Solution Structure of the K50 Class Homeodomain PITX2 Bound to DNA and Implications for Mutations That Cause Rieger Syndrome[†,‡]

Beth A. Chaney,[§] Kimber Clark-Baldwin,[§] Vrushank Dave,[||,⊥] Jun Ma,[||] and Mark Rance*[,§]

Department of Molecular Genetics, Biochemistry, and Microbiology, College of Medicine, University of Cincinnati, 231 Albert Sabin Way, Medical Sciences Building, Cincinnati, Ohio 45267-0524, Division of Developmental Biology, Cincinnati Children's Hospital Research Foundation, 3333 Burnet Avenue, Cincinnati, Ohio 45229, and Division of Pulmonary Biology, Cincinnati Children's Hospital Research Foundation, 3333 Burnet Avenue, Cincinnati, Ohio 45229

Received December 20, 2004; Revised Manuscript Received March 20, 2005

ABSTRACT: We have determined the solution structure of a complex containing the K50 class homeodomain Pituitary homeobox protein 2 (PITX2) bound to its consensus DNA site (TAATCC). Previous studies have suggested that residue 50 is an important determinant of differential DNA-binding specificity among homeodomains. Although structures of several homeodomain−DNA complexes have been determined, this is the first structure of a native K50 class homeodomain. The only K50 homeodomain structure determined previously is an X-ray crystal structure of an altered specificity mutant, Engrailed Q50K (EnQ50K). Analysis of the NMR structure of the PITX2 homeodomain indicates that the lysine at position 50 makes contacts with two guanines on the antisense strand of the DNA, adjacent to the TAAT core DNA sequence, consistent with the structure of EnQ50K. Our evidence suggests that this side chain may make fluctuating interactions with the DNA, which is complementary to the crystal data for EnQ50K. There are differences in the tertiary structure between the native K50 structure and that of EnQ50K, which may explain differences in affinity and specificity between these proteins. Mutations in the human *PITX2* gene are responsible for Rieger syndrome, an autosomal dominant disorder. Analysis of the residues mutated in Rieger syndrome indicates that many of these residues are involved in DNA binding, while others are involved in formation of the hydrophobic core of the protein. Overall, the role of K50 in homeodomain recognition is further clarified, and the results indicate that native K50 homeodomains may exhibit differences from altered specificity mutants.

The homeodomain is an evolutionarily conserved protein domain found in organisms ranging from yeast and *Drosophila* to humans (*1*−*5*). Homeodomain-containing proteins are known to play important roles in such diverse activities as embryonic pattern formation, cell-type specification, and differentiation (*3*). This domain is responsible for recognizing specific DNA sequences, thereby recruiting the corresponding transcription factors to specific target genes. Most DNA sites that homeodomains recognize consist of only 6 base pairs, with a common TAAT core sequence followed by two base pairs that have been proposed to define the specificity (*3*, *4*). The homeodomain consists of a self-folding, stable protein domain of 60 amino acids. Previous studies of homeodomains have shown that they have a compact 3-helix structure and a flexible N-terminal arm (*3*, *5*). The third helix, called the recognition helix, makes specific contacts within the major groove of the DNA. Homeodomains have been studied extensively (*1*−*5*), including genetic, biochemical, and structural analyses, due to their critical role in cellular processes and to the fact that they serve as a valuable model for probing the physical basis of protein−DNA interactions. Although the overall topology of the homeodomain motif and its docking arrangement on duplex DNA are now generally well-defined, fundamental questions remain, particularly in regard to the role of amino acid side chains in defining the specificity of homeodomain−DNA binding and the nature of the interactions of these side chains with the DNA binding sites. Homeodomains have evolved different DNA specificities in part by altering the amino acid residue at position 50, which can interact with base pairs 5 and 6, and to a lesser extent, base pair 4, in the TAATNN consensus binding site. A previous study has shown that each of 6 different amino acids tested at position 50 confers a different DNA binding specificity (*6*). Tucker-Kellogg et al. (*7*) and others have emphasized the point that the degree of specificity of a homeodomain for its particular DNA binding sites depends on the identity of the amino acid residue in position 50. Most homeodomains contain a glutamine residue at this position, and are therefore referred to as Q50 homeodomains. Q50 homeodomains prefer DNA sequences such as TAATTA and TAATGG. The homeodomain of Bicoid, which is a *Drosophila* morphogenetic protein, contains a lysine at

[†] This work is supported by Grant GM063855 from the National Institutes of Health (to M.R.) and Grant 0255347N from the American Heart Association (to J.M.).

[‡] PDB accession code for PITX2 homeodomain bound to DNA: 1YZ8.

* Author to whom all correspondence should be addressed. Tel: (513) 558-0066. Fax: (513) 558-8474. E-mail: Mark.Rance@UC.Edu.

[§] University of Cincinnati.

[||] Division of Developmental Biology, Cincinnati Children's Hospital Research Foundation.

[⊥] Division of Pulmonary Biology, Cincinnati Children's Hospital Research Foundation (present address).

10.1021/bi0473253 CCC: $30.25
Published on Web 04/26/2005

position 50 and is the founding member of the K50 class of homeodomains (*8*). The K50 class of homeodomains recognizes a consensus DNA sequence of TAATCC. Much attention has been focused on the consequences of lysine being located at position 50, largely due to the fact that the most dramatic examples of altered DNA specificity occur when a lysine is either introduced or replaced at position 50. For example, when Q50 in Engrailed is mutated to an alanine, the Q50A mutant has an affinity and specificity that are very similar to those of the wild-type protein, but when mutated to a lysine, the specificity changes from TAATTA to TAATCC, clearly demonstrating the important role played by the residue in position 50, especially in the case of K50, in defining the specificity of DNA binding (*9*−*11*). Percival-Smith et al. (*12*) investigated wild-type and Q50K mutant Fushi tarazu homeodomains in conjunction with altering the base pairs at positions 5 and 6 in the binding site, and found that differences in K_D of ∼100-fold are observed when the binding site is not the optimal one. In addition to position 50, position 47 has a role in defining specificity for some homeodomains in correlation with base pair 4 of the binding site, especially when the residue is phenylalanine or arginine (*13*, *14*).

Although structures of several homeodomains and homeodomain−DNA complexes have been determined by X-ray crystallography or NMR spectroscopy (*15*), including representatives of the wild-type Q50, S50, C50, G50, and I50 classes of homeodomains (*16*−*20*), the only experimentally determined K50 homeodomain structure available is an X-ray crystal structure of an altered specificity mutant, Engrailed Q50K (EnQ50K[1]), bound to the TAATCC site (*7*). The latter study found that the side chain of K50 projects into the major groove of the DNA and makes hydrogen bond contacts with the O6 and N7 atoms of the guanines at base pairs 5 and 6 of the complementary strand of the TAATCC binding site. This is the only case in which direct hydrogen bond contacts have been reported for amino acid residue 50 in any homeodomain−DNA complex structure. Unfortunately, the relevance of the EnQ50K studies, or analyses of other mutants such as Paired S50K (*6*) and Fushi tarazu Q50K (*21*), to the case of native K50 homeodomains is unclear in the absence of experimental structural data for a native K50 homeodomain. For example, the identity of the amino acid residue at position 54 seems to be constrained by the residue at position 50. A glutamine at position 50 allows for many different residues to be present at position 54, with Met being the most abundant. However, Met54 is never found when position 50 is lysine (*22*). Determining the biological relevance of studies of single site mutants should take into account possible covariation of residues (*23*). For example, structural studies of an EnQ50A mutant have been conducted (*11*) in order to provide additional information concerning the role of residue 50 in general, and Q50 in particular; however, a phage display selection of Engrailed mutants failed to recover a Q50A mutant: the only Q50A mutant recovered also contained a I47T mutation (*24*). Another issue concerning the Engrailed Q50K mutant is the observation that it binds to the consensus TAATCC site with an unusually high affinity, which approaches the picomolar range (*9*). There is no evidence that natural K50 class homeodomains have such a high affinity for DNA (*25*, *26*). The full-length PITX2 protein has a K_D of 50 nM (*25*). A K_D was determined for the Q50K mutant of the Fushi tarazu homeodomain, and this value was found to be 0.63 nM (*12*), which is a much lower affinity than the EnQ50K mutant. The K_D for the PITX2 homeodomain alone was found to be 2.6 ± 0.38 nM (see Supporting Information), which is comparable to the Fushi tarazu mutant, and also a much lower affinity than the Engrailed Q50K mutant. Moreover, the X-ray structure of EnQ50K reveals two conformations with the side chain of K50 contacting either the 5th or 6th position on the antisense strand of the DNA. Whether or not natural K50 homeodomains exhibit these two conformations in a static state and whether or not the side chain of K50 exhibits a fluctuating state were unknown. Another biochemical property shared by the wild-type K50 class homeodomains Bicoid and PITX2 is that they both can recognize naturally occurring DNA sites that deviate from the consensus site TAATCC (*2*, *21*, *27*−*29*). In contrast, a Q50K mutant of the Fushi tarazu homeodomain is unable to bind any nonconsensus DNA binding sites tested (*21*). Together, these considerations underscore the importance of obtaining solution structures of native K50 class homeodomains.

The question regarding side-chain conformational heterogeneity, referred to above in the context of the observations concerning the K50 side chain in the EnQ50K crystal structure, is broader in scope and of fundamental importance for understanding the full range of interactions that can occur at a protein−DNA interface. Crystallographic studies have generally indicated that there are several conserved and relatively stable contacts at the homeodomain−DNA interface. In several instances, such as the aforementioned case of K50 in the EnQ50K structure and the case of Gln50 in the crystal structure of an Even-skipped homeodomain complex (*30*), multiple, significantly populated conformations are observed for the side chain, while the nearly invariant asparagine in position 51 is observed to make very stable contacts with the N6 and N7 atoms of the adenine base in position 3 of the consensus TAAT core binding site. On the other hand, NMR studies (*31*, *32*) and molecular dynamics simulations (*31*, *33*) have provided strong indications of a dynamic, fluctuating environment encompassing some of the key amino acid side chains at the interface, most importantly, the side chains of asparagine 51 and of the position 50 residue. Billeter and co-workers (*31*) proposed that, at least in the case of Antennapedia, the homeodomain achieves specificity through a fluctuating network of short-lived contacts that allow it to recognize DNA without the entropic cost that would result if side chains were immobilized upon DNA binding. Significant interest has been expressed in the literature (*7*, *31*, *33*, *34*) for obtaining experimental data on native K50 homeodomains in order to shed further light on these fundamental issues.

For the structure determination studies reported herein, we chose the well-characterized homeodomain PITX2 as our

[1] Abbreviations: PITX2, pituitary homeobox protein 2; EnQ50K, Engrailed Q50K mutant; IPTG, isopropyl-β-D-thiogalactopyranoside; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid; PBS, phosphate buffer solution; TCB, thrombin cleavage buffer; DTT, dithiothreitol; NOESY, nuclear Overhauser enhancement spectroscopy; HSQC, heteronuclear single quantum correlation; NOE, nuclear Overhauser effect; rmsd, root-mean-square deviation; bb, backbone; PDB, Protein Data Bank; HD, homeodomain.

