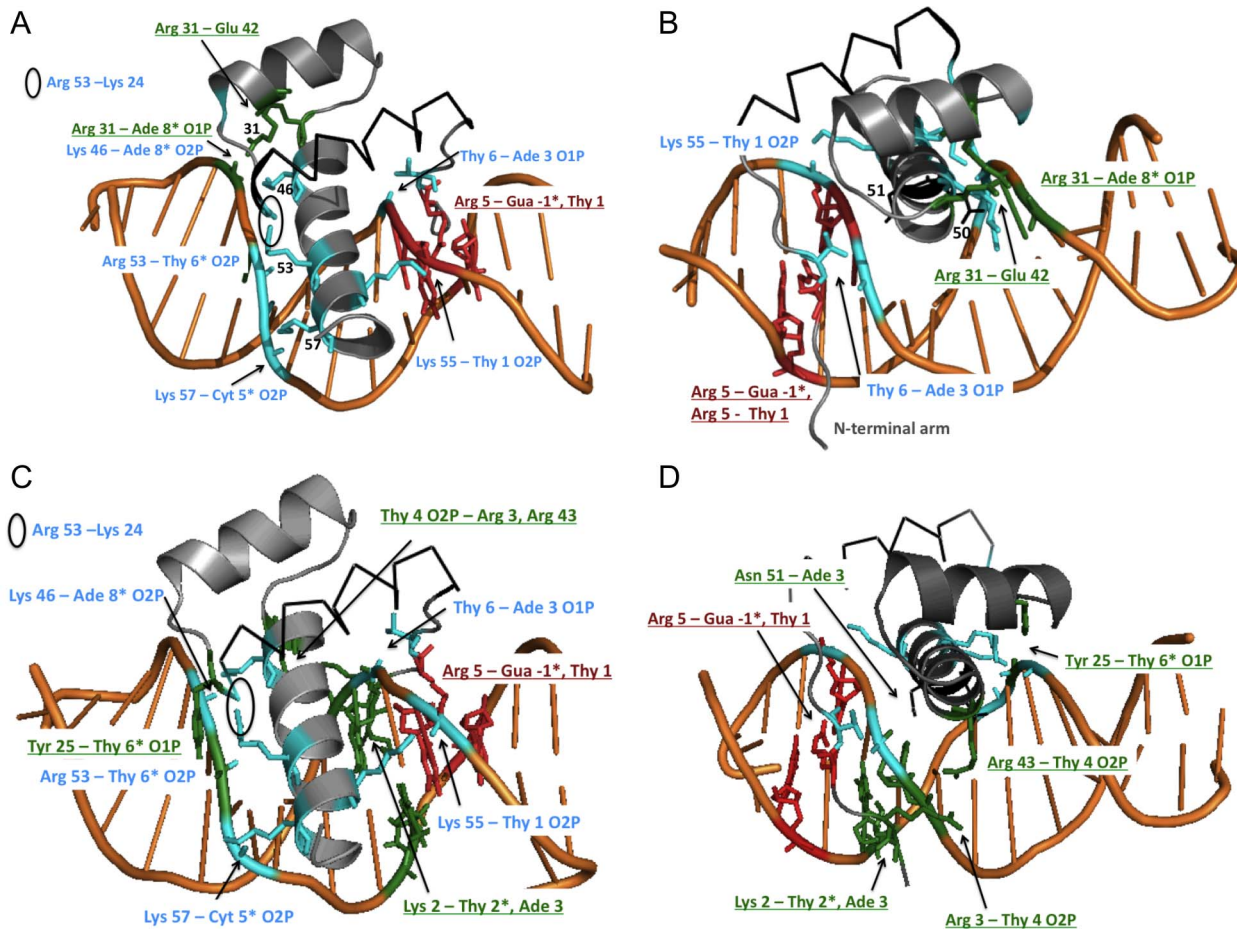


Figure 6. Interactions specific to Conformation A and B after 30 ns of the solution simulation. A) and B) Conformation A. C) and D) Conformation B. The base contacts by Arg 5 are identical in both conformations (red, underlined). Invariant (in Conformation A and B) contacts with the phosphate backbone include (cyan): Lys 46 – Ade 8* O2P, Arg 53 – Thy 6* O2P, Lys 57 – Cyt 5* O2P, Lys 55 – Thy 1 O2P, and Thy 6 – Ade 3 O1P. The intramolecular hydrogen bond Arg 53 – Lys 24 is also conserved (circled). A) Hydrogen bond interactions by Pdx1 Conformation A (grey ribbon) with the DNA during the solution simulation, facing helix 3 in the major groove. Contacts unique to Conformation A (green, underlined) include a phosphate contact by Arg 31 with Ade 8*, in the major groove opposite the N-terminal arm, and the intramolecular contact between Arg 31 and Glu 42. B) Conformation A viewed looking into the minor groove. Residues 1–4 of the N-terminal arm are highly mobile in the solution simulation. Asn 51 and Glu 50 are shown in black lines. C) Hydrogen bonds by Pdx1 Conformation B (grey) with the DNA during the solution simulation, facing helix 3 in the major groove. Contacts unique to Conformation B (green, underlined) include a phosphate contacts by Tyr 25. D) Conformation B viewed looking into the minor groove. Arg 3 and Arg 43 hydrogen bond with the backbone of Thy 4 (green, underlined), assisting in stabilizing the N-terminal arm in the minor groove and the interaction by Lys 2 with the bases of Thy 2* and Ade 3. doi:10.1371/journal.pcbi.1003160.g006

Helix 3 is bound deeper in the major groove in conformation B allowing Asn 51 to contact Ade 3 N7 directly, which may account for the wider major groove (Figure 4A, 5A).

The difference in the positioning of the homeodomain between the conformations was not apparent in the crystal structure: the distance between Asn 51 CA and Ade 3 N7 was about 6.2 Å for both conformations. The major groove width (Cyt 7* P – Thy 1 P) was different: 19.4 Å in Conformation A and 20.8 Å in Conformation B. During the crystal structure simulation, the distance between helix 3 and Ade 3 varies between 6 and 8 Å in conformation A, and between 6 and 7 Å in conformation B. The constraints of the crystal packing therefore prevented repositioning of the homeodomain.

Discussion

The MD simulation

MD has been applied to DNA/homeodomain complexes previously to study protein-DNA and water mediated contacts

[73], the role of salt bridges [74], the role of residue 50 [75,76], folding properties of the N-terminal arm [77], and other studies [67,78,79,80,81]. In general these simulations are initiated from a unique structure, assuming that the simulation will explore the relevant conformational space. In the current study we applied MD to investigate two distinct DNA binding conformations of the Pdx1 homeodomain to determine if the differences were the result of crystal packing. Simulations were carried out in the context of a crystal unit cell and in solution. The solution simulations generated two different conformations of the Pdx1/DNA complex depending on the initial conformation derived from the crystal structure. Both conformations were stable during the 50 ns simulation. The current study demonstrates the real possibility of multiple stable conformations that are not accessible during limited simulation times.

