Accelerated Publications

Structural Basis for DNA Recognition by the Human PAX3 Homeodomain[†,‡]

Gabriel Birrane, Aditi Soni, and John A. A. Ladias*

Molecular Medicine Laboratory and Macromolecular Crystallography Unit, Division of Experimental Medicine, Harvard Medical School, Boston, Massachusetts 02115

Received November 4, 2008; Revised Manuscript Received January 7, 2009

ABSTRACT: The transcription regulatory protein PAX3 binds to cognate DNA sequences through two DNA-binding domains, a paired domain and a homeodomain, and has important functions during neurogenesis and myogenesis. In humans, mutations in the *PAX3* gene cause Waardenburg syndrome, whereas a chromosomal translocation that generates a *PAX3*−*FOXO1* fusion gene is associated with the development of alveolar rhabdomyosarcoma. We have determined the crystal structure of the human PAX3 homeodomain in complex with a palindromic DNA containing two inverted TAATC sequences at 1.95 Å resolution. Two homeodomains bind to DNA as a symmetric dimer, inducing a 3° bend in the DNA helix. The N-terminal arm of the homeodomain inserts into the minor groove and makes direct and water-mediated interactions with bases and the sugar−phosphate backbone. The recognition helix fits directly into the major groove, and an elaborate network of structurally conserved water molecules mediates the majority of protein−DNA interactions. The structure elucidates the role of serine 50 in selection of the CG sequence immediately 3′ of the TAAT motif by PAX class homeodomains and provides insights into the molecular mechanisms by which certain Waardenburg syndrome-associated missense mutations could destabilize the fold of the PAX3 homeodomain whereas others could affect its interaction with DNA.

PAX3 belongs to the nine-member PAX family of metazoan transcription regulatory proteins, which are characterized by the presence of a 128-residue DNA-binding module at their N-termini, termed the paired domain (PD)[1] (*1, 2*). PAX3, PAX4, PAX6, and PAX7 contain an additional DNA-binding motif located C-terminally with respect to PD, the homeodomain (HD). During embryonic development, the *PAX3* gene is expressed in the central nervous system, craniofacial tissue, trunk neural crest, and skeletal muscle and plays critical roles in neurogenesis and myogenesis (*1, 2*). Mice carrying mutations in the *Pax3* gene have severe neural tube defects and fail to develop limb muscle (*1, 2*). In humans, mutations in the *PAX3* gene cause Waardenburg syndrome, an autosomal dominant condition characterized by craniofacial abnormalities, deafness, and pigmentary disturbances (*3*−*9*). A subset of Waardenburg syndrome-associated missense mutations occur in the PAX3 HD (Figure 1A). Furthermore, the t(2;13)(q35;q14) translocation of *PAX3* on chromosome 2 and *FOXO1* on chromosome 13 generates the chimeric transcription factor PAX3-FOXO1 that contains the PD and HD of PAX3 and the transcriptional activation domain of FOXO1, and is associated with the development of the pediatric muscle tumor alveolar rhabdomyosarcoma (*10*).

The 60-residue HD is a conserved DNA-binding module found in a large family of transcription regulatory proteins that control patterning during embryogenesis and development (*11*). A number of studies elucidated the HD structure and its mode of DNA recognition (*12*−*28*). The HD fold comprises three α-helices (α1, α2, and α3), connected by short loops, and an extended N-terminal arm which is disordered in the absence of DNA. Upon interaction with DNA, the N-terminal arm inserts into the minor groove making sequence-specific contacts, whereas the recognition helix α3 fits into the major groove where it interacts with bases in both strands of the DNA and the sugar−phosphate backbone. Previous studies have pointed to the critical role of the residue at position 50 (using the conventional HD amino acid numbering) in specifying the sequence immediately 3′ to the TAAT core, with Ser50 preferring TAATYG, Gln50 selecting TAATYNR, and Lys50 recognizing TAATCCG sequences, where Y is pyrimidine, R is purine, and N is any nucleotide (*26*−*28*).