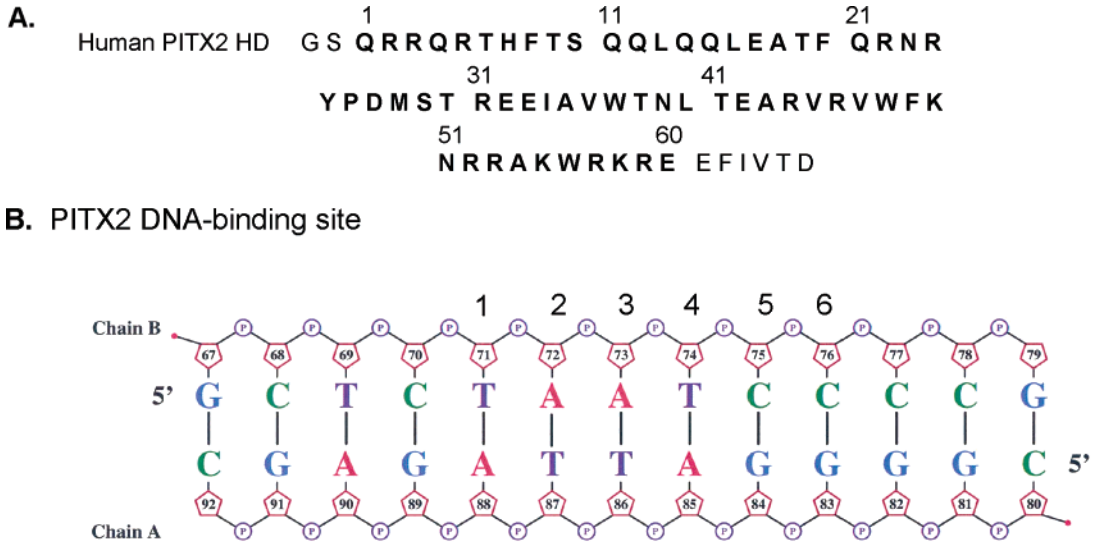

FIGURE 1: (a) Amino acid sequence of the PITX2 homeodomain used for the structural studies. The extra residues at the N- and C-termini are part of the expression system. (b) DNA sequence of the binding site used in the structural studies. The DNA sequence consists of the TAATCC binding site surrounded by residues to confer stability to the double strand.

representative model for the K50 class of homeodomains. PITX2 is a transcription factor that is found in many developing tissues in vertebrate embryos. It has been shown to be expressed in the brain, heart, pituitary, mandibular and maxillary regions, eye, gut, limbs, and umbilicus (*35*−*38*). The homeodomain is just one domain of the full-length PITX2 protein. There have been three major isoforms of PITX2 identified, and these isoforms are produced by alternative splicing and the use of different promoters (*35*, *36*, *39*−*41*). All of the isoforms contain different N-terminal domains, while the homeodomain and C-terminal domains are identical. The C-terminal region contains a transcriptional activation domain. Mice that have been genetically engineered to be homozygous for a *pitx2* null allele have a single atrium, arrested development of the pituitary gland, numerous defects of the eye, and altered development of the mandibular and maxillary regions (*41*−*43*). Mutations in *PITX2* are known to be a cause of Rieger syndrome (*44*−*48*), a genetic disease characterized by defects in the eye, facial features, and umbilicus. Sequencing of DNA from human patients has shown that many mutations in *PITX2* result in single amino acid substitutions within the homeodomain region (*45*−*48*). Although modeling and biochemical studies have provided useful insights into the effect of these mutations (refer to Table 2), no structure-based information is available on how these mutations might affect the PITX2 homeodomain properties.

In the present study, the NMR solution structure of the PITX2 homeodomain bound to its consensus DNA binding site is reported. This represents the first experimentally determined structure of a native K50 class homeodomain. The results reveal a tertiary structure similar to other homeodomains, with K50 making contacts with the two guanines adjacent to the TAAT core DNA sequence. Evidence indicates that K50 may interact with DNA in a flexible manner. The tertiary structure of PITX2 indicates that the first two helices are slightly closer to each other than in the EnQ50K mutant, and the third helix has a slightly different position relative to the other helices. We discuss how these structural properties of PITX2 might affect the biochemical functions of the PITX2 homeodomain. On the basis of the solution structure, we also discuss the effect of mutations found in Rieger syndrome patients on PITX2 homeodomain functions.

MATERIALS AND METHODS

Expression of the PITX2 Homeodomain. The expression plasmid pGEX-1λt-PITX2HD consists of a glutathione S-transferase tag and a thrombin cleavage site prior to the PITX2 homeodomain sequence. There are two extra residues at the N-terminus as a result of thrombin cleavage, and six extra residues at the C-terminus that are part of the expression system. The final protein sequence is shown in Figure 1.

The PITX2 homeodomain was obtained by growing *Escherichia coli* strain BL21-Star (Invitrogen) transformed with pGEX-1λt-PITX2HD in minimal medium [0.85 g/L NaOH, 10.5 g/L K_2HPO_4, 12 g/L Na_2HPO_4, 6 g/L KH_2PO_4, 1 g/L NaCl, 6 mg/L $CaCl_2$, 13.2 mL/L concentrated (12.2 N) HCl, nucleotides (0.5 g/L adenine, 0.65 g/L guanosine, 0.2 g/L thymine, 0.5 g/L uracil, 0.2 g/L cytosine), vitamins (1 mg/L choline chloride, 1 mg/L pyridoxal phosphate, 100 μg/L riboflavin, 50 mg/L thiamine, 50 mg/L niacin, 1 mg/L biotin), and trace elements (107 μg/L $MgCl_2 \cdot 6H_2O$, 20 μg/L $FeCl_2 \cdot 4H_2O$, 0.7 μg/L $CaCl_2 \cdot 2H_2O$, 0.26 μg/L H_3BO_3, 0.16 μg/L $MnCl_2 \cdot 4H_2O$, 16 ng/L $CuCl_2 \cdot 2H_2O$, 2.4 μg/L $Na_2MoO_4 \cdot 2H_2O$, 10 μM $FeCl_3$, 135 mM $CaCl_2$, 50 μM $ZnSO_4$)] containing 150 mg/L ampicillin, 4 g/L glucose, and 1 g/L NH_4Cl. Half a liter of bacterial culture was grown in baffled flasks in an incubator shaker at 37 °C until saturation (A_{600} ∼ 5.0). This culture was spun down (2000*g*, 10 min) and resuspended in 1 L of minimal medium enriched with 10% $^{15}N^{13}C$-Isogro, ^{13}C-glucose, and ^{15}N-ammonium chloride (Sigma-Aldrich). Expression was then induced by addition of IPTG to a final concentration of 0.1 mM, followed by growth at 20 °C for approximately 24 h.

Purification of the PITX2 Homeodomain. Cells were lysed with lysozyme and sonication. Cleared lysate was applied to a glutathione sepharose column and washed with PBS buffer, followed by thrombin cleavage buffer (50 mM trizma hydrochloride, 150 mM NaCl, 2.5 mM $CaCl_2$, pH 8.0). The resin was resuspended in thrombin cleavage buffer (TCB)

and transferred to a 50 mL conical tube. The homeodomain was cleaved from the glutathione S-transferase fusion tag using 1 mg of thrombin for 3 h at 4 °C. Nearly complete cleavage was obtained during this time as measured by SDS−PAGE. The cleaved protein was then eluted from the resin using 5 bed volumes of TCB. It was loaded onto a SP sepharose fast flow column (2 mL bed volume, Amersham), washed with washing buffer (10 mM NaH_2PO_4, 250 mM NaCl, pH 7.0), and eluted with buffer containing a higher salt concentration (10 mM NaH_2PO_4, 1 M NaCl, pH 7.0). Fractions containing the homeodomain were identified by Abs_{278} (extinction coefficient = 18350 cm^{-1} M^{-1}), pooled, and dialyzed overnight at 4 °C in 10 mM NaH_2PO_4, pH 7.0. Protein yields were ~4.5 mg/L of cell growth. The consensus DNA duplex (IdtDNA) (see Figure 1) was added to give a 1:1 protein:DNA ratio, and the complex was concentrated by burying the dialysis bag in Spectra/Gel Absorbent (Spectrum), or Aquacide (Calbiochem). Samples were dialyzed in 10 mM NaH_2PO_4, pH 7.0 after concentration. Complete protease inhibitors (Roche, 1 tablet in 3 mL, add 1 μL), leupeptin (0.3 mM), DTT (2 mM), and Pefabloc (0.2 mM, Roche) were all added to inhibit proteases. Sodium azide (6 mM stock, add 1 μL to 540 μL NMR sample) was added to prevent bacterial growth in the sample. The final sample concentration was approximately 1 mM, in 90% 10 mM NaH_2PO_4, pH 7.0 and 10% D_2O.

Determination of K_D. Measurements of K_D were taken by following the procedure of Dave et al. (*2*). The DNA probe concentrations used in this analysis were 1, 2, 4, 8,16, 20, and 40 nM. Quantitative gel shift assays were performed by measuring the bound and free fractions of the probes with a PhosphorImager as previously described (*21*). The data were analyzed using Microsoft Excel (linear regression analysis) to determine the K_D value ($-1/K_D$ = slope of the plot of bound/free against bound DNA). These results can be seen in the Supporting Information.

NMR Structure Determination. All NMR experiments were carried out on Varian Inova 600 and 800 MHz spectrometers. The sample temperature was set to 295 K. Spectra were referenced to an external DSS standard.

Protein Assignments. Protein ^1H, ^{13}C, and ^{15}N resonance assignments were obtained primarily from heteronuclear-edited NMR spectra, using conventional triple resonance ^1H-$\{^{13}C,^{15}N\}$ NMR probes. The pulse programming codes were written in-house. Approximately 92% of assignable atoms were assigned. Sequence-specific assignment of the backbone H^N, N, C′, C^α, and C^β resonances were obtained from 3D HNCO, HN(CO)CA, HNCA, CBCA(CO)NH, and HNCACB (*49*−*55*) spectra. Assignment of the aliphatic side-chain resonances was accomplished using a combination of 3D ^{15}N-edited-TOCSY-HSQC (*56*, *57*), H(CCO)NH-TOCSY, and HBHA(CBCACO)NH spectra (*55*). Aromatic ^1H and ^{13}C resonances were obtained from a combination of 2D HMQC, 2D HMQC-TOCSY, 3D HMQC-TOCSY, and 2D NOESY-HMQC spectra (*58*, *59*). A HNHA experiment was performed to assign H^α and to obtain coupling constants (*60*).

DNA Assignments. Resonance assignments for unlabeled DNA bound to ^{13}C,^{15}N-labeled protein were obtained using standard assignment methods for DNA (*61*). The data was obtained with doubly $^{13}C/^{15}N$-filtered NOESY and ω_2-filtered TOCSY experiments (*62*, *63*).

Structural Constraints. The main source of structural information was the proton−proton distance constraints identified from NOESY spectra. Three-dimensional ^{15}N-NOESY-HSQC experiments (*64*) using 50−125 ms mixing times were used for intramolecular restraints in the homeodomains, along with a ^{13}C-NOESY-HSQC experiment using a 150 ms mixing time (*55*).

Intramolecular distance restraints for the DNA were obtained from an r^{-6} scaling of cross-peak volumes in the NOESY spectra. Upper and lower bounds were calibrated on the cytosine intraresidue H5−H6 NOE and set to ± 15% of the calculated distance for base and H1′ protons. Restraint boundaries to other sugar protons were widened an additional 10% to account for effects of spin diffusion. Restraints from the longer mixing time 125 ms experiment were assigned a lower bound of 3 Å and an upper bound of 5 Å.

Intermolecular restraints between the protein and DNA were obtained from 2D $^{13}C(\omega_1)$-edited, $[^{13}C,^{15}N](\omega_2)$-filtered NOESY spectra (*63*, *65*, *66*). The NOEs were assigned manually, and only unambiguously assigned peaks were used as restraints in the docking calculation. Weak peaks were assigned an upper distance limit of 6.0 Å, while medium peaks had an upper distance limit of 5.0 Å, and stronger peaks an upper distance limit of 4.0 Å.

Data Processing and Analysis. Raw NMR data was processed using NMRPipe (*67*). Linear prediction was used in the t_1 dimension for 2D spectra and in the t_1 and t_2 dimensions for 3D spectra, using sinebell window functions for apodization and zero filling in all dimensions. Spectra were viewed and analyzed using the Sparky graphical interface (*68*). This program was used to pick peaks and integrate them using a Lorentzian function.

Structure Calculation. Referenced chemical shift assignments and peak intensities from Sparky were entered into the structure calculation program CYANA (*69*, *70*). CYANA consists of an automated NOE assignment program, CANDID, which automatically assigns all NOESY cross-peaks, taking into account nearness of chemical shift, network anchoring, ambiguous distance constraints, and constraint combination (*70*). The structure is then calculated using the DYANA algorithm, which calculates structures using torsion angle dynamics (*69*). Calibration constants for peak intensities versus upper distance limits were determined automatically by CYANA. The 20 lowest energy conformers were retained after structure calculation and used for docking to DNA.

