Structural Basis for Induced Fit Mechanisms in DNA Recognition by the Pdx1 Homeodomain[†,‡]

Antonella Longo,[§,||] Gerald P. Guanga,[§] and Robert B. Rose*[,§]

Department of Molecular and Structural Biochemistry, 128 Polk Hall, North Carolina State University, Raleigh, North Carolina 27695, and CCS-CSM, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

Received May 16, 2006; Revised Manuscript Received December 20, 2006

ABSTRACT: Pancreatic and duodenal homeobox 1 (Pdx1) is a homeodomain transcription factor belonging to the ParaHox family. Pdx1 plays an essential role in pancreatic endocrine and exocrine cell development and maintenance of adult islet β-cell function. Mutations in the human *pdx1* gene are linked to an early onset form of non-insulin-dependent diabetes mellitus, MODY-4. We demonstrate that the homeodomain reproduces the binding specificity of the full-length protein. We report the 2.4 Å resolution crystal structure of the homeodomain bound to a target DNA. The two Pdx1/DNA complexes in the asymmetric unit display conformational differences: in the DNA curvature, the orientation of the homeodomain in the major groove, and the order of the N-terminal arm. Comparing the two complexes indicates invariant protein−DNA contacts, and variant contacts that are unique to each binding orientation. An induced fit model is proposed that depends on the DNA conformation and provides a mechanism for nonlocal contributions to binding specificity.

Pancreatic and duodenal homeobox-1 (Pdx1[1]) [also known as insulin promoter factor-1 (Ipf1), islet/duodenum homeobox-1 (Idx1), and somatostatin transactivating factor-1 (Stf1)] is a homeodomain (HD) protein essential for pancreatic development, differentiation of pancreatic β-cells, and proper function of the mature endocrine pancreas (reviewed in refs *1*, *2*). Expression in the adult is primarily restricted to pancreatic β-cells, with some expression in pancreatic δ- and acinar cells, and neural cells (*1*, *3*−*6*). Pdx1, along with Gsx and Cdx, is a member of the ParaHox gene cluster, which diverged from the Homeobox (Hox) cluster through an early duplication event (*7*, *8*). The sequence of Pdx1 is highly conserved in zebrafish, *Xenopus*, and mammals (*2*). Among the Hox factors, the Pdx1 sequence resembles Hox3 most closely (*9*) (Figure 1a). All ParaHox genes are expressed in the endocrine pancreas, but only Pdx1 is specifically expressed in β-cells (*10*).

The consensus binding site for Pdx1 was defined as 5′-TAAT(T/G)-3′ (*11*). Pdx1 is one among several HD factors expressed in the pancreas, including the Parahox proteins Gsx and Cdx, the paired-box proteins Pax4 and Pax6, the NK2 proteins Nkx2.2 and Nkx6.1, Brn4, the LIM protein Isl1, and the POU proteins Hnf-1α and Hnf-1β (*12*, *13*). While the consensus binding site varies among these HD factors, many of them are capable of binding sequences which include a TAAT core (*14*−*16*). The unique role of Pdx1 in regulating β-cell development and homeostasis suggests the importance of selective DNA binding. Hox proteins depend on the interaction with other DNA-binding proteins that act as Hox cofactors to achieve DNA-binding specificity (*17*). Similarly Pdx1 interacts with different cofactors depending on the cellular context. In pancreatic δ-cells Pdx1 interacts with Pbx1 and Prep1, from the PBC and MEIS classes of the three amino acid loop extension (TALE) family of HDs, to regulate the somatostatin promoter (*18*, *19*). In neural cells, Pdx1 utilizes a different binding site on the somatostatin promoter than in δ-cells, suggesting a distinct protein complex (*6*). In pancreatic acinar cells, Pdx1 interacts with Pbx1b and Mrg1 to regulate the elastase I gene (*20*). Interaction with the TALE proteins is mediated through a conserved pentapeptide motif, FPWMK, located 10 residues N-terminal of the HD (*21*−*24*). Pdx1 is also known to interact with cofactors from other protein families. In pancreatic β-cells, where the TALE factors are not present, Pdx1 regulates the insulin gene that does not contain a Pbx binding site (*19*). Instead Pdx1 interacts with the basic helix−

[†] This work was supported by a Junior Faculty Award from the American Diabetes Association (7-03-JF-34 to R.B.R) and a Joint Faculty Agreement between UT-Battelle/Oak Ridge National Labs and North Carolina State University (A.L.). X-ray data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID (or 22-BM) beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract W-31-109-Eng-38.

[‡] Coordinates and structure factors for the Pdx1-HD/DNA complex were deposited in the Worldwide Protein Data Bank under entry code 2H1K.

* To whom correspondence should be addressed: Department of Molecular & Structural Biochemistry, North Carolina State University, 128 Polk Hall, Raleigh, NC 27695-7622. Phone: (919) 513-4191. Fax: (919) 515-2047. E-mail: bob_rose@ncsu.edu.

[§] North Carolina State University.

[||] Oak Ridge National Laboratory.

[1] Abbreviations: Antp, Antennapedia; bHLH, basic helix−loop−helix; bp, base-pair; EMSA, electrophoretic mobility shift assay; HD, homeodomain; Hox, homeobox; Idx1, islet/duodenum homeobox-1; Ipf1, insulin promoter factor-1; IPTG, isopropyl-β-D-thiogalactopyranoside; MODY-4, maturity-onset diabetes of the young type-4; Pdx1, pancreatic and duodenal homeobox-1; Pbx1, pre-B cell homeobox 1; rmsd, root-mean-square deviation; Stf1, somatostatin transactivating factor-1; TALE, three amino acid loop extension.