The AMBER force field ff99SB [57,58] used in the simulations reported in this work is considered state-of-the-art, and includes several refinements for DNA simulations [59,82,83,84]. Present

computer capabilities allow fully atomistic simulations, minimizing artifacts. The DNA and protein in these simulations are fully solvated with explicit waters; in a relatively large box under periodic boundary conditions (as opposed to spherical water clusters that may experience various surface potential discontinuities at the cluster-vacuum or cluster-continuum interface [85,86]); with a correct treatment of electrostatics [87,88]. Crystal simulations at constant pressure and temperature (NPT) that reproduce the crystallographic cell and symmetries have traditionally been used to test and tune force fields, as they allow direct comparison with experiments, and will also reproduce packing effects [52,53,54,55,89,90,91,92].

Two stable conformations of Pdx1 bound to DNA

After the solution simulation Conformation B bound DNA specifically while Conformation A bound with limited specificity. The unique interactions in the two conformations were due to different positions of the homeodomain in the major groove of the DNA, with helix 3 buried deeper in the major groove in Conformation B than in Conformation A (Figure 4A and 5A). In Conformation B Asn 51 interacts directly with Ade 3 in the major groove, and Lys 2 of the N-terminal arm contacts bases through the minor groove. The proximity of helix 3 to the DNA in Conformation B facilitates ordering bridging water molecules between the protein and DNA, with Gln 50 and Asn 51 (Figure 5A). These bridging water molecules were not observed in the Pdx1 crystal structure, but were observed in the related Antennapedia structure [71].

What determines the position of helix 3 in the major groove? We previously attributed the presence of two Pdx1/DNA conformations in the crystal structure to the curvature of the DNA in conformation B [51]. The differences between the two conformations identified in the crystal structure diminished during the crystal simulation, despite maintaining the average curvature of the DNA in conformation B. In contrast, differences between the two conformations increased during the solution simulations. A comparison of the Antp homeodomain/DNA complex by NMR and crystallography indicated contacts by Arg 43 with Ade 3, and movements of Gln 50 and Asn 51 in the NMR structure that could not be explained from the crystal structure [71]. These contacts are consistent with the motions of the Pdx1 homeodomain in the solution simulation. Clearly the crystal lattice constrained the Pdx1/DNA conformation, suggesting caution when interpreting crystal structures of protein/DNA complexes.