Docking of the Protein to the DNA. The protein was docked to the DNA using the AMBER all-atom force field with the generalized Born solvation model (*71*, *72*). The 20 CYANA structures with the lowest values of target function were docked onto canonical B-form DNA. This was chosen as the starting DNA structure, since NOESY spectra for the complex indicated the DNA to be close to B-form. Starting structures of the complex were generated by systematically placing PITX2 in varying orientations relative to the DNA, with helix 3 approximately 50 Å from the DNA. For each of the 20 lowest-energy structures, 5 different orientations of the protein relative to the DNA were selected, yielding 100 starting conformers.

The protein was docked onto the DNA by a 20 ps simulated annealing calculation (T = 600 K, time step = 1 fs) using an altered version of a procedure described

previously for docking of TFIIIA to DNA (*73*). The temperature was increased from 0 to 600 K over the first 4 ps, held at 600 K for 2 ps, and then slowly cooled to 0 K over 14 ps. The weights of the force constants were linearly increased from 0.1 to 1 during the course of the calculation. DNA base-pairing was maintained by incorporating Watson−Crick hydrogen-bonding restraints. These Watson−Crick DNA restraints were implemented as lower and upper bound restraints on base-paired heteroatom−heteroatom (2.7 to 3.1 Å) and heteroatom−proton distances (1.67 to 2.07 Å), and had a final force constant of 50 kcal mol^{-1} $Å^{-2}$. The intramolecular protein and DNA restraints had final force constants of 20 kcal mol^{-1} $Å^{-2}$. Protein and DNA angle restraints had a final force constant of 32 kcal mol^{-1} $Å^{-2}$. Protein−DNA intermolecular restraints had a final force constant of 32 kcal mol^{-1} $Å^{-2}$. Protein restraints were applied to prevent the protein conformation from being altered too much from the structure calculated by CYANA. DNA restraints were applied to prevent fraying of the DNA, and to maintain the structure close to B-form.

Structure Refinement and Analysis. The 20 structures with the lowest AMBER energy values were subjected to restrained energy minimization by the SANDER module of the AMBER 7.0 package (*71*, *72*). The 1994 version of the force field was used. Each conformer was subjected to a conjugate-gradient energy minimization calculation with solvent included.

The evaluation of the structure, i.e., analysis of geometry, stereochemistry, and energy distributions in the models, was performed using the program PROCHECK (*74*). Restraint violations were analyzed using the program AQUA (*75*). Graphics were prepared using MOLMOL (*76*).

RESULTS AND DISCUSSION

Structure Determination. Assignments of the protein backbone and side-chain ^{1}H, ^{13}C, and ^{15}N resonances were obtained from heteronuclear spectra. Restraint data derived for the PITX2 homeodomain−DNA complex are summarized in Table 1. Analysis of ^{15}N and ^{13}C heteronuclear-edited NOESY spectra recorded at various mixing times provided 1259 intramolecular distance restraints comprising 513 intraresidue, 338 sequential, 300 medium-range, and 108 long-range NOE contacts. Torsional restraints for 55 ϕ and 43 ψ angles were obtained from a 3D HNHA experiment, and from using C^{α} chemical shifts (*77*, *78*). Overall, there are 19 restraints per residue, on average, for intramolecular protein NOEs. All of these protein restraints were used for structure calculation with the program CYANA (version 1.0.6) (*69*, *70*). After the final round of structure calculation, the 20 structures with the lowest CYANA target function were used for docking to DNA and energy minimization. The final average CYANA target function for the 20 structures was 2.05 $Å^2$.

A total of 292 distance restraints between protons within the DNA were obtained from $^{13}C/^{15}N$-filtered NOE spectra. A series of 2D $^{13}C(\omega_1)$-edited, $[^{13}C,^{15}N](\omega_2)$-filtered NOESY spectra provided 27 unambiguous intermolecular restraints between the protein and the DNA. These restraints were entered into the program AMBER for docking and energy minimization, as described in Materials and Methods.

Quality of the NMR Structure. The structure of the PITX2−DNA complex was calculated by a restrained molecular dynamics docking and energy minimization procedure starting from the coordinates of the PITX2 protein calculated from CYANA and canonical B-form DNA as described in Materials and Methods. The 20 structures with the lowest total energies were selected for conformer analysis. These structures exhibited mean AMBER energies of −6268 kcal mol^{-1} and mean van der Waals and electrostatic energies of −399 and −3974 kcal mol^{-1}, respectively. The mean AMBER energies given represent the intraprotein interaction energy.

Table 1. NMR Structure Statistics[a]

NMR constraints	
protein	
distance constraints	1259
intraresidue	513
sequential	338
medium-range	300
long-range	108
dihedral constraints	98
phi	55
psi	43
DNA	292
protein−DNA (intermolecular)	27
total	1676
CYANA target function value ($Å^2$)[b]	2.05 ± 0.39
no. of violations	
distance violations (>0.30 Å)	0
dihedral angle violations (>5.0°)	1
AMBER energies (kcal/mol)[c]	
mean AMBER energy	−6268 ± 250
van der Waals	−399 ± 33
electrostatic	−3974 ± 336
Ramachandran plot (%)[d]	
residues in most favored regions	80.1
residues in additional allowed regions	14.8
residues in generously allowed regions	2.3
residues in disallowed regions	2.8
rmsd from the mean structure (Å)	
protein (bb, residues 3−58)	1.38
all heavy atoms (residues 3−58)	1.95
protein (bb, all residues)	1.85
DNA (residues 68−78, 81−91)	1.30
complex (residues 3−58, 68−78, 81−91)	1.81

[a] A total of 30 conformers were calculated, and the 20 structures with the smallest residual CYANA target function values were subjected to docking and energy minimization. [b] The value given for the CYANA target function corresponds to the value before energy minimization (the CYANA target function is not defined after energy minimization, since the conformers no longer have ECEPP standard geometry). [c] The value given represents the intra-protein interaction energy. [d] For residues 3−58.

The superposition of the structures (Figure 2) demonstrates a well-defined tertiary structure for PITX2 bound to DNA. The structures have no distance violations greater than 0.3 Å, and only 1 angle violation greater than 5°. Analysis of Ramachandran plots for the ensemble indicates that the structures generally show favorable backbone conformations within allowed conformational space, with 80.1% of the residues 3−58 within the most favored regions, 14.8% in additionally allowed regions, 2.3% in generously allowed regions, and 2.8% in disallowed regions for the 20 conformers (Table 1). The N- and C-termini are largely disordered. When superimposed, residues 3−58 have an average root-mean-square deviation (rmsd) from the mean structure of 1.38 Å for backbone (N, C^{α}, C′, and O), 1.95 Å for all heavy

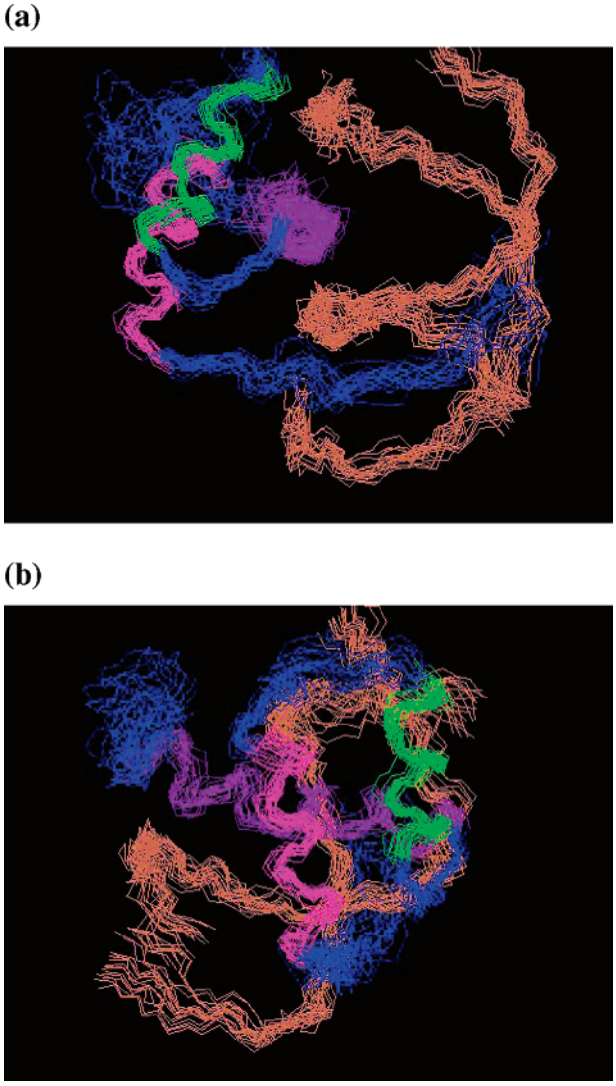

FIGURE 2: Ensemble of structures of the PITX2 homeodomain−DNA complex. (a) Ensemble of 20 structures showing the protein backbone N, C^{α}, and C′ atoms and the DNA backbone. Helix 1 is colored pink, helix 2 green, helix 3 purple, and the DNA strands are coral. Superimposition was performed using backbone atoms from protein and DNA. (b) Alternate view of the structure, rotated by approximately 90°.

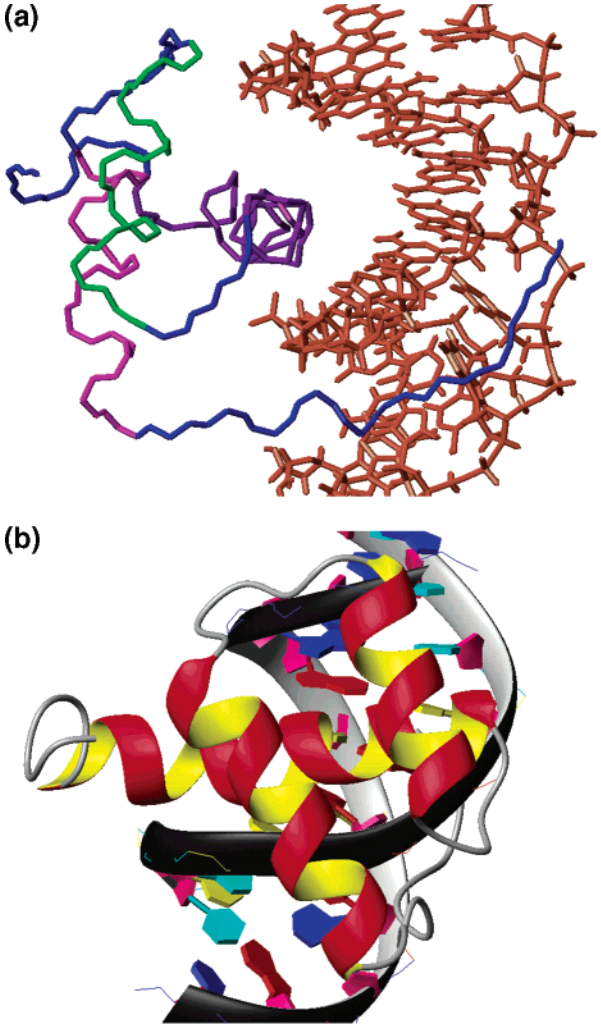

FIGURE 3: Structure of the PITX2 homeodomain−DNA complex. (a) Mean structure of the homeodomain. Helix 1 is colored pink, helix 2 green, and helix 3 purple. The DNA strands are colored coral. (b) Ribbon diagram of the mean structure of the PITX2 homeodomain−DNA complex.

atoms, and 1.85 Å for the backbone when all residues are included. The global rmsd for all DNA heavy atoms (nucleotides 68−78, 81−91) is 1.30 Å. The rmsd for the entire complex (residues 3−58; nucleotides 68−78, 81−91) is 1.81 Å.