10.1021/bi060969l CCC: $37.00
Published on Web 02/23/2007

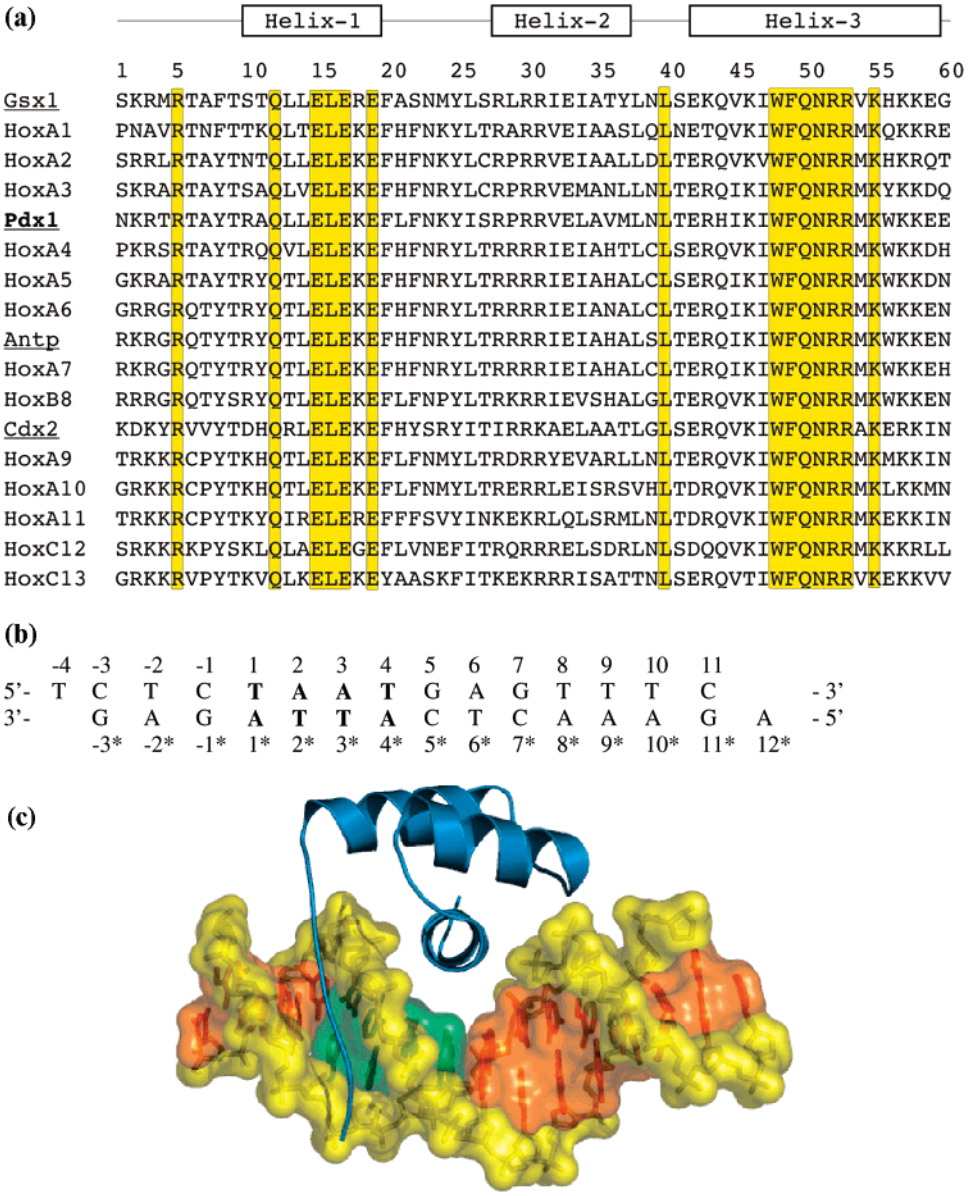


FIGURE 1: (a) Multiple sequence alignment of human Hox and ParaHox (underlined) proteins, and for *Drosophila* Antp. The sequence for the Syrian hamster Pdx1 is 100% identical to the human orthologue in the HD region. The amino acids that are strictly conserved are highlighted in yellow. The secondary structure for Pdx1 is reported above the sequences. Numbering follows the convention for the 60 residues of the HD. Sequences were aligned with the program Clustal W (*69*). (b) Sequence of the 15-bp DNA fragment used to crystallize the complex. The bases in bold (**TAAT**) are the core binding site for Pdx1. The DNA sequence was numbered 5′ to 3′ according to the TAAT strand ($N_{-3}N_{-2}N_{-1}T_1A_2A_3T_4N_5N_6N_7$, as in ref *16*). The complementary strand is labeled with ∗. (c) Ribbon drawing of the Pdx1-HD/DNA complex structure. The complex is viewed along the axis of the recognition helix. The protein is in blue; the sugar−phosphate backbone is in yellow, the bases are in orange except for the TAAT sequence in green. Model B (segids B, E, and F) is represented.

loop−helix (bHLH) factor E47/NeuroD on the rat I insulin promoter (*25*−*27*). The protein−protein interacting surfaces are not known for the Pdx1/E47/NeuroD complex, but the HD of Pdx1 is sufficient to interact with E47/NeuroD1 (*27*).

Mutations in Pdx1 have been associated with a familial form of type II diabetes, maturity-onset diabetes of the young-type 4 (MODY-4) (*28*−*30*). MODY is characterized by early disease onset (usually <25 years), autosomal dominant inheritance, and a primary defect in insulin secretion (*31*). Eight of the Pdx1 mutations are missense mutations. One of these mutations, R197H, is located within the HD region (*32*).

As a step toward characterizing DNA binding specificity by Pdx1, we determined the three-dimensional structure of the Pdx1 HD bound to its cognate DNA. We show that the HD itself is sufficient to dictate the DNA binding specificity observed for the full-length protein. The crystal structure contains two nonequivalent protein/DNA complexes, each demonstrating a number of unique DNA contacts. We discuss the possibility of distinct induced conformations of the DNA/Pdx1 complex for modulating DNA binding specificity on different promoters.

EXPERIMENTAL PROCEDURES

Protein Cloning, Expression, and Purification. The gene for the Syrian hamster (*Mesocricetus auratus*) *pdx1* was a gift of Prof. German (pBAT14.shPdx1) (*27*). Residues 146−206 (1−60 of the HD plus 1 extra-residue, an Asp, at the C-terminus) were extracted by PCR and cloned into a

pET24b vector (Promega) modified in our laboratory to encode an N-terminal His-tag followed by a thrombin cleavage site (referred to as pET24b-6H). The resulting plasmid, pET24c-6H-Pdx1$_{(146-206)}$, was transformed into the *Escherichia coli* strain BL21(*DE3*). The protein was overexpressed after induction with 100 mM IPTG. Cells were lysed and the Pdx1 HD was purified from the supernatant by nickel-resin affinity chromatography according to the manufacturer's instructions (Qiagen). The six-histidine tag was removed by incubating the purified protein with thrombin at room temperature overnight. After thrombin cleavage, two extra amino acids, Gly and Ser, remain at the N-terminus of the protein. After a second purification step through a heparin column (Amersham-Biosciences), the protein was dialyzed into a buffer containing 100 mM Tris/HCl pH 7.6, 150 mM NaCl, and 10% glycerol, and concentrated with a centrifuge filter device (Amicon from Millipore). High purity salt free (HPSF) oligonucleotides for crystallization were purchased from MWG Biotech. Complementary oligonucleotides were mixed, heated to 100 °C, and then annealed by slow cooling in a water bath. The protein/DNA complex was formed by mixing protein and DNA at a 1:1.2 ratio; the final concentration for the complex was 0.77 mM. The complex was directly used for crystallization without further purification.

Crystallization. Initial crystallization conditions were identified using commercial crystallization screens (Hampton Research, Emerald Biostructures, Nextal Biotechnologies) with the sitting drop vapor diffusion method and optimized. Eight different oligonucleotide sequences of varying lengths and end configuration (blunt or single base overhang) were screened. The best crystals were obtained with the microbatch-under-oil method: 1 μL of the protein/DNA complex was mixed with 1 μL of a buffer containing 100 mM imidazole pH 7.0, 280 mM $MgCl_2$, and 21.4% PEG 4000. The drops were covered with a mixture of paraffin oil (90%) and silicon oil (10%) and placed at 8 °C. Crystals grew overnight and reached maximum size in 2−3 days.

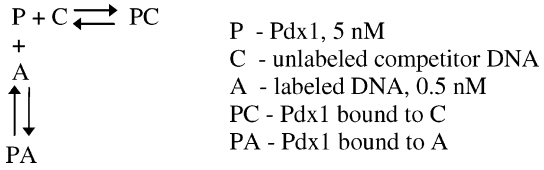
Data Collection and Analysis. Crystals were flash-frozen in liquid nitrogen after brief equilibration with a cryoprotectant solution containing 125 mM imidazole pH 7.0, 350 mM $MgCl_2$, 125 mM NaCl, 26.8% PEG 4000, 7.7% PEG 400, and 6.25% glycerol. Data were collected at the SER-CAT ID-22 beam line at the Advanced Photon Source (APS) in Argonne, IL, with a MAR 300 detector. Data were indexed and scaled using MOSFLM (*33*). The space group is $P2_12_12_1$ with unit cell dimensions of $a = 58.28$ Å, $b = 61.92$ Å, and $c = 96.44$ Å and two complexes in the asymmetric unit.

Structure Determination, Model Building, and Refinement. The Pdx1 structure was solved by molecular replacement using the crystallographic structure for the *Drosophila* Antennapedia (Antp) HD/DNA complex (pdb id: 9ANT) (*34*) as a search model. The HD sequences for Pdx1 and Antp are 73% identical (Figure 1a). All nonidentical residues were mutated to alanines in the search model: the DNA sequence although different was not mutated. CNS was used to search for molecular replacement solutions (*35*). The model was built using O (*36*) and refined with CNS (*35*) to a final *R*-factor of 22.2% and *R*-free of 27.3%.

Figures of protein and DNA models were generated with PyMOL (*37*). The curvature of the DNA was calculated between backbone phosphates using the program Curves (*38*). Structures were superimposed using either PyMOL (*37*) or O (*36*).