All HDs of the PAX/paired class have a serine at position 50 and bind cooperatively to DNA containing two inverted TAAT motifs separated by either two base pairs (P2 site) or three base pairs (P3 site), showing a stronger preference for P2 elements (*26*). By contrast, the *Drosophila* paired HD

[†] This work was supported by a charitable gift from Mortimer B. Zuckerman, Grants GM065520, DK062162, and AG021964 from the National Institutes of Health, Grants DAMD170210300, DAMD170310563, W81XWH0510622, and W81XWH0710178 from the U.S. Department of Defense, a grant from the Dystonia Medical Research Foundation, and Temple Discovery Award TLL035927 from the Alzheimer's Association to J.A.A.L.

[‡] The atomic coordinates and structure factors have been deposited in the Protein Data Bank as entry 3CMY.

* To whom correspondence should be addressed: Harvard Institutes of Medicine, 4 Blackfan Circle, Boston, MA 02115. Telephone: (617) 667-0064. Fax: (617) 975-5241. E-mail: johnladias@gmail.com.

[1] Abbreviations: A, adenine; C, cytosine; G, guanine; HD, homeodomain; PD, paired domain; T, thymine.

10.1021/bi802052y CCC: $40.75
Published on Web 01/21/2009

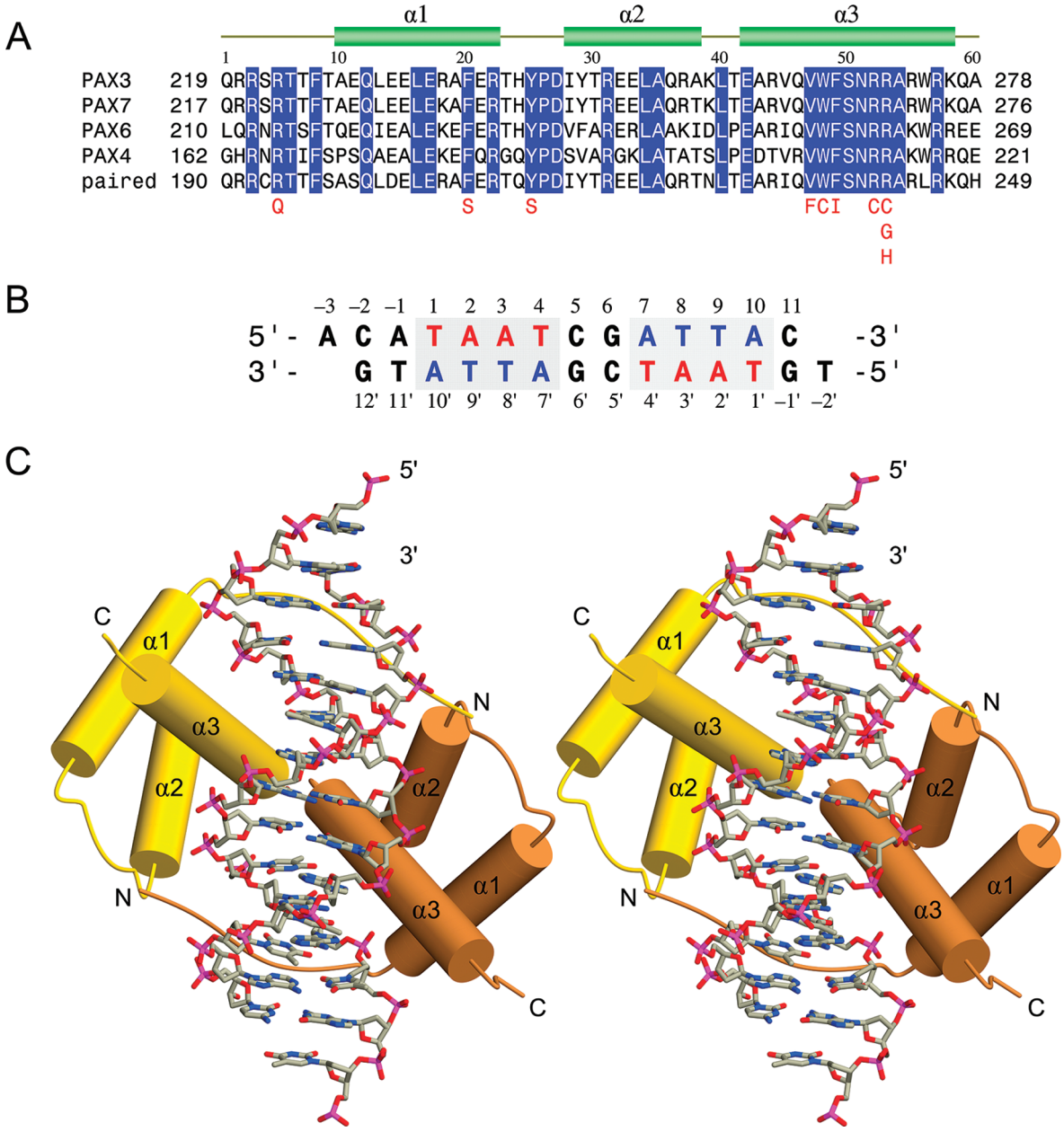

FIGURE 1: (A) Sequence alignment of the HDs of human PAX3 (GenBank accession number NP_852122), PAX7 (NP_002575), PAX6 (NP_000271), and PAX4 (NP_006184) and *Drosophila* paired (NP_523556). The sequences were aligned using CLUSTAL W (*38*). Helices are depicted as green cylinders. Identical amino acids in all sequences are shown as white letters on a blue background. Missense mutations associated with the Waardenburg syndrome are shown at the bottom. (B) Sequence of the DNA fragment used in the cocrystallization with the PAX3 HD. The TAAT core sequence is shown with red letters. (C) Stereoview of the PAX3 HD−DNA complex. The DNA is represented as a stick model, and the α-helices of the two HDs are represented as brown and yellow cylinders. Carbon, nitrogen, oxygen, and phosphorus atoms are colored gray, blue, red, and magenta, respectively. This figure was made using BOBSCRIPT (*39*) and POV-Ray (www.povray.org).