Tertiary Structure of the PITX2 Homeodomain−DNA Complex. The overall tertiary structure of the PITX2 homeodomain is similar to other homeodomains, supporting previous findings that this tertiary structure is well conserved among homeodomains (*1*, *3*, *79*, *80*). The tertiary structure of the PITX2 homeodomain is composed of three α helices (Figure 3). Helix 1 (residues 10−20) is followed by a loop region, and then helix 2 (residues 28−37) runs antiparallel to helix 1. Helix 2 and helix 3 (residues 42−58) form a helix−turn−helix motif. Helix 3 is approximately perpendicular to helices 1 and 2, and fits into the major groove of the DNA. The N-terminus of the homeodomain makes contacts within the minor groove of the DNA.

The helices of the PITX2 homeodomain are held together by a core of eight tightly packed hydrophobic amino acids (F8, L13, L16, F20, L40, V45, W48, and F49). These amino acids are either invariant (W48 and F49) or highly conserved in all homeodomains (*3*, *81*, *82*). In a threading analysis performed previously for the PITX2 homeodomain (*83*), it was hypothesized that the tertiary structure of the PITX2 homeodomain would be similar to other homeodomains, mainly because many of these hydrophobic amino acids that are present in other homeodomains are also present in the PITX2 homeodomain. The threading analysis threaded the PITX2 homeodomain sequence to the Engrailed homeodomain structure, so the overall tertiary structure ended up being very close to that of Engrailed. The threading analysis did not provide a PDB file that we could analyze in detail, and in any case is not necessarily indicative of the true molecular structure. For this threading analysis, the focus was the role of Rieger mutations in causing disease, and there was no discussion of the role of K50 in determining the DNA-binding affinity and specificity of the homeodomain, which is something best addressed via an experimentally determined structure rather than a threading model. This study also did not analyze the K50 Rieger mutants, so there is no indication what the structure of this side chain was in their analysis. In the absence of an experimentally determined

structure, the threading model was most useful for visualizing some of the intramolecular interactions that stabilize the tertiary structure, and the predicted interactions are consistent with our experimental data. While we cannot compare our PITX2 tertiary structure directly to that of the threaded structure, we can compare it to EnQ50K and other homeodomain structures. In our experimentally determined structure, the first helix is closer to the second helix when measured from the backbone nitrogen of L16 to the backbone nitrogen of I34 and compared to the EnQ50K, Antennapedia, wild-type Engrailed, Fushi tarazu, vnd/NK-2, and MATα2 homeodomains (*7*, *84*−*88*). As far as this distance is concerned, PITX2 is an outlier compared to the other six homeodomains. This distance is a range of 9.60−10.70 Å for Antennapedia conformers, 9.43 Å for the crystal EnQ50K structure, 9.54 Å for wild-type Engrailed, 9.30−11.10 Å for Fushi tarazu, 8.67 Å for vnd/NK-2, and 10.9 Å for MATα2. However, for PITX2 this distance range over the 20 conformers is only 7.55−8.58 Å, which is an average of 1.8 Å closer. In view of the rmsd for the PITX2 protein backbone atoms (residues 3−58) of 1.38, this result is still significant, when compared with the ranges of distances seen in the structures of the other homeodomains. This difference is especially significant when considering that the rmsd for helices 1 and 2 alone is only 0.78 Å. The range for the distance between L16 and I34 for all of the other homeodomains together is 8.67−11.10 Å, and the PITX2 distance range is completely outside of this.

In addition to the narrower distance between the first and second helices in the PITX2 structure, there are several other differences between the PITX2 and EnQ50K structures. In particular, the third helix of PITX2 is positioned about 0.5 Å lower (closer to the N-terminus of helix 1 and C-terminus of helix 2) than in EnQ50K (Figure 4). This difference in orientation of the three helices causes slightly different contacts to be made between the first and third helices, and may provide a partial explanation for the decreased stability of this homeodomain. Unlike other homeodomains that are stable in the free form (*32*, *86*, *89*−*92*), the PITX2 homeodomain is unstable in the absence of DNA in that it irreversibly aggregates at micromolar concentrations, which suggests a possible lack of stable tertiary structure in the free form. This may be due to slightly different hydrophobic interactions within the core of the protein, and the absence of other stabilizing interactions such as the salt bridge linking residues 19 and 30, which can be present in most homeodomains (*93*) but is not possible in PITX2. One difference seen here is that F49, which is nearly invariant among homeodomains, points slightly upward toward the loop region of the homeodomain, instead of pointing toward the interior of the protein. The orientation of the first helix in relation to the third would cause a steric clash with F49 if it were in an orientation similar to that of other homeodomains. While there is still an interaction involving F49 and F20 within the hydrophobic core of the PITX2 homeodomain, the orientations of the side chains themselves are different. This differing orientation may lessen the strength of the interaction between the first and third helices, which may affect the stability of the protein in the absence of DNA. This difference in orientation may be due to any number of differing residues between the two homeodomains (see Supporting Information). One possibility is a proline residue

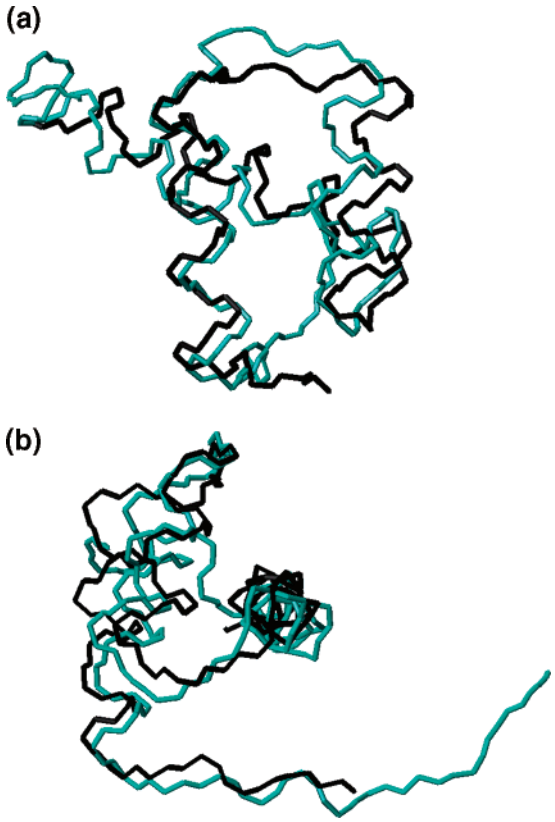

FIGURE 4: Overlay of PITX2 homeodomain and EnQ50K homeodomain structures. Cyan corresponds the structure of the PITX2 homeodomain, and black corresponds to the structure of the Engrailed mutant homeodomain. (a) Helices 1 and 2 are approximately 1.8 Å closer to each other in PITX2 than in other homeodomains. (b) Alternate view, rotated by approximately 90°. Helix 3 is about 0.5 Å lower in PITX2 than in EnQ50K.

that is found in the loop region between helices 1 and 2 in PITX2, but is not present in Engrailed or Fushi tarazu.

In the PITX2 structure, the N- and C-terminal segments −2 to 2 and 60 to 68 (Figure 1) appear disordered (Figure 2), which is to be expected on the basis of a lack of medium-range and long-range constraints for these residues. Analysis of ^{15}N relaxation data (unpublished) indicates that residues −2 to 2 and 59 to 68 are more mobile in solution, explaining the observed disorder and lack of restraint information for these regions.

Our study also reveals structural information about the DNA when it is bound to the protein. Distance restraints obtained from the experiments described above for assigning the DNA were entered into AMBER during the docking procedure. Visual inspection of the structure of the PITX2 homeodomain−DNA complex indicates that there is a slight widening of the minor groove of the DNA compared to B-form DNA, and a concomitant narrowing of the major groove. Previous structures of protein−DNA complexes have indicated that changes in DNA structure are possible upon protein binding (*94*). A more thorough, quantitative analysis of the DNA structure when PITX2 is bound will not be possible until a high-resolution structure of the DNA is determined, using isotopically labeled DNA (*95*).

Protein−*DNA Recognition*. Analysis of the filtered NOESY experiments produced 27 unambiguous distance restraints between the protein and the DNA (see Supporting Information for a list). These include contacts that have been seen

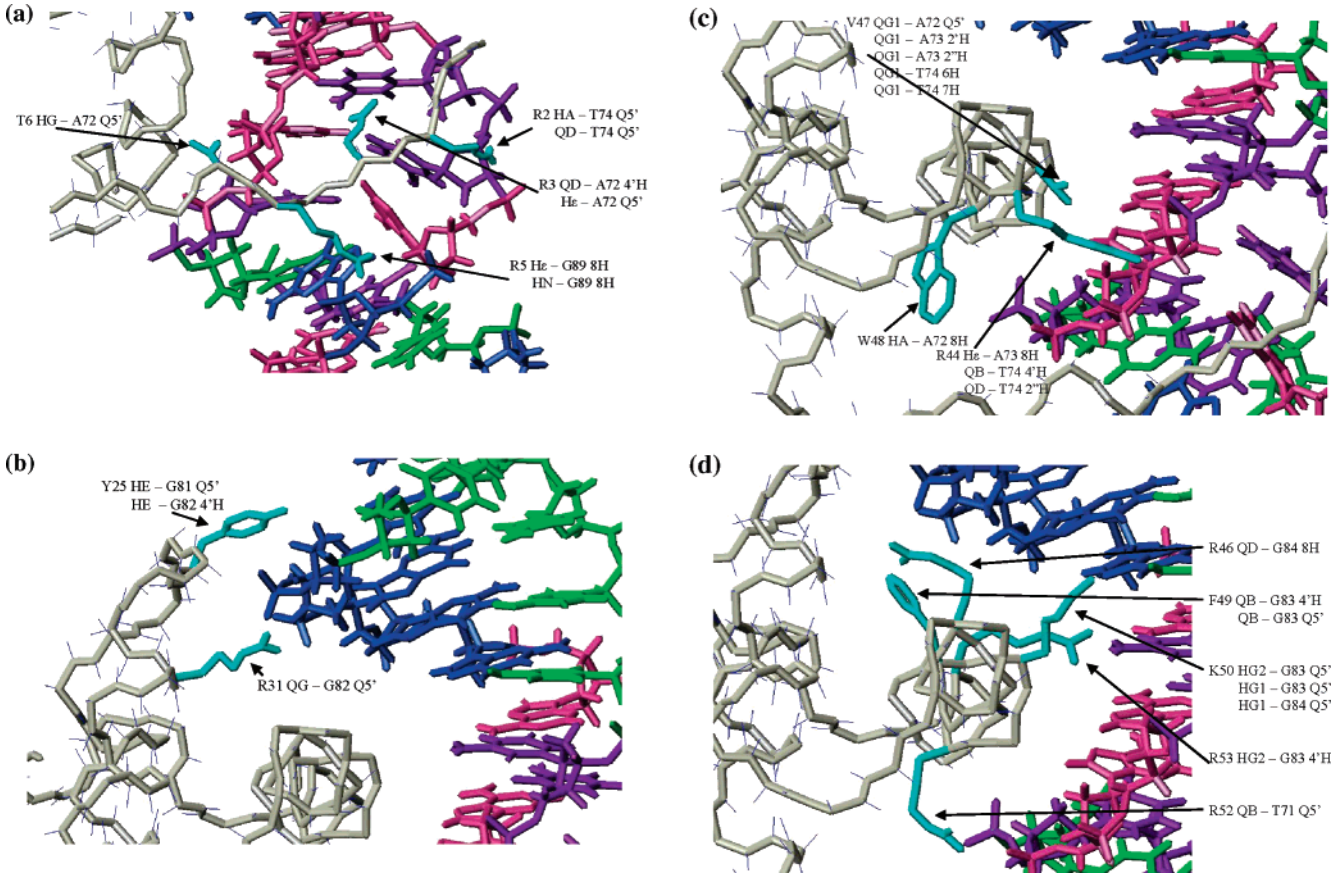

FIGURE 5: Detailed view of the protein−DNA interface and protein−DNA contacts. The backbone of the protein is shown in beige, with the C^{α}−H^{α} bonds shown with small blue lines. Side chains of the protein are illustrated in cyan. On the DNA, blue corresponds to guanine residues, green to cytosine, pink to adenine, and purple to thymine. (a) View of the protein−DNA NOE contacts between the N-terminus of the PITX2 homeodomain and the minor groove of the DNA. (b) View of the protein−DNA NOEs between Y25, R31, and the DNA. (c, d) View of protein−DNA NOE contacts between residues in the third helix and the major groove of the DNA.

in other biochemical and structural studies of homeodomains. Many of the residues that interact with the DNA are arginines, including R3 and R5 at the N-terminus, R31 in the second helix, and R46, R52, and R53 in the third helix (Figure 5). Other residues that were found to make DNA contacts are Y25 and F49. A number of NOESY peaks were also seen between K50 and the DNA, and these contacts are discussed further below.