DNA-Binding Assays. Electrophoretic mobility shift assays (EMSA) were performed using the purified Pdx1 HD. Double-stranded synthetic oligonucleotides containing the A1 sequence and labeled with an infrared dye (IRDye 800, LiCor) were annealed following the manufacturer's protocol. Unlabeled double-stranded oligonucleotides of the A1, albumin (Alb), GLUT2, and E1 sequences (synthesized by MWG Biotech) were diluted in 2.5 mM HEPES, pH 7.9, 15 mM KCl, 1% glycerol, and 0.5 mM DTT, and annealed as described in ref *39*. For competition experiments, 0.5 nM labeled A1 was incubated with 5 nM of the Pdx1 HD and increasing concentrations (1 nM to 500 nM) of competitor DNA for 30 min on ice in EMSA binding buffer (10 mM HEPES buffer pH 7.1, 75 mM KCl, 2.5 mM $MgCl_2$, 2.5 mM DTT, 0.1 mM EDTA, 10% glycerol, 0.5% Tween-20). Samples were run on a 6% native polyacrylamide gel using a Tris-glycine-EDTA buffer and analyzed with the Odyssey Infrared Imaging System (LI-COR Biosciences).

The apparent dissociation constants of the competitor DNA (K_C) were calculated from the fraction of labeled DNA shifted in the presence of the competitor. The formula was derived from the following scheme:



Because the competitor DNA concentrations were much greater than the concentration of Pdx1, the equation was derived with C constant considering A as a competitor: $PC/P_{tot} = C/[K_C(1 + A/K_A) + C]$, with K_A the dissociation constant for A, and P_{tot} the total concentration of Pdx1. An expression for P was derived using $K_C = (P)(C)/PC$. An expression for PA was then derived using $K_A = (A)(P)/PA$:

$$PA = (1/2)[(K_A + A_{tot} + P_{tot} + K_A C/K_C) - [(K_A + A_{tot} + P_{tot} + K_A C/K_C)^2 - 4P_{tot}A_{tot}]^{1/2}]$$

where A_{tot} is the total amount of labeled DNA.

The fraction of Pdx1 bound to A1 was fit to the data in Microsoft Excel (2002) minimizing the root-mean-square difference using the Solver routine. First K_A was derived with A as a competitor and setting $K_A = K_C$. This value was used subsequently to derive K_C for the albumin, GLUT2, and E1 competitors. The equation was fit to 3−4 independent experiments, and the average and standard deviations were calculated.

RESULTS

Overview of the Structure. The 61 amino acids of the Pdx1 HD (Figure 1a) were crystallized bound to a 15-bp DNA fragment including a single base overhang at the 5′ ends (sequence in Figure 1b). The core sequence (5′-CTC**TAAT**-GAG-3′) is identical to the Pdx1 binding site in the mouse glucokinase promoter, and matches the consensus binding site (*39*). The final model of the Pdx1/DNA complex was built into a 2.4 Å 2FoFc map, containing two complexes in

Table 1: Data and Refinement Statistics

Data Statistics	
wavelength (Å)	1.0
space group	$P2_12_12_1$
unit cell	$a = 58.28$ Å
	$b = 61.92$ Å
	$c = 96.44$ Å
resolution	2.42 Å
highest bin	2.57−2.42
reflections	93557 (13762)[a]
completeness (%)	99.6 (97.6)[a]
R_{merge} (%)	7.9 (35.1)[a]
$I/\sigma I$	(2.1)[a]
redundancy	5.3
Refinement Statistics	
R_{factor} (%)	22.7
R_{free} (%)	27.8
no. of protein atoms	1008
no. of solvent atoms	23
rmsd from ideal values	
bond length (Å)	0.011
bond angle (deg)	1.33
average *B* factor, all atoms (Å^2)	61.04

[a] Numbers in parentheses refer to the highest resolution bin.

the asymmetric unit (Table 1). The two complexes were refined separately. One complex will be referred to as model A consisting of Pdx1, segid A, bound to DNA, segids C and D; the other complex will be referred to as model B consisting of Pdx1, segid B, bound to DNA, segids E and F.

The entire DNA duplex is ordered in both models. However the root-mean-square deviation (rmsd) between the two models is large (1.2 Å between the sugar C1′ atoms). The protein backbone traces of model A and model B are identical, with an rmsd of 0.4 Å (residues 10−55). Different residues are disordered in the two Pdx1 models. Amino acids 1 to 3 are disordered in model A, while electron-density is visible for the complete sequence of the N-terminal arm of model B except for the side chain of Asn 1 (see 2Fo-Fc density map in Figure 5a). At the C-termini, the side chains of the last two residues are disordered in model A, and the last four residues are disordered in model B. The remaining amino acids are visible, except for side chains of Arg 10 and Arg 43 in model A and Glu 42 in model B. The final structure contains 23 water molecules: 13 associated with model A (numbered S1−13), and 10 associated with model B (numbered S21−30).

The overall conformation of Pdx1 is similar to that of other HD structures (Figure 1c). The protein folds into three α-helices, with helices 1 and 2 antiparallel to each other and perpendicular to helix 3, and a flexible N-terminal arm. Helix 3, also termed the recognition helix, binds in the major groove of the DNA, while the N-terminal arm contacts the DNA bases through the minor groove. Although the overall structure is similar in the two complexes, there are differences in the conformation of the DNA, the orientation of Pdx1 relative to the DNA, and the conformation of protein side chains. As a consequence, the Pdx1 monomers from the two complexes form many similar DNA contacts (Figure 4a), but also distinct ones (Figure 4b).

Crystal Packing and DNA Distortion. The large rmsd measured between the DNA from the two models can be attributed to a different DNA curvature. In model A the DNA is bent around helix 3 of the HD by about 16° (Figure 3a); in model B the curvature is similar to that of model A near the TAAT core, but the DNA axis curves sharply from bases 6−11 so that the overall curvature is 34°. The increased curvature of the DNA in model B could not be explained by contacts with Pdx1, since the DNA curves away from the protein. In fact helix 3 of Pdx1 makes more contacts with the less bent DNA in model A, to the DNA phosphate backbone of bases Cyt 5*, Cyt 7*, and Ade 8*.

Instead the increased DNA bending in model B is a consequence of crystal packing. In the crystal the DNA forms pseudo-continuous helices, base-pairing through the single base overhangs at both ends (Figure 2a). Adjacent DNA units are related by a crystallographic 2-fold axis, curving in the opposite direction to generate the repeating unit of the crystal. The two HDs in the asymmetric unit are related by a rotation of 154° and interact through an extensive nonsymmetric interface (Figure 2a and b). Therefore the distortion in the DNA bound to Pdx1 model B results from satisfying two packing interfaces: between Pdx1 monomers which orients the bound DNA at less than 180°, and the DNA base-pairing which requires the two DNA double helices in model A and model B to traverse the same overall axial rise. As a comparison crystal packing in the structure of the Antp HD (*34*) is more typical, with the proteins interacting through identical surfaces (burying 910 Å^2) related by a 180° rotation and no DNA distortion.

Although the two Pdx1 monomers interact through a relatively large surface area, there is no evidence from gel shift assays that a Pdx1 dimer exists in solution. Instead, the hydrophobic surface may be involved in interactions with other proteins, such as E47/NeuroD1 (*27*).

The Recognition Helix and the Major Groove. Helix 3 of Pdx1 forms specific interactions in the major groove with the bases of Ade 2, Ade 3, and Thy 4 of the TAAT core, and the bases of Cyt 5*, Thy 6*, and Cyt 7* (Figure 4a). Recognition of the TAAT core in the two complexes is through van der Waals contacts made by Ile 47 with Ade 3 and Thy 4, and through two hydrogen bonds by Asn 51 with Ade 3. Asn 51 also forms a hydrogen bond with Ade 2. Bases Cyt 5*, Thy 6*, and Cyt 7* are recognized through van der Waals contacts with Gln 50 and Met 54. Gln 50 is conserved among all Hox-like HDs (*40*), and Asn 51 is conserved among all homeodomains.

Although the rmsd between the two Pdx1 HDs is very low, the proteins are oriented slightly differently relative to the DNA, differing by a 2.4° rigid body rotation (Figure 3b). This is a small rotation, but it is sufficient to cause a displacement of the backbone atoms between 0.4 and 0.7 Å altering the strength of hydrogen bonds and influencing DNA contacts (Figure 4b). The model A-specific contacts include 7 additional hydrogen bonds mostly located at the C-terminal end of helix 3, while the model B-specific contacts are primarily focused around Arg 43, Asn 51 and water S29 toward the N-terminus of helix 3, and by the N-terminal arm (Figure 4b). The Pdx1-specific residue, His 44, is significantly more buried in model B, contributing to van der Waals interactions with the DNA. Superimposing the TAAT core sequences of model A and B demonstrated that the curvature of the DNA in model B prevented formation of model A-specific contacts, including the contacts by Arg 31, Gln 50, Lys 57, and waters S5 and S12.