What properties of the two conformations in the crystal structure directed the solution simulations toward the specific (starting from Conformation B) versus less specific (starting from Conformation A) complexes? The average DNA sequence was straight during the solution simulation of both conformations, indicating that DNA bending was not the primary cause. In the crystal structure helix 3 was oriented at slightly different angles relative to the DNA in the two conformations. The specific phosphate contacts formed by Conformation A in the crystal structure are accessible to Conformation B during the solution simulation. Already in the crystal structure the contacts are less specific in Conformation A. The configuration that defines Conformation A includes: Gln 50 contacting the phosphate backbone at base 7*, Asn 51 contacting Ade 3 but not Ade 2, and the disordered N-terminal residues (Figure 1B). The contacts that define Conformation B include: Asn 51 contacting Ade 2 and Ade 3 in the major groove, Gln 50 making a water mediated contact with DNA bases at positions 5 and 6, and the ordered N-terminal arm with Lys 2 contacting Thy 2* and Ade 3.

The DNA “bound” state consists of multiple conformations

The MD simulations presented here suggest multiple conformations are possible for the N-terminal arm in the minor groove and for the helix-turn-helix domain in the major groove. In both conformations, Arg 5 contacts Gua -1* and Thy 1 through the minor groove (Figure 6). The most stable (longest-lived) configuration for the N-terminal arm of Pdx1 consists of Lys 2 inserted in the minor groove and Arg 3 outside of the minor groove contacting the phosphate backbone and Arg 43 (Figure S4A–D). In Conformation A, Arg 3 inserts in the minor groove and contacts the base Thy 2* for some time in the crystal simulation (Figure 3B). This configuration resembles the configuration in the Scr-Exd DNA complex (the *Drosophila* homolog of Hox5-Pbx1) with a 14 residue N-terminal extension of Scr, including the YPWM Pbx1 binding motif [25]. In that structure the N-terminal arm was ordered with Arg 3 inserted in the minor groove but contacting the phosphate backbone. The authors suggested that Arg 3 is positioned by a His residue along the N-terminal extension. Therefore while binding of the Pdx1 monomer may favor base contacts by Lys 2 in the minor groove, other protein interactions may favor Arg 3 positioned in the minor groove.

The MD simulation also distinguishes two orientations of the helix-turn-helix domain in the major groove. An alternate orientation of the recognition helix was previously characterized for the Mata2 homeodomain bound to a nonspecific DNA sequence [93]. In this structure the homeodomain was rotated with respect to the consensus binding site, altering interactions in the major groove and eliminating contacts by the N-terminal arm in the minor groove. A second paper noted that the Hox homeodomains in the HoxA9-Pbx1 and HoxB1-Pbx1 complexes were oriented differently in the major groove, altering base contacts [94,95]. In contrast to these examples, *the two conformations of the Pdx1 homeodomain are bound to the same DNA sequences of the consensus-binding site*. In the less specific Conformation A of Pdx1, Arg 5 makes base-specific contacts through the minor groove, like the specific conformation. Many of the same phosphate contacts position helix 3 in the major groove, by Thy 6, Arg 31, Arg 53, Lys 55, and Lys 57. But helix 3 of Conformation A is too far from the DNA bases to form direct hydrogen bonds; instead Gln 50 and Asn 51 contact the phosphate backbone (Figure 4A).

One interpretation of the partially specific Conformation A is that it represents a DNA binding intermediate in search of the specific DNA binding conformation B. The Pdx1 homeodomain binds nonspecific DNA with just 20-fold lower affinity than the consensus site [23,51]. Other homeodomains also bind DNA with low specificity, as noted for Mata2 and Antennapedia [96,97]. The stability of the less-specific Conformation A during the MD simulation suggests it might be populated when the Pdx1 monomer binds nonspecific DNA sequences.

Pdx1 binds to thousands of DNA sites *in vivo*, as measured by ChIP-Seq, including sequences distinct from the consensus binding sequence [13,98]. In binding a specific DNA sequence, both conformations A and B may be present as two of an ensemble of DNA-bound conformations. In this scenario the DNA sequence and other protein interactions stabilize a subset of this ensemble. The diversity of interactions might explain the myriad of functions accomplished by Pdx1.

Coordination between the major and minor groove

In our previous paper we proposed that Arg 43 and Arg 3 bridge between the major and minor grooves to order the N-terminal arm in Conformation B, suggesting some synergistic interactions between the helical and N-terminal domains. Many

studies conclude that the N-terminal arm contributes to DNA binding specificity of homeodomains [24,45,46,47,99,100,101]. Synergy between the major and minor groove has been noted for chimeric homeodomains, which generally require mutations in the N-terminal arm and the recognition helix to change specificity between homeodomain factors [24,99]. In a survey of all *Drosophila* homeodomains, specificity determinants for DNA binding originate from both the recognition helix and N-terminal residues [20,102].