A detailed picture of the protein−DNA interface is shown in Figure 5. This figure illustrates the orientations of some of the side chains that are important in DNA binding, particularly within the third helix (the specific atom contacts are indicated in Figure 5). Figure 1b outlines the numbering of the DNA used in the following discussion. Figure 5a illustrates the protein−DNA NOE contacts seen within the N-terminal arm. NOE contacts were seen in the minor groove between R2, R3, and R5 and DNA residues A72, T74, and G89. Although the NOESY-derived distance constraints indicate contact between residues R3 and R5 and the minor groove, ^{15}N relaxation data (unpublished) indicates that this region of the N-terminus does retain some degree of mobility; similar results were reported for the Even-skipped homeodomain, on the basis of refined atomic B factors (*30*). Broad line widths were observed for the backbone NH resonances of His7 and Phe8, which are indicative of slow time scale motions in this region of the homeodomain and could possibly render undetectable possible NOEs from these residues to the DNA.

In the second helix, R31 has a NOE contact from the H^{γ} position to G82 Q5′ (Q refers to a pseudoatom representation, to indicate that stereospecific assignment of the 5′ and 5″ protons was not made), as can be seen in Figure 5b. HBPLUS analysis (*96*) indicates that R31 is making a hydrogen bond contact with the phosphate backbone of this nucleotide. In the loop between helices 1 and 2, Y25 H^{ϵ} is making NOE contacts with G81 Q5′ and G82 4′H. In the third helix, V47 $Q^{\gamma 1}$ is making conserved NOE contacts to A72 Q5′, A73 2′H and 2″H, and T74 6H. Residue W48 has a NOE contact between H^{α} and A72 8H. R44 H^{ϵ} is making contact with DNA proton A73 8H, R44 Q^{β} with T74 4′H, and R44 Q^{δ} with T74 2″H. HBPLUS analysis indicates that R44 is making a backbone hydrogen bond contact to the phosphate of T74. Residues 44, 47, and 48 are illustrated in Figure 5c. In the third helix, R46 and R52 appear to be making conserved contacts with the DNA backbone. R46 extends upward, and R52 extends downward to make these contacts (Figure 5d). R46 Q^{δ} has a NOE contact with G84 8H. R52 Q^{β} has a NOE contact with T71 Q5′. R53 Q^{γ} makes a NOE contact with G83 4′H. All of these NOEs could be due to the close proximity of the atoms while the side chains form hydrogen bonds with backbone phosphate groups. NOEs are also seen between F49 Q^{β} and G83 4′H and Q5′. K50 will be discussed further below, but as can be seen in Figure 5d, there are NOE contacts between the K50 side chain and atoms from G83 and G84.

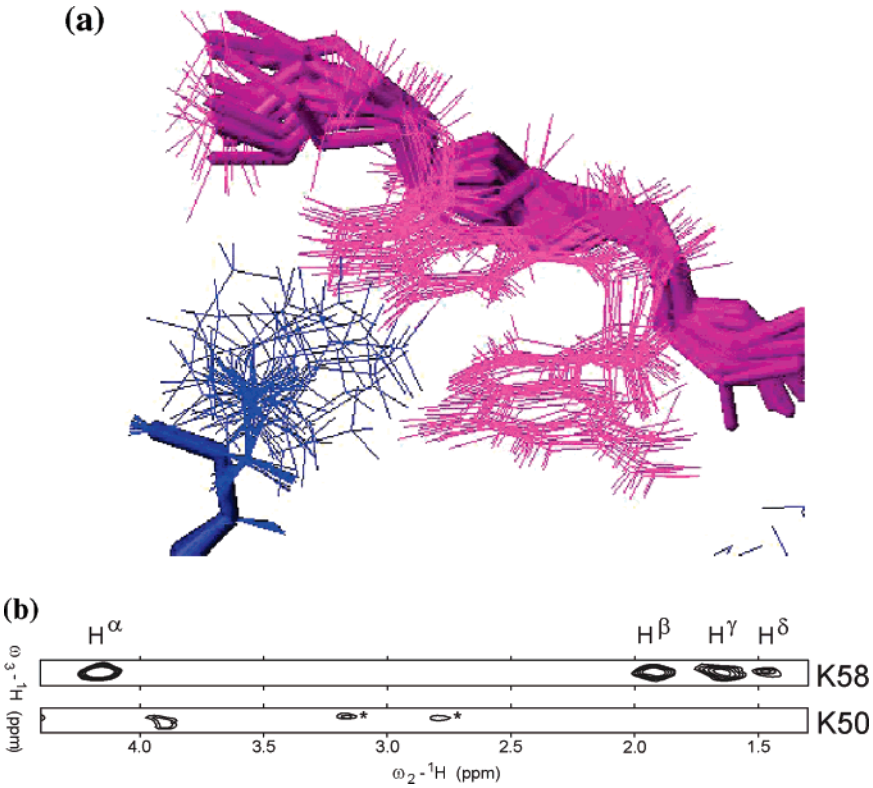

FIGURE 6: (a) View of the 20 conformers, with only the K50, G83, and G84 backbone and side-chain atoms shown to illustrate the extent of disorder of the K50 side chain, implying possible mobility of this side chain in interacting with the DNA. K50 atoms are shown in blue, and G83 and G84 atoms are shown in pink. Backbone atoms are bolder than side-chain atoms. (b) Strips from an H(CCO)NH-TOCSY spectrum showing proton resonances for the side chains of K58 and K50. Line broadening of resonances in the K50 side chain, leading to the weak or missing signals, is indicative of possible motion of this side chain. The asterisks in the K50 slice indicate peaks breaking through from an adjacent ^{15}N plane. The sole K50 peak is for the H^{α} proton resonance.

Other residues that were observed in the calculated structures to be in close contact with the DNA, but without NOEs being seen in the NMR data, are N51, K55, R57, and K58. N51 is nearly invariant among homeodomains (*82*) and is found herein to make the same highly conserved interaction within the major groove with base A73. This residue has been shown in crystal structures to form a pair of hydrogen bonds with this adenine at the N7 and N6 positions, while NMR studies have indicated possible rapidly interchanging conformations (*32*, *97*). NMR studies have shown this close interaction, but no NOEs are seen, possibly due to line-broadening effects (*97*). While no NOEs are seen between K55, R57, or K58 and the DNA, HBPLUS analysis of the complex indicates that there are possible interactions present. K55 may be forming a hydrogen bond with the phosphate of T71. R57 may be contacting the phosphate of G84. K58 may be contacting the phosphate of C70. Due to the usual sensitivity limitations in the edited/filtered NMR experiments employed to identify intermolecular NOEs, it is quite likely that a number of anticipated NOEs fall at or below the threshold for detection.

The Role of Lysine at Position 50. No previous structures have been described for any native K50 class homeodomains. However, the X-ray crystal structure of the Q50K mutant of the Engrailed homeodomain bound to DNA has been reported (*7*), and the side chain of K50 was found to project into the major groove of the DNA, making hydrogen bond contacts with the O6 and N7 atoms of the guanines at base pairs 5 and 6 of the complementary strand of the TAATCC binding site. Our structure of the PITX2 homeodomain marks the first experimentally determined structure of a native K50 class homeodomain, and is important for validating results seen in the studies of non-native proteins. When binding to the consensus site, the position of K50 is very similar to that seen in the EnQ50K structure, with the side chain of K50 extending outward and making contacts with the two guanines adjacent to the TAAT core sequence on the antisense strand (Figure 5d). NOEs are observed between the K50 Q^{γ} and the Q5′ protons of G83 and G84. The N^{ζ} of the K50 side chain is likely making hydrogen bond contacts to the O6 and N7 atoms of G83 and G84, according to analysis by HBPLUS (*96*).

NMR spectroscopy allows one to obtain information about the mobility of the protein backbone and side chains. A key finding in the present study was that the side chain of K50 potentially mediates recognition by fluctuating between multiple conformations. The conformational heterogeneity can be seen in Figure 6a. This preliminary evidence is based on averaging of NOEs and broadening of resonances for this residue. The averaging of NOEs was dealt with as ambiguous distance constraints within the structure calculation in CYANA, and these constraints were satisfied in all structures of the family. When results from an H(CCO)NH-TOCSY experiment are compared between the K50 and K58 side chains (Figure 6b), peaks are easily seen for the K58 side-chain resonances, but only the H^{α} resonance is seen for the K50 side chain. The extra peaks in the K50 strip of Figure 6b are from another residue on an adjacent nitrogen plane and are strong enough to show up as residual peaks on this plane. The broadening of resonances for this side chain made

it difficult to assign using typical heteronuclear-edited NMR spectra. Instead, assignments were made using NOESY spectra and eliminating assignments from nearby residues, until only K50 resonances were left. In principle, it is possible that the line broadening of K50 side-chain resonances could be caused by ring current effects from aromatic bases in the DNA, or by mobility of other nearby protons in the DNA binding site. However, no anomalous line broadening was observed for DNA proton resonances in the vicinity of the K50 side chain. In addition, results similar to those reported here have been seen in other DNA-binding proteins in which side-chain mobility appears to cause line broadening of resonances (vide infra) (*32*, *97*−*99*). These results, in combination with the multiple conformations observed for K50 in EnQ50K, provide compelling evidence that the side chain of K50 is mobile. Preliminary ^{15}N relaxation measurements of the homeodomain backbone dynamics (unpublished) did not show anything unusual in the region of K50. Some degree of side-chain mobility at the protein−DNA interface would be expected to confer an entropic advantage for binding to the DNA. It has been estimated previously that the entropic cost of keeping a lysine side chain static during binding is 3 kcal mol^{-1} (*100*). This possible entropic component cannot be assessed until a detailed thermodynamic study is performed for this complex. This hypothesis of K50 side-chain mobility will be explored further in the future, but for now, it is complementary to the data for the EnQ50K mutant (*7*). The crystal structure indicates that there are two alternate conformations for the K50 side chain, one in which the side chain points to base pairs 5 and 6, and one in which the side chain is oriented slightly more toward base pair 5. It must be pointed out that this X-ray structure was solved at cryogenic temperatures, so there is the possibility that there is a freezing out of a subset of conformational populations. It is possible that these results indicate two static, nearly isoenergetic conformations for this side chain of EnQ50K, rather than a dynamic fluctuation between two conformations. The *B*-factors in this case provide no evidence for distinguishing between these possibilities. The *B*-factors are low for the side chain of K50 in the 1.9 Å crystal structure of EnQ50K, varying over the range 20.8 to 23.6, which are the lowest values in the protein, aside from the aromatic ring of F49. *B*-factors of about 20 indicate uncertainties of about 0.5 Å. Typically, *B*-factors of 60 or greater in high-resolution crystal structures indicate possible mobility of a side chain. So, according to the crystal results, the side-chain position of K50 is well defined in the crystal, in contrast to the possible mobility of the K50 side chain seen in our results. The true nature of the side-chain conformation and dynamics may involve a combination of the states revealed by the two different experimental approaches, so that the K50 side chain has two predominant conformations, and fluctuates between these alternatives.