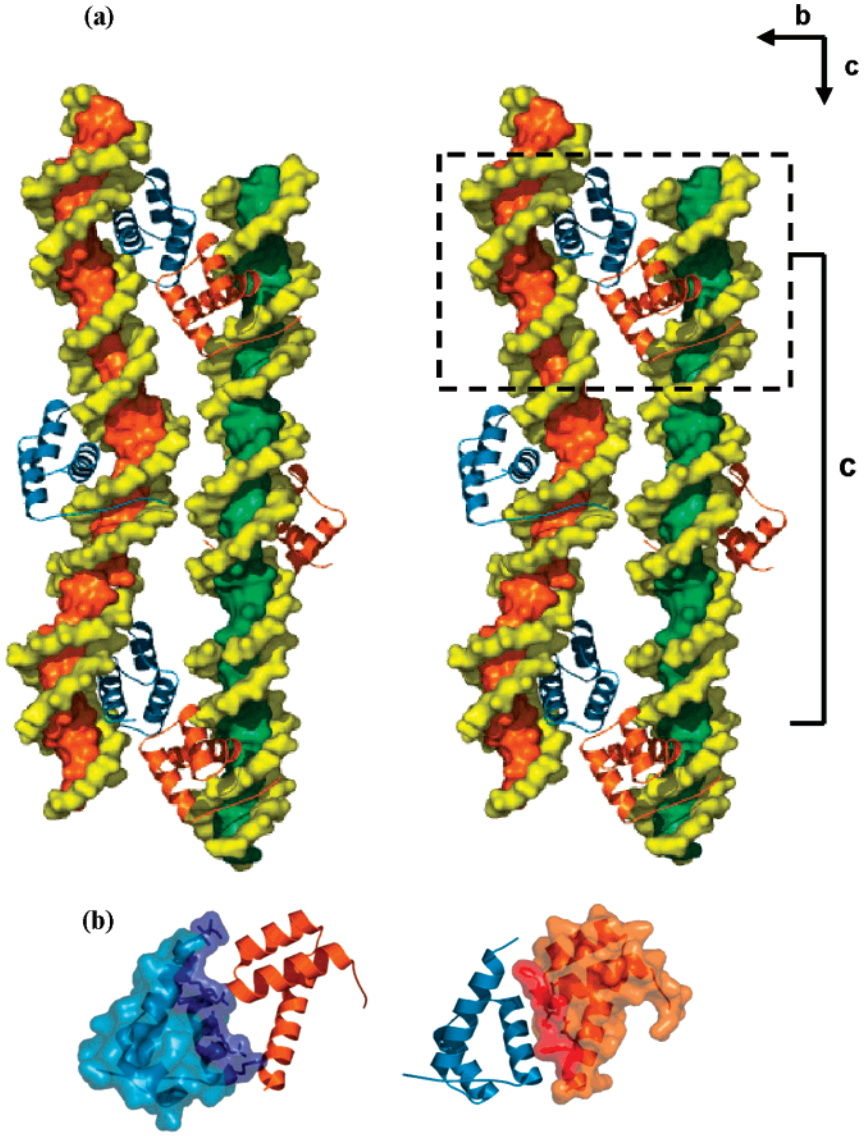


FIGURE 2: (a) Packing forces in the crystal result in increased DNA bending in model B. The figure is oriented looking directly into the crystallographic *a*-axis. The DNA duplexes stack end to end to form pseudo-continuous helices with the average helical axes oriented along the crystallographic *c*-axis. The increased curvature of the DNA associated with model B is apparent in the figure. The two Pdx1-HDs in the asymmetric unit (contained in the square) pack at a 154° angle. The DNA backbone is in yellow; protein and DNA bases from model A are orange and green respectively; protein and DNA bases from model B are blue and orange respectively (cross-eyed stereo). (b) Protein−protein interactions between models A and B are through a large asymmetric surface. The Pdx1 interface buries a total surface area of 1110 Å^2, representing roughly 10% of the 9500 Å^2 combined surface of both Pdx1 HDs, and more than half the 1900 Å^2 buried by the protein/DNA interface of model B. Left: The interface from Pdx1 of model B consists of residues from the C-terminal half of helix 3 and solvent exposed residues from helix 1 (Connolly surface in dark blue). Right: The interface from Pdx1 of model A consists of the same residues from helix 3 with residues from the loop between helix1 and 2 (Connolly surface in red). Model A in orange, model B in blue.

The N-Terminal Arm and the Minor Groove. The N-terminal arm (residues 1−9) of HD proteins primarily contacts the core TAAT bases of the DNA through the minor groove and contributes to binding specificity (*41*). In Pdx1, the N-terminal sequence contains three basic residues, Lys 2, Arg 3, and the strictly conserved Arg 5. Arg 5 forms hydrogen bonds with the bases of Thy 1 and Gua -1*, and van der Waals contact with Ade 2 in both Pdx1 complexes. Residues 1−3 are only ordered in model B. Arg 3 does not contact the DNA bases, but positions the N-terminal arm by reaching into the major groove to interact with the Nϵ nitrogen of Arg 43 on helix 3 (Figure 5b). In turn, the guanidinium nitrogens of Arg 43 form hydrogen bonds with both backbone phosphate oxygens of Ade 3 in the major groove. Arg 43 also forms a hydrogen bond with His 44. In model A neither Arg 43 nor Arg 3 is ordered, and there is a different rotamer for His 44. Lys 2 protrudes into the minor groove to make contact with the base of Thy 2* and forms a weak hydrogen bond (3.5 Å) with Ade 3. We propose that the orientation of the HD in the major groove correlates with ordering of the N-terminal arm and contacts by Arg 43 and Lys 2 with the DNA.

Since the Antp sequence also contains Lys 2, Arg 3, and Arg 43, we compared our structure with the Antp-HD/DNA crystal structure (*34*). In the Antp structure, the side chain of Arg 43 is ordered but it interacts with the phosphate backbone of Thy 4 instead of Ade 3 as in model B of Pdx1 (Figure 5c). The first four residues of the N-terminal arm are disordered in Antp. Overlapping the TAAT sequence of the two structures indicates that the HD orientation relative

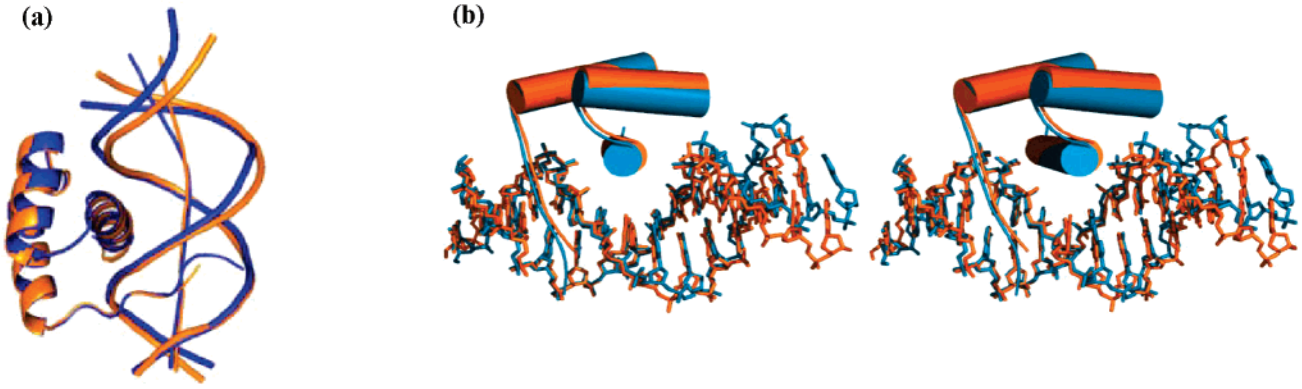


FIGURE 3: (a) Comparison of the DNA curvature in model A and B. The Cα atoms of the HDs of Pdx1 from model A (orange) and B (blue) were superimposed (residues 10−55). The curvature of the DNA in model B (blue, 34°) was significantly greater than the curvature of the DNA in model A (orange, 16°). The axis was derived using Curves (*38*). (b) Comparison of the orientation of Pdx1 in model A and model B relative to the DNA. To compare the orientations of the two HDs, model A was first superimposed onto model B using backbone atoms of Pdx1 (residues 10−55, rmsd 0.4 Å). From this orientation, we identified the transformation that superimposed the DNA of model A onto the DNA of model B using ribose C1′ atoms of the TAAT core residues (rmsd = 0.21 Å) (*36*). The proteins in the two models are oriented 2.4° apart relative to the DNA. Model A in orange, model B in blue (cross-eyed stereo).