Like other Hox factors, Pdx1 binds DNA cooperatively with PBC class homeodomains, such as Pbx1 [30]. Extensions of the N-terminal arm to the “YPWM” motif enhance DNA binding specificity of Hox factors, exposing “latent specificity” among the eight Hox paralogs through interactions with Pbx1 [5,27,103]. But minor groove contacts do not explain all of the sequence preferences observed. For example a comparison of two structures by the *Drosophila* Hox-Pbx1 heterodimer Scr-Exd bound to different DNA sequences demonstrated conformational changes in the extended N-terminal linker as well as contacts in the major groove [25]. In the context of Pbx1, the consensus-binding site for the Hox factors is generally not TAAT, necessitating different DNA interactions by the N-terminal arm and recognition helix. The MD simulations reported here suggest that the DNA and protein context may promote “specific binding” by restricting the ensemble of accessible conformations available to the homeodomain on the DNA.

Even though longer MD simulations are needed to probe “rare” conformational transitions and to completely characterize the relative stability of the different conformations, the fact that completely independent X-ray studies support the existence of these two conformations lends validity to our conclusions. These can be summarized as follows. Conformation A represents a partially specific DNA bound configuration with a single base contact by Arg 5 in the minor groove. Conformation B represents the specific Pdx1 conformation, forming additional direct and water-mediated contacts with DNA bases by Asn 51 and Gln 50 in the major groove, and by Lys 2 in the minor groove. These conformations differ in the position of helix 3 in the major groove and indicate some of the inherent flexibility of homeodomains in binding DNA. The stability of both conformations suggests they both play a role in the free energy landscape of the complex: either as stable minima or a kinetically trapped intermediate (Conformation A) in search of a global minimum (Conformation B). Flexibility in DNA binding of the homeodomain may be important in allowing Pdx1 to fulfill its multiple functional roles, particularly in binding non-consensus DNA sequences or in the presence of DNA binding partners. A source of diversity of homeodomain function may derive from distinct bound states with differing degrees of DNA binding specificity. Further structural and MD studies of Pdx1 to different DNA sequences and in the presence of partner proteins are necessary to characterize DNA binding in the context of authentic enhancers.

Supporting Information

Figure S1 Packing of the Pdx1/DNA complex in the unit cell of the crystal structure. Each asymmetric unit contains two Pdx1 monomers in Conformation A (yellow) and Conformation B (magenta), and two DNA helices (colored black, red, green and

blue in asymmetric unit 1, 2, 3 and 4, respectively). The packing constraints differ for each model during the crystal simulation. B1 and A4 are the most constrained by crystal contacts, including the N-terminal arm; in A1 and B4 the helices are constrained but not the N-terminal arm; and A2, A3, B2 and B3 are not constrained by crystal contacts.

(TIF)

Figure S2 Different contacts between Conformations A and B in the crystal simulation. A) In Conformation A the hydrogen bond between Asn 51 and Ade 3 is lost in all models. B) In Conformation B the contact between Asn 51 and Ade 3 is more consistent than Conformation A, but still lost in all but model B2. C) In Conformation A when Asn 51 is not contacting the base of Ade 3 it frequently forms a hydrogen bond with the phosphate backbone of Ade 2. This contact is favored in the solution simulation of Conformation A, and is not formed in Conformation B (see Figure 3). D) In Conformation A Lys 46 contacts the phosphate backbone of Ade 8*. This is one of the backbone-specific contacts in Conformation A. E) In Conformation B the side chain of Lys 46 is more mobile than in Conformation A.

(TIF)

Figure S3 Contacts stabilizing the N-terminal arm in Conformation B during the crystal simulation. The N-terminal arm is stabilized by contacts by Arg 43 in the major groove and Arg 3 in the minor groove. A) In Conformation B, Arg 43 contacts the phosphate backbone of Ade 3. This contact is broken in model B4 (blue) after the N-terminal arm escapes the minor groove. B) Arg 3 generally contacts the phosphate backbone of Thy 4, and C) Arg 43 contacts His 44 in Conformation B, but D) not in Conformation A. E) In Conformation A Arg 43 is generally mobile except in models A2 (red) and A3 (green) in which Arg 43 contacts the phosphate backbone of Thy 4. In model A2 (red) this contact correlates with insertion of Arg 3 into the minor groove, before 30 ns (Figure 2B).