Although a more detailed characterization of the side-chain dynamics in the PITX2−DNA interface must await data from experimental NMR relaxation measurements and molecular dynamics simulations, substantial support for our observation of flexibility in the K50 side chain already exists from studies of related systems. Significant broadening of side-chain resonances at the protein−DNA interface was observed in studies of homeodomain−DNA complexes of Antennapedia (*97*) and NK-2 (*32*). Moreover, flexibility in lysine side chains appears to be a significant feature of various modes of protein−DNA interactions. Foster and co-workers (*98*) have reported clear indications of substantial, conformational fluctuations in lysine side chains in the interface of the zinc-finger protein TFIIIA with its DNA binding site, including the observation of broadened resonances and multiple NOE contacts that strongly suggest rapid conformational averaging. Significant line-broadening effects were also reported for a lysine side chain in NMR studies of the telomeric DNA complex of trf1 (*99*). In addition to NMR studies, molecular dynamics simulations of wild-type (*31*) and a Q50K mutant (*33*) of the Antennapedia homeodomain bound to DNA provide further evidence in support of a dynamic homeodomain−DNA interface. For example, the Q50K simulations indicated that the side chain of K50 exhibited very pronounced mobility, with several arrangements of the lysine side-chain torsion angles allowing for frequent contacts, both hydrogen-bonding and hydrophobic interactions, with base pairs 5 and 6 in the TAATCC binding site. In this case, the lysine in the Q50K mutant provides both entropic and enthalpic contributions to protein−DNA affinity. A general observation arising from the known structures of homeodomain−DNA complexes is that the region of position 50 is not in intimate contact with the bases of the major groove. Such a relatively unrestrained arrangement allows for relatively long-range contacts to be formed in multiple, possibly isoenergetic ways.

Previous studies have shown that the lysine at position 50 is critical for its binding to the TAATCC DNA binding site (*82*). In contrast, homeodomains with a glutamine at position 50 bind to TAATGG sites with a higher affinity. The glutamine at position 50 appears to have a more modest role. When this residue is mutated to an alanine, the Q50A mutant has an affinity and specificity very similar to those of the wild-type protein, but when mutated to a lysine, the specificity changes (*5*). These studies, along with the current results, indicate that the interaction between K50 and the two guanines at positions 5 and 6 is vital to the affinity and specificity of the protein. The current model for specific homeodomain−DNA interactions consists of a fluctuating network of hydrogen bonds formed between polar groups of the protein and the DNA, and the interfacial water (*31*). These interactions are further complemented by hydrophobic contacts. The possible fluctuating hydrogen-bonding interactions between K50 and the DNA and subsequent strict specificity of this class of homeodomains are consistent with this model. Investigation of side-chain−base interactions has shown that lysine−guanine interactions are very common (*101*). K50 homeodomains may have such a strong specificity for the TAATCC site because the orientation of the lysine is in an ideal position for the charged group to make hydrogen-bonding contacts with the two guanines. In contrast, these hydrogen bonds cannot be made with cytosines, which are in these positions for the Q50 binding site TAATGG (*101*). The N7 of guanine is the most electronegative region of the major groove (*102*), and the favorable interactions that the lysine can make with both guanines in a mobile model may determine why K50 homeodomains are so specific for the TAATCC binding site, rather than other binding sites.

Analysis of Residues Mutated in Rieger Syndrome. There have been 9 mutations found in the PITX2 homeodomain

Table 2. PITX2 Homeodomain Mutations (*35*, *45*, *48*, *103*−*109*)

mutation	disease	properties
L16Q	Rieger syndrome	unstable, no activation, no DNA binding
T30P	Rieger syndrome	no activation, only binds consensus
R31H	iridogoniodysgenesis	reduced activation, only binds consensus site
V45L	Rieger syndrome	<10-fold reduction in DNA binding, 200% increase in activation
R46W	iris hypoplasia	reduced binding to nonconsensus site, reduced activation
K50E	Rieger syndrome	no DNA binding or activation, dominant negative
K50Q	Rieger syndrome	not known
R52C	Rieger syndrome	not known
R53P	Rieger syndrome	no nonconsensus binding, no activation, dominant negative

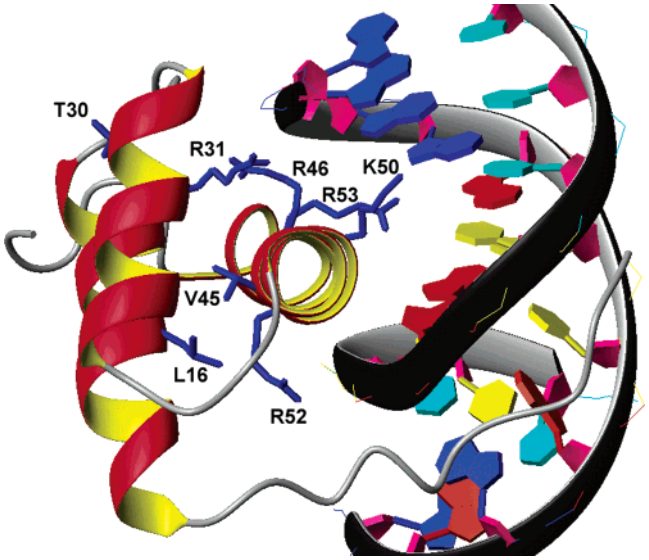

FIGURE 7: Ribbon diagram of the PITX2 homeodomain−DNA complex showing the positions of the side chains for the residues known to be mutated in Rieger syndrome and related disorders.

in Rieger syndrome and related disorders (*35*, *45*, *48*, *103*−*109*). These mutations, along with their known biochemical effects, are listed in Table 2. The consequences of these mutations vary. Some mutations cause a total lack of DNA binding, while others can still bind DNA, albeit with a decreased affinity. These consequences are directly reflected in the severity of the disease. A model of the PITX2 homeodomain structure was created previously by threading analysis, which allowed predictions to be made regarding the role of Rieger syndrome mutations in PITX2 dysfunction, although it is not necessarily an indication of the true molecular structure (*83*).

The orientations of the side chains altered in Rieger syndrome patients are shown in Figure 7. Analysis of these orientations provides insights into the role of each side chain, and how mutations in these positions could alter the structure and function of the protein. Future studies will focus on analyzing the mutant proteins by NMR spectroscopy. The side chain of highly conserved L16 points toward the interior hydrophobic core of the protein, and is probably involved in stabilizing both the formation of this core and the overall tertiary structure of the protein; the L16Q mutation would therefore be expected to destabilize or disrupt this hydrophobic core. The side chain of T30 extends outward from the second helix, away from the DNA, so it does not appear to play a role in DNA recognition. Biochemical studies have shown that this mutant can still bind consensus DNA, but no longer activates transcription of a reporter gene (*25*). This residue may perform an activation function by interacting with other proteins, which could easily be disrupted by the effects of the proline mutation. An interesting observation is that, in many homeodomains, residue 30 is involved in a salt bridge to residue 19, whereas this is not possible for PITX2. The side chain of R31, as described above, appears to contact the DNA backbone phosphate of G82. Therefore mutating this residue, even to another positively charged residue, may disrupt this interaction with the DNA and may disrupt a possible salt bridge with E42. The histidine side chain at this position in the mutant may not have favorable steric interactions with the DNA. The side chain of V45 points toward the interior of the protein from the third helix. Like L16, this side chain appears to be involved in formation of the hydrophobic core of the protein. Unlike the L16Q mutant, the V45L mutant has the unusual characteristic of having a greatly heightened activation function, while having a reduced DNA-binding ability. It is possible that this mutant affects the protein in a way that alters these two functions separately, with a different fold of the protein that allows for a more efficient interaction with other proteins. For example, altered interactions of the PITX2 homeodomain with the C-terminal tail of the full-length PITX2 protein could have differential effects on DNA binding and activation (*110*). The DNA-binding functions of R46, K50, R52, and R53 were discussed in detail above. Mutating these residues would disrupt many favorable interactions with the DNA, and biochemical studies have indicated that these mutations interfere with DNA binding. Overall, these results are similar to the threading analysis, but provide a more direct and detailed understanding of the roles of these residues.

Many of the residues in the PITX2 homeodomain found to be altered in Rieger syndrome are involved in contacting the DNA. Other residues are involved in forming the hydrophobic core of the protein, which stabilizes the global fold. The analysis of mutations causing structural changes could be very relevant for the understanding and prediction of dysfunctions caused by mutations in homeodomains, as several homeodomains are known to be involved in various diseases (*111*−*115*).

CONCLUDING REMARKS

The structure previously determined for the Engrailed Q50K mutant (*7*) provided some interesting insights into the possible role of lysine at position 50. The presence of hydrogen bonds between position 50 and the DNA had not been seen previously. But many questions remained unanswered concerning the role of lysine in a native K50 homeodomain. For example, the Engrailed mutant has a dissociation constant of 0.0088 nM (*7*), representing an unusually high affinity for homeodomain−DNA interactions. Previous studies have indicated that proteins with excessively high affinities for DNA or RNA can cause functional defects (*116*, *117*). The unusually high affinity of EnQ50K for DNA suggests that it may have properties that make it different from natural K50 homeodomains. Unlike the Engrailed mutant, the native K50 class homeodomains PITX2 and Bicoid have properties that make them unstable in free forms,

and have affinities within the normal nanomolar range (*25*, *26*, Supporting Information). When DNA is not present, these proteins will irreversibly aggregate and precipitate out of solution at micromolar concentrations. These differences in biochemical properties between the mutant and natural K50 proteins suggest the importance of understanding the structural properties of lysine at position 50 in the context of a native K50 class protein.

But the question still remains as to what causes these differences. The authors of the EnQ50K structure found that the mutant bound to DNA more tightly and specifically than did the native protein (*7*). They hypothesized that this was due to very specific hydrogen bonds between the K50 side chain and the guanines at positions 5 and 6 on the antisense strand. In our study, we found that the native K50 homeodomain PITX2 has a slightly different tertiary structure, with helix 1 being closer to helix 2 than in other homeodomains, including the EnQ50K mutant. Helix 3 is angled about 0.5 Å closer to the N-terminus of helix 1 and C-terminus of helix 2 than EnQ50K. This appears to cause a difference in the way that helix 1 and helix 3 can interact, and previous studies have shown that this interaction between the helices stabilizes the global fold of the homeodomain (*3*, *81*, *82*). Another Q50K mutant, this time of Fushi tarazu, is unable to bind nonconsensus DNA sites that PITX2 and Bicoid are able to recognize (*21*). It is currently unknown whether the Engrailed mutant can bind nonconsensus sites. These differences in affinity and specificity may involve any of the differing residues between these homeodomains. Positions 50 and 54 have been shown to be involved in recognizing nonconsensus DNA sites (*22*), and it is possible that other residues are also involved. Within the third helix, position 52 of Engrailed is a lysine. In PITX2, Bicoid, and Fushi tarazu, this residue is an arginine. We do not know whether having lysine residues at both positions 50 and 52 could contribute to the unnaturally tight binding of EnQ50K, but this is a possibility.

The current study of the solution structure of the PITX2 homeodomain reveals possible fluctuating interactions between the K50 side chain and the DNA. It is possible that this mobile side chain may allow the protein to sample multiple DNA binding sites, and enable binding to the nonconsensus sites, though at a slightly lower affinity. It will be interesting in the future to determine if other natural K50 class proteins share similar properties with PITX2. Future studies will focus on analyzing Rieger mutants of the PITX2 homeodomain, and analyzing the structural features of this protein when bound to nonconsensus DNA binding sites. This will allow a greater understanding of the roles of specific residues in consensus and nonconsensus DNA binding, and a greater understanding of how proteins can recognize multiple DNA sites to activate transcription of genes.

ACKNOWLEDGMENT

We would like to especially thank Dr. Jeffery C. Murray and Dr. Elena V. Semina from the University of Iowa for providing the original *pitx2* clone. We would also like to thank Dr. Jack Howarth for computer support and maintaining the NMR facility at the University of Cincinnati.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was originally published 4/26/05 with an incorrect entry in Table 1. The corrected version was also published 4/26/05.