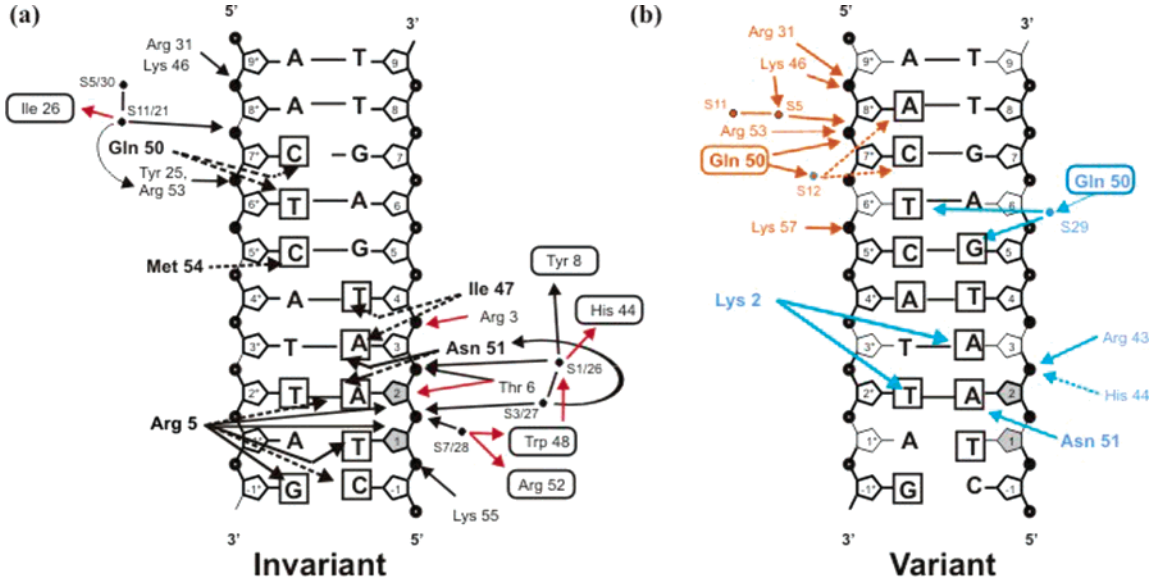


FIGURE 4: Diagrams showing the invariant (a) and variant (b) contacts between DNA and Pdx1. Direct Pdx1/DNA interactions as well as indirect protein/DNA interactions mediated through water molecules are represented. Sugars are represented as pentagons and phosphates as black circles; hydrogen bonds as solid lines and van der Waals contacts as dashed lines. Black arrows represent contacts by protein side chains; red arrows represent contacts by the protein backbone. Amino acid side chains contacting phosphates or sugars of the DNA are shown in normal text; amino acids contacting DNA bases are in bold; amino acids contacting DNA through a water molecule are in boxes. The DNA bases in contact with protein are marked with boxes around the base name. Sugars in contact with protein are shaded gray. (b) Residues participating in contacts specific to model A are in orange text and those specific to model B are in blue text.

to the DNA is intermediate between the two Pdx1 models and that the Arg 43 side chain is not in a conformation that would allow hydrogen bonding with Arg 3 (Figure 5c). In the NMR structure of the Antp/DNA complex, NOEs were reported between Arg 3 and Arg 43 (*42*). It appears that the conformation of the DNA in the crystal structure of Antp altered the DNA contact by Arg 43, disrupting the Arg 3/Arg 43 interaction. This may explain why the N-terminal arm is disordered in the Antp crystal structure but ordered in model B of Pdx1.

Binding Specificity. In order to determine if the HD of Pdx1 is sufficient to reproduce the DNA binding specificity of full-length Pdx1, we performed EMSAs using three sequences containing a TAAT core sequence and different flanking regions. We measured the relative affinity of the Pdx1 HD for the sequences. The resulting Kds for the HD were as follows: A1, 1.2 ± 0.6 nM; Glut2, 3.6 ± 1.0 nM; and albumin, 7.6 ± 2.1 nM (Figure 6). The dissociation constant of a nonspecific sequence, E1, was measured as 19.3 ± 7.8 nM. The values obtained for the HD Pdx1 follow the same pattern as found in a previous study for full-length Pdx1 bound to the same DNA sequences (A1, 5.9 ± 0.9 nM; Glut2, 24.7 ± 4.2 nM; Albumin, 24.6 ± 1.8 nM; E1, 142 nM) indicating that the DNA binding specificity of Pdx1 is primarily determined by the HD (*39*).

DISCUSSION

The ParaHox protein Pdx1 fulfills multiple functional roles in the developing pancreas, mature pancreatic β- and δ-cells, and neural cells (*1*, *4*−*6*). In different cells Pdx1 competes

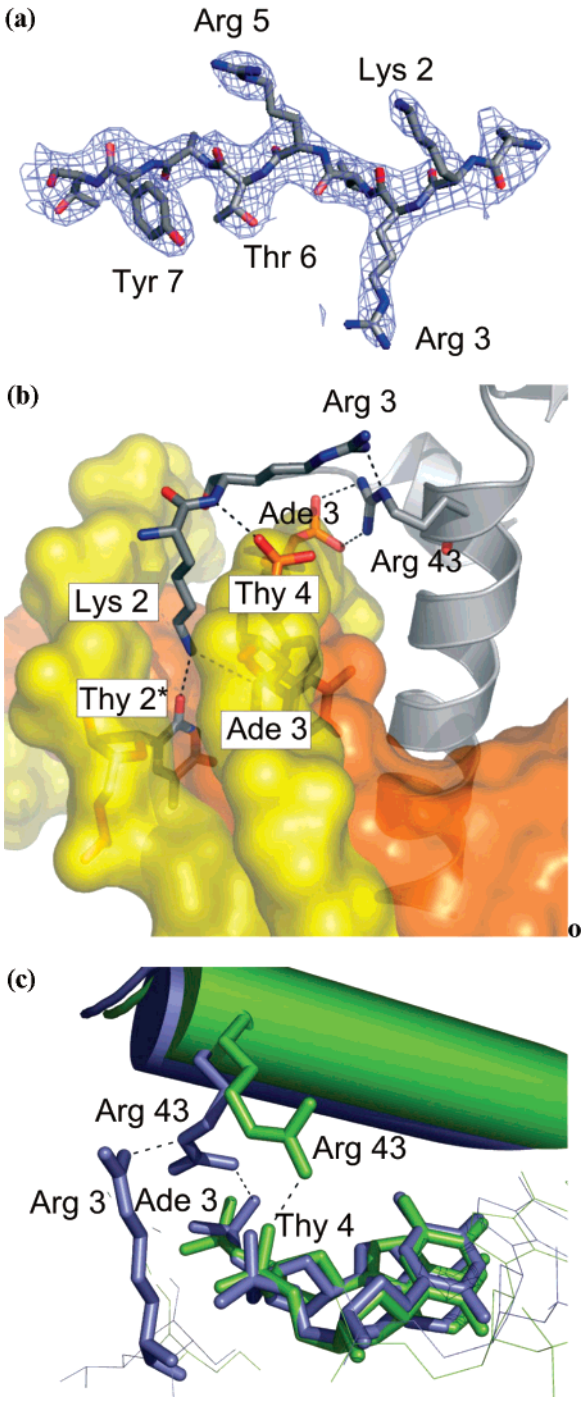
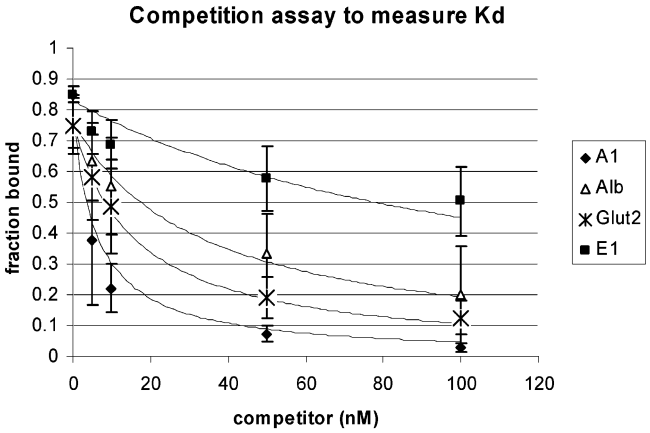