(TIF)

Figure S4 Contacts stabilizing the N-terminal arm in Conformation B during the solution simulation. A) Lys 2 remains in the minor groove in Conformation B (red) for about 35 ns, contacting the base of Thy 2*. The N-terminal arm is disordered in conformation A (black). B) Arg 43 contacts the phosphate backbone of Thy 4 in Conformation B (red). C) Arg 3 contacts the phosphate backbone of Thy 4 in Conformation B for about 25% of the trajectory. D) Arg 43 and Arg 3 may interact through pi-pi stacking in Conformation B only. E) In Conformation A, Arg 43 contacts the phosphate backbone of Ade 3 during about 1/3 of the solution simulation (black). F) In Conformation A Glu 42 interacts with Arg 31 (black) and stabilizes the phosphate contact between Arg 31 and Ade 8*, the only specific phosphate contact remaining in Conformation A. G) A hydrogen bond between Tyr 25 OH and Thy 6* O1P is unique to Conformation B (red).

(TIF)

Author Contributions

Conceived and designed the experiments: VB RBR CS. Performed the experiments: VB. Analyzed the data: VB DW. Wrote the paper: VB RBR CS.

References

1. Sarai A, Kono H (2005) Protein-DNA Recognition Patterns and Predictions. *Annu Rev of Bioph and Biom* 34: 379–398.
2. Spolar RS, Record MT, Jr. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263: 777–784.
3. Dessain S, Gross CT, Kuziora MA, McGinnis W (1992) Antp-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. *EMBO J* 11: 991–1002.