SUPPORTING INFORMATION AVAILABLE

Sequence alignment of homeodomains (Table S1), binding affinity of the PITX2 homeodomain (Figure S1), HSQC with the residues labeled (Figure S2), chemical shift assignments for the PITX2 homeodomain and its DNA binding site (Tables S2-S7), protein−DNA contacts (Table S8), and Ramachandran plot for the PITX2 homeodomain (Figure S3). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Gehring, W. J., Muller, M., Affolter, M., Percival-Smith, A., Billeter, M., Qian, Y. Q., Otting, G., and Wüthrich, K. (1990) The structure of the homeodomain and its functional implications, *Trends Genet. 6*, 323−329.
2. Dave, V., Zhao, C., Yang, F., Tung, C., and Ma, J. (2000) Reprogrammable recognition codes in Bicoid homeodomain-DNA interaction, *Mol. Cell. Biol. 20*, 7673−7684.
3. Gehring, W. J., Affolter, M., and Burglin, T. (1994) Homeodomain Proteins, *Annu. Rev. Biochem. 63*, 487−526.
4. Treisman, J., Harris, E., Wilson, D., and Desplan, C. (1992) The Homeodomain: A New Face for the Helix-Turn-Helix?, *BioEssays 14*, 145−150.
5. Wolberger, C. (1993) Transcription factor structure and DNA binding, *Curr. Opin. Struct. Biol. 3*, 3−10.
6. 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.
7. Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T., and Pabo, C. O. (1997) Engrailed (Gln50→Lys) homeodomain-DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity, *Structure 5*, 1047−1054.
8. Hanes, S. D., and Brent, R. (1989) DNA Specificity of the Bicoid activator protein is determined by homeodomain recognition helix residue 9, *Cell 57*, 1275−1283.
9. 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.
10. Fraenkel, E., Rould, M. A., Chambers, K. A., and Pabo, C. O. (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.
11. Grant, R. A., Rould, M. A., Klemm, J. D., and Pabo, C. O. (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.
12. Percival-Smith, A., Muller, M., Affolter, M., and Gehring, W. J. (1990) The Interaction with DNA of Wild-Type and Mutant *Fushi tarazu* Homeodomains, *EMBO J. 9*, 3967−3974.
13. Tron, A. E., Bertoncini, C. W., Palena, C. M., Chan, R. L., and Gonzalez, D. H. (2001) Combinatorial Interactions of Two Amino Acids with a Single Base Pair Define Target Site Specificity in Plant Dimeric Homeodomain Proteins, *Nucleic Acids Res. 29*, 4866−4872.
14. Pomerantz, J. L., and Sharp, P. A. (1994) Homeodomain determinants of major groove recognition, *Biochemistry 33*, 10851−10858.
15. Banerjee-Basu, S., Moreland, T., Hsu, B. J., Trout, K. L., and Baxevanis, A. D. (2003) The homeodomain resource: 2003 update, *Nucleic Acids Res. 31*, 304−306.
16. Otting, G., Qian, Y. Q., Muller, M., Affolter, M., Gehring, W., and Wüthrich, K. (1988) Secondary structure determination for the Antennapedia homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif, *EMBO 7*, 4305−4309.

17. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) Crystal Structure of the MATa1/MATα2 Homeodomain Heterodimer Bound to DNA, *Science 270*, 262−269.
18. Cox, M., van Tilborg, P. J., de Laat, W., Boelens, R., van Leeuwen, H. C., van der Vliet, P. C., and Kaptein, R. (1995) Solution structure of the Oct-1 POU homeodomain determined by NMR and restrained molecular dynamics, *J. Biomol. NMR 6*, 23−32.
19. Piper, D. E., Batchelor, A. H., Chang, C., 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.
20. Tejada, M. L., Jia, Z., May, D., and Deeley, R. G. (1999) Determinants of the DNA-binding specificity of the Avian homeodomain protein, AKR, *DNA Cell Biol. 18*, 791−804.
21. Zhao, C., Dave, V., Yang, F., Scarborough, T., and Ma, J. (2000) Target Selectivity of Bicoid is Dependent on Nonconsensus Site Recognition and Protein-Protein Interaction, *Mol. Cell. Biol. 20*, 8112−8123.
22. Pellizzari, L., Tell, G., Fabbro, D., Pucillo, C., and Damante, G. (1997) Functional interference between contacting amino acids of homeodomains, *FEBS Lett. 407*, 320−324.
23. Clarke, N. D. (1995) Covariation of residues in the homeodomain sequence family, *Protein Sci. 4*, 2269−2278.
24. Simon, M. D., Sato, K., Weiss, G. A., and Shokat, K. M. (2004) A phage display selection of engrailed homeodomain mutants and the importance of residue Q50, *Nucleic Acids Res. 32*, 3623−3631.
25. Amendt, B. A., Sutherland, L. B., Semina, E. V., and Russo, A. F. (1998) The Molecular Basis of Rieger Syndrome, *J. Biol. Chem. 273*, 20066−20072.
26. Ma, X., Yuan, D., Diepold, K., Scarborough, T., and Ma, J. (1996) The *Drosophila* morphogenetic protein Bicoid binds DNA cooperatively, *Development 122*, 1195−1206.
27. Yuan, D., Ma, X., and Ma, J. (1999) Recognition of Multiple Patterns of DNA Sites by Drosophila Homeodomain Protein Bicoid, *J. Biochem. 125*, 809−817.
28. Hjalt, T. A., Amendt, B. A., and Murray, J. C. (2001) PITX2 Regulates Procollagen Lysyl Hydroxylase (PLOD) Gene Expression: Implications for the Pathology of Rieger Syndrome, *J. Cell Biol. 152*, 545−552.
29. Espinoza, H. M., Cox, C. J., Semina, E. V., and Amendt, B. A. (2002) A molecular basis for differential developmental anomalies in Axenfeld-Rieger syndrome, *Hum. Mol. Genet. 11*, 743−753.
30. Hirsch, J. A., and Aggarwal, A. K. (1995) Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity, *EMBO J. 14*, 6280−6291.
31. Billeter, M., Guntert, P., Luginbuhl, P., and Wüthrich, K. (1996) Hydration and DNA Recognition by Homeodomains, *Cell 85*, 1057−1065.
32. Tsao, D. H., Gruschus, J. M., Wang, L. H., Nirenburg, M., and Ferretti, J. A. (1994) Elongation of helix III of the NK-2 homeodomain upon binding to DNA: a secondary structure study by NMR, *Biochemistry 33*, 15053−15060.
33. Gutmanas, A., and Billeter, M. (2004) Specific DNA recognition by the *Antp* homeodomain: MD simulations of specific and nonspecific complexes, *Proteins: Struct. Funct., Bioinf. 57*, 772−782.
34. Duan, J., and Nilsson, L. (2002) The role of residue 50 and hydration water molecules in homeodomain DNA recognition, *Eur. Biophys. J. 31*, 306−316.
35. Semina, E. V., Reiter, R., Leysenns, N. J., Alward, W. L., Small, K. W., Datson, N. A., Siegel-Bartelt, J., Bierke-Nelson, D., Bitoun, P., Zabel, B. U., Carey, J. C., and Murray, J. C. (1996) Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome, *Nat. Genet. 14*, 392−398.
36. Gage, P. J., and Camper, S. A. (1997) Pituitary homeobox 2, a novel member of the bicoid-related family of homeobox genes, is a potential regulator of anterior structure formation, *Hum. Mol. Genet. 6*, 457−464.
37. Mucchielli, M., Mitsiadis, T. A., Raffo, S., Brunet, J. F., Proust, J. P., and Goridis, C. (1997) Mouse Otlx2/RIEG expression in the odontogenic epithelium precedes tooth initiation and requires mesenchyme-derived signals for its maintenance, *Dev. Biol. 189*, 275−284.
38. Hjalt, T. A., Semina, E. V., Amendt, B. A., and Murray, J. C. (2000) The PITX2 protein in mouse development, *Dev. Dyn. 218*, 195−200.
39. Arakawa, H., Nakamura, T., Zhadanov, A. B., Fidanza, V., Yano, T., Bullrich, F., Shimizu, M., Blechman, J., Mazo, A., Canaani, E., and Croce, C. M. (1998) Identification and Characterization of the ARP1 Gene, a Target for the human acute leukemia ALL1 gene, *Proc. Natl. Acad. Sci. U.S.A. 95*, 4573−4578.
40. Kitamura, K., Miura, H., Yanazawa, M., Miyashita, T., and Kato, K. (1997) Expression patterns of Brx1 (Rieg gene), Sonic hedgehog, Nkx2.2, Dlx1 and Arx during zona limitans intrathalamica and embryonic ventral lateral geniculate nuclear formation, *Mech. Dev. 67*, 83−96.
41. Gage, P. J., Suh, H., and Camper, S. A. (1999) Dosage requirement of PITX2 for development of multiple organs, *Development 126*, 4643−4651.
42. Lu, M.-F., Pressman, C., Dyer, R., Johnson, R. L., and Martin, J. F. (1999) Function of Rieger syndrome gene in left-right asymmetry and craniofacial development, *Nature 401*, 276−278.
43. Lin, C. R., Kioussi, C., O'Connell, S., Briata, P., Szeto, D., Liu, F., Izpisua-Belmonte, J. C., and Rosenfeld, M. G. (1999) PITX2 regulates lung asymmetry, cardiac positioning and pituitary and tooth morphogenesis, *Nature 401*, 279−282.
44. Xia, K., Wu, L., Liu, X., Xi, X., Liang, D., Zheng, D., Cai, F., Pan, Q., Long, Z., Dai, H., Hu, Z., Tang, B., Zhang, Z., and Xia, J. (2004) Mutation in PITX2 is associcated with ring dermoid cornea, *J. Med. Genet. 41*, e129.
45. Priston, M., Kozlowski, K., Gill, D., Letwin, K., Buys, Y., Levin, A. V., Walter, M. A., and Heon, E. (2001) Functional analyses of two newly identified PITX2 mutants reveal a novel molecular mechanism for Axenfeld-Rieger syndrome, *Hum. Mol. Genet. 10*, 1631−1638.
46. Lines, M. A., Kozlowski, K., Kulak, S. C., Allingham, R. R., Heon, E., Ritch, R., Levin, A. V., Shields, M. B., Damji, K. F., Newlin, A., and Walter, M. A. (2004) Characterization and prevalence of PITX2 microdeletions and mutations in Axenfeld-Rieger malformations, *Invest. Ophthalmol. Vis. Sci. 45*, 828−833.
47. Phillips, J. C. (2002) Four novel mutations in the PITX2 gene in patients with Axenfeld-Rieger syndrome, *Ophthalmic Res. 34*, 324−326.
48. Kulak, S. C., Kozlowski, K., Semina, E. V., Pearce, W. G., and Walter, M. A. (1998) Mutation in the RIEG1 gene in patients with iridogoniodysgenesis syndrome, *Hum. Mol. Genet. 7*, 1113−1117.
49. Grzesiek, S., and Bax, A. (1992) Improved 3D triple-resonance NMR techniques applied to a 31-Kda Protein, *J. Magn. Reson. 96*, 432−440.
50. Ikura, M., Kay, L. E., and Bax, A. (1990) A novel-approach for sequential assignment of H-1, C-13, and N-15 spectra of larger proteins: heteronuclear triple-resonance 3-dimensional NMR spectroscopy: application to calmodulin, *Biochemistry 29*, 4659−4667.
51. Kay, L. E., Xu, G. Y., and Yamazaki, T. (1994) Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation, *J. Magn. Reson., Ser. A 109*, 129−133.
52. Muhandiram, D. R., and Kay, L. E. (1994) Gradient-enhanced triple-resonance 3-dimensional NMR experiments with improved sensitivity, *J. Magn. Reson., Ser. B 103*, 203−216.
53. Grzesiek, S., and Bax, A. (1992) Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR, *J. Am. Chem. Soc. 114*, 6291−6293.
54. Wittekind, M., and Mueller, L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the α-carbon and β-carbon resonances in proteins, *J. Magn. Reson., Ser. B 101*, 201−205.
55. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog. NMR Spectrosc. 34*, 93−158.
56. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) Backbone ^1H and ^{15}N resonance assignments of the N-terminal SH3 domain of *drk* in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques, *J. Biomol. NMR 4*, 845−858.
57. Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc. 114*, 10663−10665.
58. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) Overcoming the overlap problem in the assignment of H-1-NMR spectra of larger proteins by use of 3-dimensional heteronuclear H-1-N-15 Hart-