FIGURE 5: The N-terminal arm. (a) 2Fo-Fc electron density contoured at 1 σ is shown for residues 1 to 9 of the N-terminal arm of Pdx1 in model B. Density is visible for all side chains except for Asn 1. (b) Network of interactions between Lys 2, Thy 2*, Arg 3, Arg 43 and Ade 3 that order the N-terminal arm in model B. Lys 2 interacts with Thy 2* and Ade 3 in the minor groove, Arg 43 with Ade 3 in the major groove; Arg 3 connects the N-terminal arm with helix 3 through an interaction with Arg 43. Hydrogen bonds are shown as black dashed lines. A Connolly surface of the DNA is colored as in Figure 1: orange for bases, yellow for backbone. (c) Arg 3 and the N-terminal arm are disordered in the Antp crystal structure (pdb id: 9ANT) (*34*). The Antp-HD/DNA structure (in green) was aligned with the Pdx1-HD/DNA structure (model B, in blue) by using the C1′ atoms of the TAAT DNA base pairs (rmsd 0.36 Å). Arg 43 from helix 3 interacts with Thy 4 in Antp, instead of Ade 3 as in Pdx1. In the Antp crystal structure the orientation of the homeodomain in the major groove prevents the interaction between Arg 43 and Arg 3, similar to model A of Pdx1.



A1	CAGGCCCT	TAAT	GGGCCAAACGGCA
GLUT2	ATCAACT	TAAT	AATAACAGTAGA
Alb	GAAGGTT	TAAT	TCCCAGTCACAT
E1	CTCCAGCCATCTGCCGACCC		

FIGURE 6: DNA binding selectivity by the Pdx1 HD is the same as for the full-length protein. The dissociation constants for the Pdx1 HD were measured for the three TAAT-containing sequences shown: the high affinity A1 site from the rat insulin I promoter, mouse albumin (Alb) A-box, and mouse GLUT2 (GLUT) A-box. The fourth sequence, E1, derived from the E1 element of the rat insulin I promoter, does not contain a TAAT sequence and binds nonspecifically to Pdx1. The binding affinities for the full-length protein were measured by Liberzon et al. (*39*). The curves show the fit to the data from the equation for the fraction of labeled A1 DNA bound in the presence of competitor DNA. Dissociation constants were measured by competitions with the labeled A1 sequence by EMSAs (mean ± standard deviation, $n = 3$ or 4). The competitor concentrations of 0 nM, 50 nM, and 100 nM were used to fit the equation to ensure that the competitor concentration remained constant (since Pdx1 = 5 nM).

with a diverse set of homeodomains as well as the ParaHox factors Cdx and Gsx (*10*, *12*, *13*). To enhance its binding specificity and selectivity Pdx1, like other HD factors (*43*), associates with cofactors, including Pbx1 and Prep1, Pbx1 and Mrg1, and E47/NeuroD1 (*18*, *20*, *26*). The presence of two conformations of the Pdx1 HD reported in the current paper suggests that induced conformational changes in the HD may modulate DNA binding specificity in response to DNA bending.

Induced Fit Model for Pdx1 Binding to DNA. In an induced fit model local folding is coupled to site-specific DNA binding. Induced folding is a property of many transcription factors, and has been proposed as a mechanism for increasing DNA binding specificity (*44*−*46*). DNA-induced conformational changes in the N- and C-termini and side chains of HDs have been discussed when comparing the unbound and bound structures of the engrailed and Antp HDs (*42*, *47*). NMR studies showed that N-terminal residues 1−6 are disordered in solution, and become ordered upon DNA binding (*48*, *49*). The structure of a nonspecific DNA complex of the Matα2-HD indicates that the DNA sequence can influence the orientation of the HD with respect to the DNA (*50*). Finally, structures of POU proteins demonstrate that cooperative DNA binding can result from conformational changes in the DNA with no direct protein−protein contacts (*51*).

A comparison of the two Pdx1 HD complexes in the current structure indicates that the orientation of the HD relative to the DNA can vary. The HD of Pdx1 in model A

is rotated by 2.4° with respect to the DNA relative to the HD in model B. Both the Pdx1/DNA complexes are specific since the side chains contact the bases of the TAAT core sequence (*50*). While many of the protein−DNA contacts are invariant between the two Pdx1 HD complexes, each complex also forms distinct DNA interactions. Binding specificity may vary because of differences in base contacts: by Lys 2 of the N-terminal arm in model B, by an additional hydrogen bond by Asn 51, and by differences in water-mediated contacts by Gln 50 (see Figure 4). In addition, hydrogen bonding by Pdx1 to the DNA backbone differs between model A and model B, suggesting that the overall orientation of the HD can affect the DNA binding affinity as well. The orientation of the HD may be an important parameter in determining DNA binding specificity for other homeodomains. Comparing the HoxA9/Pbx1 structure with the HoxB1/Pbx1 structure indicated a rotation of helix 3 of the Hox HD relative to the DNA, which we calculated to be 3° (*52*). This rotation may contribute to the higher affinity for DNA of HoxA9 than HoxB1 (*53*).

The N-Terminal Arm. The N-terminal arm is an important determinant of DNA binding specificity (*41*), as demonstrated by chimeras generated between HoxD4 and HoxA1 (*54*), between Antp and Sex comb reduced (Scr) (*55*, *56*), and between Ubx and Dfd (*57*−*59*). Among Hox proteins residues of the N-terminal arm are more variable than those of helix 3, another indication of the importance of the arm in determining specificity (*9*, *60*). In contrast, the identity of the first 3 residues has been shown not to contribute to DNA binding specificity for some Hox/Pbx complexes (*61*, *62*).

The first 3−4 N-terminal residues are often disordered in crystal structures of HD/DNA complexes. In the Pdx1 structure, these residues are ordered in model B but not in model A. The rotation of the Pdx1 HD induces ordering of residues 1−4 of the N-terminus in model B, through cooperative ordering of a network of interacting side chains involving Arg 3 in the minor groove and Arg 43 and His 44 in the major groove (Figure 5b). The Thy 2*/Ade 3 - Lys 2 - Arg 3 - Arg 43 - His 44 - Ade 3 network of interactions depends on the orientation of helix 3 in the major groove. This ordering is not possible in model A where the recognition helix is shifted away from Ade 3 by 0.7 Å and Arg 43 is disordered. This suggests that the overall conformation of the DNA that determines the orientation of the Pdx1 HD in the major groove can facilitate ordering of the N-terminal arm. The induced order among networks of interacting side chains may contribute to cooperative recognition of DNA.

We propose that the two Pdx1 complexes represent two binding modes of Pdx1 that are relevant for different promoters: one that depends on the identity of residues 1−3 and one that does not. This agrees with an induced fit model in which the binding conformation of the HD varies in response to the conformation of the DNA, which in turn may vary in response to binding of other transcription factors (*63*). The order of the N-terminal arm in HD structures may not be the result of local interactions, but a more global response to the DNA conformation or to the DNA sequence.

DNA Binding Specificity and Comparison with Other HD Structures. DNA binding specificity of HD proteins *in vitro* is largely determined by the HD itself (*41*). We demonstrate that the HD of Pdx1 is sufficient to discriminate sequences outside of the TAAT core sequence, like the full-length protein (see Figure 6) (*39*).