4. Holland PW, Booth HA, Bruford EA (2007) Classification and nomenclature of all human homeobox genes. *BMC Biol* 5: 47.
5. Mann RS, Lelli KM, Joshi R (2009) Hox specificity unique roles for cofactors and collaborators. *Curr Top Dev Biol* 88: 63–101.
6. Clarke ND, Kissinger CR, Desjarlais J, Gilliland GL, Pabo CO (1994) Structural studies of the engrailed homeodomain. *Protein Sci* 3: 1779–1787.
7. Fraenkel E, Rould MA, Chambers 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.
8. Qian YQ, Otting G, Billeter M, Muller M, Gehring W, et al. (1993) Nuclear magnetic resonance spectroscopy of a DNA complex with the uniformly ¹³C-labeled Antennapedia homeodomain and structure determination of the DNA-bound homeodomain. *J Mol Biol* 234: 1070–1083.
9. Farber PJ, Mittermaier A (2011) Concerted dynamics link allosteric sites in the PBX homeodomain. *J Mol Biol* 405: 819–830.
10. Garcia-Fernandez J (2005) The genesis and evolution of homeobox gene clusters. *Nat Rev Genet* 6: 881–892.
11. Hueber SD, Lohmann I (2008) Shaping segments: Hox gene function in the genomic age. *Bioessays* 30: 965–979.
12. Donaldson IJ, Amin S, Hensman JJ, Kutejova E, Rattray M, et al. (2012) Genome-wide occupancy links Hoxa2 to Wnt-beta-catenin signaling in mouse embryonic development. *Nucleic Acids Res* 40: 3990–4001.
13. Khoo C, Yang J, Weinrott SA, Kaestner KH, Naji A, et al. (2012) Research resource: the pdx1 cisrome of pancreatic islets. *Mol Endocrinol* 26: 521–533.
14. Boyer DF, Fujitani Y, Gannon M, Powers AC, Stein RW, et al. (2006) Complementation rescue of Pdx1 null phenotype demonstrates distinct roles of proximal and distal cis-regulatory sequences in pancreatic and duodenal expression. *Dev Biol* 298: 616–631.
15. Guz Y, Montminy MR, Stein R, Leonard J, Gamer LW, et al. (1995) Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny. *Development* 121: 11–18.
16. McKinnon CM, Docherty K (2001) Pancreatic duodenal homeobox-1, PDX-1, a major regulator of beta cell identity and function. *Diabetologia* 44: 1203–1214.
17. Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, et al. (1996) PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* 122: 983–995.
18. Stoffers DA, Ferrer J, Clarke WL, Habener JF (1997) Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. *Nat Genet* 17: 138–139.
19. Weng J, Macfarlane WM, Lehto M, Gu HF, Shepherd LM, et al. (2001) Functional consequences of mutations in the MODY4 gene (IPF1) and coexistence with MODY3 mutations. *Diabetologia* 44: 249–258.
20. Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, et al. (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* 133: 1266–1276.
21. Carr A, Biggin MD (1999) A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in *Drosophila* embryos. *EMBO J* 18: 1598–1608.
22. Joshi R, Sun L, Mann R (2010) Dissecting the functional specificities of two Hox proteins. *Genes Dev* 24: 1533–1545.
23. Liberzon A, Ridner G, Walker MD (2004) Role of intrinsic DNA binding specificity in defining target genes of the mammalian transcription factor PDX1. *Nucleic Acids Res* 32: 54–64.
24. Puppini C, Fabbro D, Pellizzari L, Damante G (2011) Using the recognition code to swap homeodomain target specificity in cell culture. *Mol Biol Rep* 38: 5349–5354.
25. Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, et al. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 131: 530–543.
26. Mann RS, Chan SK (1996) Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet* 12: 258–262.
27. Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, et al. (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* 147: 1270–1282.
28. Wilson DS, Desplan C (1999) Structural basis of Hox specificity. *Nat Struct Biol* 6: 297–300.
29. Goudet G, Delhalle S, Biemar F, Martial JA, Peers B (1999) Functional and cooperative interactions between the homeodomain PDX1, Pbx, and Prepl factors on the somatostatin promoter. *J Biol Chem* 274: 4067–4073.
30. Peers B, Sharma S, Johnson T, Kamps M, Montminy M (1995) The pancreatic islet factor STF-1 binds cooperatively with Pbx to a regulatory element in the somatostatin promoter: importance of the FPWMK motif and of the homeodomain. *Mol Cell Biol* 15: 7091–7097.
31. Swift GH, Liu Y, Rose SD, Bischof IJ, Steelman S, et al. (1998) An endocrine-exocrine switch in the activity of the pancreatic homeodomain protein PDX1 through formation of a trimeric complex with PBX1b and MRG1 (MEIS2). *Mol Cell Biol* 18: 5109–5120.
32. Glick E, Leshkowitz D, Walker MD (2000) Transcription factor BETA2 acts cooperatively with E2A and PDX1 to activate the insulin gene promoter. *J Biol Chem* 275: 2199–2204.