mann-Hahn multiple quantum coherence and nuclear Overhauser multiple quantum coherence spectroscopy-application to interleukin-1-beta, *Biochemistry 28*, 6150−6156.
59. Zerbe, O., Szyperski, T., Ottiger, M., and Wüthrich, K. (1996) Three-dimensional ^{1}H-TOCSY-relayed ct-[^{13}C,^{1}H]-HMQC for aromatic spin system identification in uniformly ^{13}C-labeled proteins, *J. Biomol. NMR 7*, 99−106.
60. Vuister, G. W., and Bax, A. (1993) Quantitative J correlations: A new approach for measuring homonuclear three-bond J($H^{N}H^{\alpha}$) coupling constants in ^{15}N-enriched proteins, *J. Am. Chem. Soc. 115*, 7772−7777.
61. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
62. Otting, G., and Wüthrich, K. (1990) Heteronuclear filters in two-dimensional [^{1}H,^{1}H]-NMR spectroscopy: combined use with isotope labeling for studies of macromolecular conformation and intermolecular interactions, *Q. Rev. Biophys. 23*, 39−96.
63. Breeze, A. L. (2000) Isotope-filtered NMR methods for the study of biomolecular structure and interactions, *Prog. NMR Spectrosc. 36*, 323−372.
64. Talluri, S., and Wagner, G. (1996) An optimized 3D NOESY-HSQC, *J. Magn. Reson., Ser. B 112*, 200−205.
65. Stuart, A. C., Borzilleri, K. A., Withka, J. M., and Palmer, A. G. (1999) Compensating for variations in ^{1}H−^{13}C scalar coupling constants in isotope-filtered NMR experiments, *J. Am. Chem. Soc. 121*, 5346−5347.
66. Lee, W., Revington, M. J., Arrowsmith, C., and Kay, L. E. (1994) A pulsed field gradient isotope-filtered 3D ^{13}C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes, *FEBS Lett. 350*, 87−90.
67. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−293.
68. Goddard, T. D., Kneller, D. G. *SPARKY 3*, University of California, San Francisco.
69. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol. 273*, 283−298.
70. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA, *J. Mol. Biol. 319*, 209−227.
71. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kalman, P. A. (1996) *AMBER7*, University of California, San Francisco.
72. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., deBolt, S., Ferguson, D., Seibel, G., and Kollman, P. A. (1995) AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules, *Comput. Phys. Commun. 91*, 1−41.
73. Wuttke, D. S., Foster, M. P., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (1997) Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: Determinants of affinity and sequence specificity, *J. Mol. Biol. 273*, 183−206.
74. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
75. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR 8*, 477−486.
76. Konradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55.
77. Spera, S., and Bax, A. (1991) Empirical correlation between protein backbone conformation and C^{α} and C^{β} ^{13}C nuclear magnetic resonance shifts, *J. Am. Chem. Soc. 113*, 5490−5492.
78. Luginbuhl, P., Szyperski, T., and Wüthrich, K. (1995) Statistical basis for the use of $^{13}C^{\alpha}$ chemical shifts in protein structure determination, *J. Magn. Reson. B 109*, 229−233.
79. Scott, M. P., Tamkun, J. W., and Hartzell, G. W. (1989) The structure and function of the homeodomain, *Biochim. Biophs. Acta 989*, 25−48.
80. Billeter, M. (1996) Homeodomain-type DNA recognition, *Prog. Biophys. Mol. Biol. 66*, 211−225.
81. Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., and Wüthrich, K. (1989) The Structure of the Antennapedia Homeodomain Determined by NMR Spectroscopy in Solution: Comparison with Prokaryotic Repressors, *Cell 59*, 573−580.
82. Kornberg, T. B. (1993) Understanding the Homeodomain, *J. Biol. Chem. 268*, 26813−26816.
83. Banerjee-Basu, S., and Baxevanis, A. (1999) Threading analysis of the PITX2 homeodomain: predicted structural effects of mutations causing Rieger Syndrome and Iridogoniodysgenesis, *Hum. Mutat. 14*, 312−319.
84. Billeter, M., Qian, Y. Q., Otting, G., Muller, M., Gehring, W. J., and Wüthrich, K. (1993) Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex, *J. Mol. Biol. 234*, 1084−1093.
85. Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions, *Cell 63*, 579−590.
86. Qian, Y. Q., Furukubo-Tokunaga, K., Resendez-Perez, D., Muller, M., Gehring, W. J., and Wüthrich, K. (1994) Nuclear magnetic resonance solution structure of the *Fushi tarazu* homeodomain from *Drosophila* and comparison with the *Antennapedia* homeodomain, *J. Mol Biol. 238*, 333−345.
87. Gruschus, J. M., Tsao, D. H., Wang, L. H., Nirenberg, M., and Ferretti, J. A. (1997) Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity, *Biochemistry 36*, 5372−5380.
88. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions, *Cell 67*, 517−528.
89. Damante, G., Tell, G., Leonardi, A., Fogolari, F., Bortolotti, N., DiLauro, R., and Formisano, S. (1994) Analysis of the conformation and stability of rat TTF-1 homeodomain by circular dichroism, *FEBS. Lett. 354*, 293−296.
90. Carra, J. H., and Privalov, P. L. (1997) Energetics of folding and DNA binding of the MAT alpha 2 homeodomain, *Biochemistry 36*, 526−535.
91. Otting, G., Qian, Y. Q., Muller, M., Affolter, M., Gehring, W., and Wüthrich, K. (1988) Secondary structure determination for the *Antennapedia* homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif, *EMBO J. 7*, 4305−4309.
92. Yamamoto, K., Yee, C. C., Shirakawa, M., and Kyogoku, Y. (1992) Characterization of the bacterially expressed Drosophila engrailed homeodomain, *J. Biochem. Tokyo 111*, 793−797.
93. Iurcu-Mustata, G., Van Belle, D., Wintjens, R., Prévost, M., and Rooman, M. (2001) Role of salt bridges in homeodomains investigated by structural analyses and molecular dynamics simulations, *Biopolymers 59*, 145−159.
94. Jones, S., van Heyningen, P., Berman, H. M., and Thornton, J. M. (1999) Protein-DNA interactions: A structural analysis, *J. Mol. Biol. 287*, 877−896.
95. Fernandez, C., Szyperski, T., Billeter, M., Ono, A., Iwai, H., Kainosho, M., and Wüthrich, K. (1999) Conformational changes of the BS2 operator DNA upon complex formation with the *Antennapedia* homeodomain studied by NMR with 13C/15N-labeled DNA, *J. Mol. Biol. 292*, 609−617.
96. McDonald, I. K., and Thornton, J. M. (1994) Satisfying Hydrogen Bonding Potential in Proteins, *J. Mol. Biol. 238*, 777−793.
97. Qian, Y. Q., Otting, G., Billeter, M., Müller, M., Gehring, W., and Wüthrich, 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.
98. Foster, M. P., Wuttke, D. S., Radhakrishnan, I., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (1997) Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA, *Nat. Struct. Biol. 4*, 605−608.
99. Nishikawa, T., Okamura, H., Nagadoi, A., König, P., Rhodes, D., and Nishimura, Y. (2001) Solution structure of a telomeric DNA complex of human TRF1, *Structure 9*, 1237−1251.
100. Doig, A. J., and Sternberg, M. J. E. (1995) Side-chain conformational entropy in protein folding, *Protein Sci. 4*, 2247−2251.
101. Mandel-Gutfreund, Y., Schueler, O., and Margalit, H. (1995) Comprehensive analysis of hydrogen bonds in regulatory protein-

DNA complexes: in search of common principles, *J. Mol. Biol. 253*, 370−382.
102. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
103. Saadi, I., Semina, E. V., Amendt, B. A., Harris, D. J., Murphy, K. P., Murray, J. C., and Russo, A. F. (2001) Identification of a Dominant Negative Homeodomain Mutation in Rieger Syndrome, *J. Biol. Chem. 276*, 23034−23041.
104. Heon, E., Sheth, B. P., Kalenak, J. W., Sunden, S. L., Streb, L. M., Taylor, C. M., Alward, W. L., Sheffield, V. C., and Stone, E. M. (1995) Linkage of autosomal dominant iris hypoplasia to the region of the Rieger Syndrome locus (4q25), *Hum. Mol. Genet. 4*, 1435−1439.
105. Alward, W. L. M., Semina, E. V., Kalenak, J. W., Heon, E., Sheth, B. P., Stone, E. M., and Murray, J. C. (1998) Autosomal Dominant Iris Hypoplasia is Caused by a Mutation in the Rieger Syndrome (RIEG/PITX2) Gene, *Am. J. Opthalmol. 125*, 98−100.
106. Chisholm, E. A., and Chudley, A. E. (1983) Autosomal dominant iridogoniodysgenesis with associated somatic anomalies: four-generation family with Rieger's syndrome, *Br. J. Ophthalmol. 67*, 529−534.
107. Walter, M. A., Mirzayans, F., Mears, A. J., Hickey, K., and Pearce, W. G. (1996) Autosomal-dominant iridogoniodysgenesis and Axenfeld-Rieger syndrome are genetically distinct, *Ophthalmology 103*, 1907−1915.
108. Murray, J. C., Bennett, S. R., Kwitek, A. E., Small, K. W., Schinzel, A., Alward, W. L., Weber, J. L., Bell, G. I., and Buetow, K. H. (1992) Linkage of Rieger Syndrome to the region of the epidermal growth factor gene on chromosome 4, *Nat. Genet. 2*, 46−49.
109. Quentien, M., Pitoia, F., Gunz, G., Guillet, M., Enjalbert, A., and Pellegrini, I. (2002) Regulation of Prolactin, GH, and Pit-1 Gene Expression in Anterior Pituitary by PITX2: An Approach Using PITX2 Mutants, *Endocrinology 143*, 2839−2851.
110. Amendt, B. A., Sutherland, L. B., and Russo, A. F. (1999) Multifunctional role of the Pitx2 homeodomain protein C-terminal tail, *Mol. Cell. Biol. 19*, 7001−7010.
111. Boncinelli, E. (1997) Homeobox genes and disease, *Curr. Opin. Genet. Dev. 7*, 331−337.
112. Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B. R. (1996) Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13, *Science 272*, 548−551.
113. Nakamura, T., Largaespada, D. A., Lee, M. P., Johnson, L. A., Ohyashiki, K., Toyama, K., Chen, S. J., Willman, C. L., Chen, I. M., Feinberg, A. P., Jenkins, N. A., Copeland, N. G., and Shaughnessy, J. D., Jr. (1996) Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)-(p15;p15) in human myeloid leukaemia, *Nat. Genet. 12*, 154−158.
114. D'Elia, A. V., Tell, G., Paron, I., Pellizzari, L., Lonigro, R., and Damante, G. (2001) Missense Mutations of Human Homeoboxes: A Review, *Hum. Mutat. 18*, 361−374.
115. Borrow, J., Shearman, A. M., Stanton, V. P., Jr., Becher, R., Collins, T., Williams, A. J., Dube, I., Katz, F., Kwong, Y. L., Morris, C., Ohyashiki, K., Toyama, K., Rowley, J., and Housman, D. E. (1996) The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9, *Nat. Genet. 12*, 159−167.
116. Watanabe, K., and Lambowitz, A. M. (2004) High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA, *RNA 10*, 1433−1443.
117. Monsalve, M., Calles, B., Mencia, M., Rojo, F., and Salas, M. (1998) Binding of phage Φ29 Protein p4 to the Early A2c Promoter: Recruitment of a Repressor by the RNA Polymerase, *J. Mol. Biol. 283*, 559−569.

BI0473253