The Pdx1 structure is the first HD structure reported with a lysine at position 2 instead of an arginine. Lys 2 is also found in the ParaHox factor Gsh1, and the Hox factors HoxA3−HoxA5, and HoxA7 (see alignment in Figure 1a). Arg 2 can form direct and water-mediated hydrogen bonds with core bases at positions 2, 3, and 4 (*52*, *64*, *65*). The indirect water-mediated contacts by Arg 2 have been proposed to increase the flexibility for binding site recognition of core residues at sites 2 and 4, to TNA(T/C) (*52*, *66*, *67*). Lys 2 makes fewer contacts with the TAAT core than does Arg 2: a direct hydrogen bond to the base of Thy 2*, and a van der Waals contact with Ade 3 (Figure 4b). The direct contact by Lys 2 explains the preference by Pdx1 for Ade 2.

Arg 5 and Asn 51 form hydrogen bonds with Thy 1 and with Ade 2 (in model B) and Ade 3 respectively. The Pdx1 HD does not contact base 4 through the minor groove, only via a van der Waals contact by Ile 47 through the major groove. This suggests that Pdx1 can accommodate variation in the core sequence at position 4.

In our structure the Pdx1 HD contacts bases outside the TAAT core sequence: one base 5′ (position −1) and three bases 3′ (position 5−7). The conserved Arg 5 residue forms a direct hydrogen bond with Gua −1*. Similar to HoxB1 and HoxA9, there are no direct hydrogen bonds to bases 5−7 (*52*, *68*). Instead two indirect water-mediated contacts are observed (by water S12 or S29), and van der Waals contacts by Met 54 and Gln 50 (Figure 4a). Hydrogen bonds are formed to each of the phosphate backbones of Cyt 5*, Thy 6*, Cyt 7*, and Ade 8*. These contacts likely contribute to DNA binding affinity. As has been pointed out previously, the number of interactions with the phosphate backbone may explain the high affinity of HDs for nonspecific DNA sequences (*50*, *52*, *68*).

CONCLUSIONS

Pdx1 fulfills multiple functional roles in different cellular contexts. Sequence-specific binding is coupled to conformational changes in both protein and DNA. The conformation of the DNA and therefore the DNA accessibility to the HD may vary with different Pdx1 binding partners. Since it is difficult to rationalize DNA binding specificity by the HD from direct side chain contacts alone, we propose an induced fit mechanism in which the binding configuration of the HD depends on the overall DNA conformation. From our analysis of the Pdx1 structure, the orientation of the HD within the major groove can vary by several degrees and influence DNA contacts through cooperative ordering of side chain networks, ordering of the N-terminal arm, and water-mediated contacts.

Structural studies of Pdx1 with alternative DNA sequences, and of Pdx1 in complex with protein partners, will be required to test this proposal.

ACKNOWLEDGMENT

We thank Dr. German for providing the pBAT14.shPdx1 plasmid.

REFERENCES

1. McKinnon, C. M., and Docherty, K. (2001) Pancreatic duodenal homeobox-1, PDX-1, a major regulator of beta cell identity and function, *Diabetologia 44*, 1203−1214.

2. Hui, H., and Perfetti, R. (2002) Pancreas duodenum homeobox-1 regulates pancreas development during embryogenesis and islet cell function in adulthood, *Eur. J. Endocrinol. 146*, 129−141.
3. Wu, K. L., Gannon, M., Peshavaria, M., Offield, M. F., Henderson, E., Ray, M., Marks, A., Gamer, L. W., Wright, C. V., and Stein, R. (1997) Hepatocyte nuclear factor 3beta is involved in pancreatic beta-cell-specific transcription of the pdx-1 gene, *Mol. Cell. Biol. 17*, 6002−6013.
4. Leonard, J., Peers, B., Johnson, T., Ferreri, K., Lee, S., and Montminy, M. R. (1993) Characterization of somatostatin transactivating factor-1, a novel homeobox factor that stimulates somatostatin expression in pancreatic islet cells, *Mol. Endocrinol. 7*, 1275−1283.
5. Perez-Villamil, B., Schwartz, P. T., and Vallejo, M. (1999) The pancreatic homeodomain transcription factor IDX1/IPF1 is expressed in neural cells during brain development, *Endocrinology 140*, 3857−3860.
6. Schwartz, P. T., Perez-Villamil, B., Rivera, A., Moratalla, R., and Vallejo, M. (2000) Pancreatic homeodomain transcription factor IDX1/IPF1 expressed in developing brain regulates somatostatin gene transcription in embryonic neural cells, *J. Biol. Chem. 275*, 19106−19114.
7. Ferrier, D. E., Dewar, K., Cook, A., Chang, J. L., Hill-Force, A., and Amemiya, C. (2005) The chordate ParaHox cluster, *Curr. Biol. 15*, R820−822.
8. Garcia-Fernandez, J. (2005) The genesis and evolution of homeobox gene clusters, *Nat. Rev. Genet. 6*, 881−892.
9. Banerjee-Basu, S., and Baxevanis, A. D. (2001) Molecular evolution of the homeodomain family of transcription factors, *Nucleic Acids Res. 29*, 3258−3269.
10. Rosanas-Urgell, A., Marfany, G., and Garcia-Fernandez, J. (2005) Pdx1-related homeodomain transcription factors are distinctly expressed in mouse adult pancreatic islets, *Mol. Cell. Endocrinol. 237*, 59−66.
11. Miller, C. P., McGehee, R. E., Jr., and Habener, J. F. (1994) IDX-1: a new homeodomain transcription factor expressed in rat pancreatic islets and duodenum that transactivates the somatostatin gene, *EMBO J. 13*, 1145−1156.
12. Rudnick, A., Ling, T. Y., Odagiri, H., Rutter, W. J., and German, M. S. (1994) Pancreatic beta cells express a diverse set of homeobox genes, *Proc. Natl. Acad. Sci. U.S.A. 91*, 12203−12207.
13. Habener, J. F., Kemp, D. M., and Thomas, M. K. (2005) Minireview: transcriptional regulation in pancreatic development, *Endocrinology 146*, 1025−1034.
14. Mendel, D. B., and Crabtree, G. R. (1991) HNF-1, a member of a novel class of dimerizing homeodomain proteins, *J. Biol. Chem. 266*, 677−680.
15. Laughon, A. (1991) DNA binding specificity of homeodomains, *Biochemistry 30*, 11357−11367.
16. Wilson, D. S., Sheng, G., Jun, S., and Desplan, C. (1996) Conservation and diversification in homeodomain-DNA interactions: a comparative genetic analysis, *Proc. Natl. Acad. Sci. U.S.A. 93*, 6886−6891.
17. Moens, C. B., and Selleri, L. (2006) Hox cofactors in vertebrate development, *Dev. Biol. 291*, 193−206.
18. Goudet, G., Delhalle, S., Biemar, F., Martial, J. A., and Peers, B. (1999) Functional and cooperative interactions between the homeodomain PDX1, Pbx, and Prep1 factors on the somatostatin promoter, *J. Biol. Chem. 274*, 4067−4073.
19. Peers, B., Sharma, S., Johnson, T., Kamps, M., and 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.
20. Swift, G. H., Liu, Y., Rose, S. D., Bischof, L. J., Steelman, S., Buchberg, A. M., Wright, C. V., and MacDonald, R. J. (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.
21. Phelan, M. L., Rambaldi, I., and Featherstone, M. S. (1995) Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif, *Mol. Cell. Biol. 15*, 3989−3997.
22. Neuteboom, S. T., Peltenburg, L. T., van Dijk, M. A., and Murre, C. (1995) The hexapeptide LFPWMR in Hoxb-8 is required for cooperative DNA binding with Pbx1 and Pbx2 proteins, *Proc. Natl. Acad. Sci. U.S.A. 92*, 9166−9170.
23. Dutta, S., Gannon, M., Peers, B., Wright, C., Bonner-Weir, S., and Montminy, M. (2001) PDX:PBX complexes are required for normal proliferation of pancreatic cells during development, *Proc. Natl. Acad. Sci. U.S.A. 98*, 1065−1070.
24. Chang, C. P., Shen, W. F., Rozenfeld, S., Lawrence, H. J., Largman, C., and Cleary, M. L. (1995) Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins, *Genes Dev. 9*, 663−674.
25. Glick, E., Leshkowitz, D., and Walker, M. D. (2000) Transcription factor BETA2 acts cooperatively with E2A and PDX1 to activate the insulin gene promoter, *J. Biol. Chem. 275*, 2199−2204.
26. Peers, B., Leonard, J., Sharma, S., Teitelman, G., and Montminy, M. R. (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.
27. Ohneda, K., Mirmira, R. G., Wang, J., Johnson, J. D., and German, M. S. (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.
28. Stoffers, D. A., Ferrer, J., Clarke, W. L., and Habener, J. F. (1997) Early-onset type-II diabetes mellitus (MODY4) linked to IPF1, *Nat. Genet. 17*, 138−139.
29. Macfarlane, W. M., Frayling, T. M., Ellard, S., Evans, J. C., Allen, L. I., Bulman, M. P., Ayers, S., Shepherd, M., Clark, P., Millward, A., Demaine, A., Wilken, T., Docherty, K., and Hattersley, A. T. (2000) Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes, *J. Clin. Invest. 106*, 717.
30. Weng, J., Macfarlane, W. M., Lehto, M., Gu, H. F., Shepherd, L. M., Ivarsson, S. A., Wibell, L., Smith, T., and Groop, L. C. (2001) Functional consequences of mutations in the MODY4 gene (IPF1) and coexistence with MODY3 mutations, *Diabetologia 44*, 249−258.
31. Mitchell, S. M., and Frayling, T. M. (2002) The role of transcription factors in maturity-onset diabetes of the young, *Mol. Genet. Metab. 77*, 35−43.
32. Macfarlane, W. M., Frayling, T. M., Ellard, S., Evans, J. C., Allen, L. I., Bulman, M. P., Ayres, S., Shepherd, M., Clark, P., Millward, A., Demaine, A., Wilkin, T., Docherty, K., and Hattersley, A. T. (1999) Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes, *J. Clin. Invest. 104*, R33−39.
33. Leslie, A., Brick, P., and Wonacott, A. (1992) Recent changes to the MOSFLM package for processing film and image plate data, *Joint CCP4/ESF-EAMCB Newsl. Protein Crystallogr*.
34. Fraenkel, E., and Pabo, C. O. (1998) Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex, *Nat. Struct. Biol. 5*, 692−697.
35. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta. Crystallogr. D Biol. Crystallogr. 54*, 905−921.
36. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta. Crystallogr. A 47*, 110−119.
37. DeLano, W. L. (2002) The PyMOL Molecular Graphics System. http://www.pymol.org.
38. Lavery, R., and Sklenar, H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids, *J. Biomol. Struct. Dyn. 6*, 63−91.
39. Liberzon, A., Ridner, G., and Walker, M. D. (2004) Role of intrinsic DNA binding specificity in defining target genes of the mammalian transcription factor PDX1, *Nucleic Acids Res. 32*, 54−64.
40. Ades, S. E., and Sauer, R. T. (1994) Differential DNA-binding specificity of the engrailed homeodomain: the role of residue 50, *Biochemistry 33*, 9187−9194.
41. Mann, R. S. (1995) The specificity of homeotic gene function, *Bioessays 17*, 855−863.
42. Qian, Y. Q., Otting, G., Billeter, M., Muller, M., Gehring, W., and Wuthrich, K. (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.
43. Pearson, J. C., Lemons, D., and McGinnis, W. (2005) Modulating Hox gene functions during animal body patterning, *Nat. Rev. Genet. 6*, 893−904.