33. Ohmeda K, Mirmira RG, Wang J, Johnson JD, German MS (2000) The homeodomain of PDX-1 mediates multiple protein-protein interactions in the formation of a transcriptional activation complex on the insulin promoter. *Mol Cell Biol* 20: 900–911.
34. Peers B, Leonard J, Sharma S, Teitelman G, Montminy MR (1994) Insulin expression in pancreatic islet cells relies on cooperative interactions between the helix loop helix factor E47 and the homeobox factor STF-1. *Mol Endocrinol* 8: 1798–1806.
35. Liu Y, Matthews KS, Bondos SE (2008) Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the *Drosophila* hox protein ultrabithorax. *J Biol Chem* 283: 20874–20887.
36. Kishi A, Nakamura T, Nishio Y, Maegawa H, Kashiwagi A (2003) Sumoylation of Pdx1 is associated with its nuclear localization and insulin gene activation. *Am J Physiol Endocrinol Metab* 284: E830–840.
37. Macfarlane WM, McKinnon CM, Felton-Edkins ZA, Cragg H, James RF, et al. (1999) Glucose stimulates translocation of the homeodomain transcription factor PDX1 from the cytoplasm to the nucleus in pancreatic beta-cells. *J Biol Chem* 274: 1011–1016.
38. Rafiq I, da Silva Xavier G, Hooper S, Rutter GA (2000) Glucose-stimulated preproinsulin gene expression and nuclear translocation of pancreatic duodenal homeobox-1 require activation of phosphatidylinositol 3-kinase but not p38 MAPK/SAPK2. *J Biol Chem* 275: 15977–15984.
39. Li X, McGinnis W (1999) Activity regulation of Hox proteins, a mechanism for altering functional specificity in development and evolution. *Proc Natl Acad Sci U S A* 96: 6802–6807.
40. Mann RS (2005) The specificity of homeotic gene function. *BioEssays* 17: 855–863.
41. Baird-Titus JM, Clark-Baldwin K, Dave V, Caperelli CA, Ma J, et al. (2006) The solution structure of the native K50 Bicoid homeodomain bound to the consensus TAAATCC DNA-binding site. *J Mol Biol* 356: 1137–1151.
42. Grant RA, Rould MA, Klemm JD, Pabo CO (2000) Exploring the role of glutamine 50 in the homeodomain-DNA interface: crystal structure of engrailed (Gln50→ala) complex at 2.0 Å. *Biochemistry* 39: 8187–8192.
43. Gehring WJ, Affolter M, Burglin T (1994) Homeodomain proteins. *Annu Rev Biochem* 63: 487–526.
44. Toth-Petroczy A, Simon I, Fuxreiter M, Levy Y (2009) Disordered tails of homeodomains facilitate DNA recognition by providing a trade-off between folding and specific binding. *J Am Chem Soc* 131: 15084–15085.
45. Damante G, Di Lauro R (1991) Several regions of Antennapedia and thyroid transcription factor 1 homeodomains contribute to DNA binding specificity. *Proc Natl Acad Sci U S A* 88: 5388–5392.
46. Lin L, McGinnis W (1992) Mapping functional specificity in the Dfd and Ubx homeo domains. *Genes Dev* 6: 1071–1081.
47. Zeng W, Andrew DJ, Mathies LD, Horner MA, Scott MP (1993) Ectopic expression and function of the Antp and Scr homeotic genes: the N terminus of the homeodomain is critical to functional specificity. *Development* 118: 339–352.
48. Givaty O, Levy Y (2009) Protein sliding along DNA: dynamics and structural characterization. *J Mol Biol* 385: 1087–1097.
49. Vuzman D, Azia A, Levy Y (2010) Searching DNA via a “Monkey Bar” mechanism: the significance of disordered tails. *J Mol Biol* 396: 674–684.
50. Vuzman D, Levy Y (2010) DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc Natl Acad Sci U S A* 107: 21004–21009.
51. Longo A, Guanga GP, Rose RB (2007) Structural basis for induced fit mechanisms in DNA recognition by the Pdx1 homeodomain. *Biochemistry* 46: 2948–2957.
52. Babin V, Baucom J, Darden TA, Sagui C (2006) Molecular dynamics simulations of polarizable DNA in crystal environment. *Int J Quantum Chem* 106: 3260–3269.
53. Babin V, Baucom J, Darden TA, Sagui C (2006) Molecular dynamics simulations of DNA with polarizable force fields: Convergence of an ideal B-DNA structure to the crystallographic structure. *J Phys Chem B* 110: 11571–11581.
54. Baucom J, Transue T, Fuentes-Cabrera M, Krahn JM, Darden TA, et al. (2004) Molecular dynamics simulations of the d(CCAACGTTGG)(2) decamer in crystal environment: comparison of atomic point-charge, extra-point, and polarizable force fields. *J Chem Phys* 121: 6998–7008.
55. Bevan DR, Li L, Pedersen LG, Darden TA (2000) Molecular dynamics simulations of the d(CCAACGTTGG)(2) decamer: influence of the crystal environment. *Biophys J* 78: 668–682.
56. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242.
57. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, et al. (2008) AMBER, version 10. San Francisco (California): University of California.
58. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, et al. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* 65: 712–725.
59. Perez A, Marchan I, Svobol D, Sponer J, Cheatham TE, et al. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys J* 92: 3817–3829.
60. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935.

61. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, et al. (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J Am Chem Soc* 117: 5179–5197.
62. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J Chem Phys* 98: 10089–10092.
63. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, et al. (1995) A Smooth Particle Mesh Ewald Method. *J Chem Phys* 103: 8577–8593.
64. Brunger A, Brooks CL, Karplus M (1984) Stochastic Boundary-Conditions for Molecular-Dynamics Simulations of St2 Water. *Chem Phys Lett* 105: 495–500.
65. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys* 81: 3684–3690.
66. Lu XJ, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31: 5108–5121.
67. Gutmanas A, Billeter M (2004) Specific DNA recognition by the Antp homeodomain: MD simulations of specific and nonspecific complexes. *Proteins* 57: 772–782.
68. DeLano WL (2002) The PyMOL Molecular Graphics System, version 0.99. South San Francisco (California): Schrodinger, LLC.
69. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612.
70. Billeter M, Qian YQ, Otting G, Muller M, Gehring W, et al. (1993) Determination of the Nuclear Magnetic Resonance Solution Structure of an Antennapedia Homeodomain-DNA Complex. *J Mol Biol* 234: 1084–1097.
71. Fraenkel E, Pabo CO (1998) Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex. *Nat Struct Biol* 5: 692–697.
72. Hirsch JA, Aggarwal AK (1995) Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. *EMBO J* 14: 6280–6291.
73. Roy S, Thakur AR (2010) 20 ns molecular dynamics simulation of the antennapedia homeodomain-DNA complex: water interaction and DNA structure analysis. *J Biomol Struct Dyn* 27: 443–456.
74. Iurcu-Mustata G, Van Belle D, Wintjens R, Prevost M, Rooman M (2001) Role of salt bridges in homeodomains investigated by structural analyses and molecular dynamics simulations. *Biopolymers* 59: 145–159.
75. Duan J, Nilsson L (2002) The role of residue 50 and hydration water molecules in homeodomain DNA recognition. *Eur Biophys J* 31: 306–316.
76. Zhao X, Huang XR, Sun CC (2006) Molecular dynamics analysis of the engrailed homeodomain-DNA recognition. *J Struct Biol* 155: 426–437.
77. Rundgren H, Mark P, Laaksonen A (2007) Molecular dynamics simulations of conserved Hox protein hexapeptides. I. Folding behavior in water solution. *J Mol Struct: Theochem* 810: 113–120.
78. Del Vecchio P, Carullo P, Barone G, Pagano B, Graziano G, et al. (2008) Conformational stability and DNA binding energetics of the rat thyroid transcription factor 1 homeodomain. *Proteins* 70: 748–760.
79. Flader W, Wellenzohn B, Winger RH, Hallbrucker A, Mayer E, et al. (2003) Stepwise induced fit in the pico- to nanosecond time scale governs the complexation of the even-skipped transcriptional repressor homeodomain to DNA. *Biopolymers* 68: 139–149.
80. 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.
81. Yang SY, Yang XL, Yao LF, Wang HB, Sun CK (2011) Effect of CpG methylation on DNA binding protein: molecular dynamics simulations of the homeodomain PITX2 bound to the methylated DNA. *J Mol Graph Model* 29: 920–927.
82. Cheatham TE (2004) Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr Opin Struc Biol* 14: 360–367.
83. Cheatham TE, Kollman PA (2000) Molecular dynamics simulation of nucleic acids. *Annu Rev Phys Chem* 51: 435–471.
84. Dixit SB, Beveridge DL, Case DA, Cheatham TE, Giudice E, et al. (2005) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys J* 89: 3721–3740.
85. Herce DH, Darden T, Sagui C (2003) Calculation of ionic charging free energies in simulation systems with atomic charges, dipoles, and quadrupoles. *J Chem Phys* 119: 1609191.
86. Wagoner JA, Pande VS (2011) A smoothly decoupled particle interface: new methods for coupling explicit and implicit solvent. *J Chem Phys* 134: 214103.
87. Karttunen J, Rottler J, Vattulainen I, Sagui C (2008) Electrostatics in biomolecular simulations: Where are we now and where are we heading? In: Feller SE, editor. *Current topics in membranes*. Burlington: Academic Press. pp. 49–89.
88. Sagui C, Darden TA (1999) Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu Rev Biophys Biomol Struct* 28: 155–179.
89. Lee H, Darden T, Pedersen L (1995) Accurate crystal molecular dynamics simulations using particle-mesh-Ewald: RNA dinucleotides -ApU and GpC. *Chem Phys Lett* 243: 229–235.
90. Lee H, Darden TA, Pedersen LG (1995) Molecular dynamics simulation studies of a high resolution Z-DNA crystal. *J Chem Phys* 102: 3830–3834.
91. York DM, Wlodawer A, Pedersen LG, Darden TA (1994) Atomic-level accuracy in simulations of large protein crystals. *Proc Natl Acad Sci U S A* 91: 8715–8718.
92. York DM, Yang W, Lee H, Darden T, Pedersen L (1995) Toward the accurate modeling of DNA: the importance of long-range electrostatics. *J Am Chem Soc* 117: 5001–5002.
93. Aishima J, Wolberger C (2003) Insights into nonspecific binding of homeodomains from a structure of MATalpha2 bound to DNA. *Proteins* 51: 544–551.
94. LaRonde-LeBlanc NA, Wolberger C (2003) Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes Dev* 17: 2060–2072.
95. Piper DE, Batchelor AH, Chang CP, Cleary ML, Wolberger C (1999) Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* 96: 587–597.
96. Affolter M, Percival-Smith A, Muller M, Leupin W, Gehring WJ (1990) DNA binding properties of the purified Antennapedia homeodomain. *Proc Natl Acad Sci U S A* 87: 4093–4097.
97. Vershon AK, Jin Y, Johnson AD (1995) A homeo domain protein lacking specific side chains of helix 3 can still bind DNA and direct transcriptional repression. *Genes Dev* 9: 182–192.
98. Le Lay J, Matsuoka TA, Henderson E, Stein R (2004) Identification of a novel PDX-1 binding site in the human insulin gene enhancer. *J Biol Chem* 279: 22228–22235.
99. Damante G, Pellizzari L, Esposito G, Fogolari F, Viglino P, et al. (1996) A molecular code dictates sequence-specific DNA recognition by homeodomains. *EMBO J* 15: 4992–5000.
100. Dragan AI, Li Z, Makeyeva EN, Milgotina EI, Liu Y, et al. (2006) Forces driving the binding of homeodomains to DNA. *Biochemistry* 45: 141–151.
101. Ekker SC, Jackson DG, von Kessler DP, Sun BI, Young KE, et al. (1994) The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. *EMBO J* 13: 3551–3560.
102. Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, et al. (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* 133: 1277–1289.
103. Morgan R, In der Rieden P, Hooiveld MH, Durston AJ (2000) Identifying HOX paralog groups by the PBX-binding region. *Trends Genet* 16: 66–67.