44. Spolar, R. S., and Record, M. T., Jr. (1994) Coupling of local folding to site-specific binding of proteins to DNA, *Science 263*, 777−784.
45. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol. 6*, 197−208.
46. Gunther, S., Rother, K., and Frommel, C. (2006) Molecular flexibility in protein-DNA interactions, *Biosystems 85*, 126−136.
47. Fraenkel, E., Rould, M. A., Chambers, K. A., and 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, *J. Mol. Biol. 284*, 351−361.
48. Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., and Wuthrich, K. (1989) The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors, *Cell 59*, 573−580.
49. Qian, Y. Q., Otting, G., Furukubo-Tokunaga, K., Affolter, M., Gehring, W. J., and Wuthrich, K. (1992) NMR structure determination reveals that the homeodomain is connected through a flexible linker to the main body in the Drosophila Antennapedia protein, *Proc. Natl. Acad. Sci. U.S.A. 89*, 10738−10742.
50. Aishima, J., and Wolberger, C. (2003) Insights into nonspecific binding of homeodomains from a structure of MATalpha2 bound to DNA, *Proteins 51*, 544−551.
51. Klemm, J. D., and Pabo, C. O. (1996) Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage, *Genes Dev. 10*, 27−36.
52. LaRonde-LeBlanc, N. A., and Wolberger, C. (2003) Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior, *Genes Dev. 17*, 2060−2072.
53. Pellerin, I., Schnabel, C., Catron, K. M., and Abate, C. (1994) Hox proteins have different affinities for a consensus DNA site that correlate with the positions of their genes on the hox cluster, *Mol. Cell. Biol. 14*, 4532−4545.
54. Phelan, M. L., Sadoul, R., and Featherstone, M. S. (1994) Functional differences between HOX proteins conferred by two residues in the homeodomain N-terminal arm, *Mol. Cell. Biol. 14*, 5066−5075.
55. Zeng, W., Andrew, D. J., Mathies, L. D., Horner, M. A., and Scott, M. P. (1993) Ectopic expression and function of the Antp and Scr homeotic genes: the N terminus of the homeodomain is critical to functional specificity, *Development (Cambridge, England) 118*, 339−352.
56. Furukubo-Tokunaga, K., Flister, S., and Gehring, W. J. (1993) Functional specificity of the Antennapedia homeodomain, *Proc. Natl. Acad. Sci. U.S.A. 90*, 6360−6364.
57. Frazee, R. W., Taylor, J. A., and Tullius, T. D. (2002) Interchange of DNA-binding modes in the deformed and ultrabithorax homeodomains: a structural role for the N-terminal arm, *J. Mol. Biol. 323*, 665−683.
58. Lin, L., and McGinnis, W. (1992) Mapping functional specificity in the Dfd and Ubx homeo domains, *Genes Dev. 6*, 1071−1081.
59. Mann, R. S., and Hogness, D. S. (1990) Functional dissection of Ultrabithorax proteins in D. melanogaster, *Cell 60*, 597−610.
60. Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E., and Beachy, P. A. (1994) The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins, *EMBO J. 13*, 3551−3560.
61. Remacle, S., Shaw-Jackson, C., Matis, C., Lampe, X., Picard, J., and Rezsohazy, R. (2002) Changing homeodomain residues 2 and 3 of Hoxa1 alters its activity in a cell-type and enhancer dependent manner, *Nucleic Acids Res. 30*, 2663−2668.
62. Phelan, M. L., and Featherstone, M. S. (1997) Distinct HOX N-terminal arm residues are responsible for specificity of DNA recognition by HOX monomers and HOX.PBX heterodimers, *J. Biol. Chem. 272*, 8635−8643.
63. Smith, D. L., Desai, A. B., and Johnson, A. D. (1995) DNA bending by the a1 and alpha 2 homeodomain proteins from yeast, *Nucleic Acids Res. 23*, 1239−1243.
64. Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA, *Cell 82*, 709−719.
65. Hovde, S., Abate-Shen, C., and Geiger, J. H. (2001) Crystal structure of the Msx-1 homeodomain/DNA complex, *Biochemistry 40*, 12013−12021.
66. Shen, W. F., Rozenfeld, S., Lawrence, H. J., and Largman, C. (1997) The Abd-B-like Hox homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target, *J. Biol. Chem. 272*, 8198−8206.
67. Chang, C. P., Brocchieri, L., Shen, W. F., Largman, C., and Cleary, M. L. (1996) Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus, *Mol. Cell. Biol. 16*, 1734−1745.
68. Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., and 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.
69. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673−4680.

BI060969L