carrying the S50Q mutation exhibits increased cooperativity for P3 sites but does not bind to P2 DNA (*26*). The structure of the mutant paired HD(S50Q) bound to a P3 site revealed that two HDs bind to DNA as a dimer, inducing an overall bend of ~21° in the DNA (*15*). However, the structural basis for P2 recognition by wild-type PAX HDs has remained elusive.

Here we present the crystal structure of the wild-type PAX3 HD homodimer in complex with P2 DNA. The HD adopts the canonical fold comprising three α-helices, with the recognition helix fitting into the major groove. The N-terminal arm inserts into the adjacent minor groove and participates in DNA recognition and HD homodimerization. The DNA is in nearly straight B-form, and a number of water molecules mediate the majority of protein−DNA interactions. The structure elucidates the role of Ser50 in selection of the CG sequence 3′ to the TAAT motif by PAX class HDs and illustrates how a number of missense mutations associated with the Waardenburg syndrome adversely affect the structure and function of the PAX3 HD.

EXPERIMENTAL PROCEDURES

Protein Purification. A DNA fragment encoding the human PAX3 HD (residues 219−278) was amplified using the polymerase chain reaction and cloned into a modified

pET-6H vector (*29*). The PAX3 HD protein carrying an N-terminal hexahistidine tag followed by a tobacco etch virus (TEV) protease site (ENLYFQ↓G) was expressed in *Escherichia coli* BL21(DE3) cells grown at 37 °C until they reached an absorbance at 600 nm of 0.6, followed by induction with 0.4 mM IPTG for 20 h at 20 °C. The cells were harvested by centrifugation, resuspended in phosphate-buffered saline supplemented with protease inhibitor cocktail tablets (Roche Applied Science) on ice, and lysed on an EmulsiFlex-C3 homogenizer (Avestin). Soluble 6His-PAX3 HD protein was purified on Ni-NTA resin (Qiagen), eluted with 200 mM imidazole, and dialyzed against a buffer containing 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 1 mM DTT, and 0.5 mM EDTA. The hexahistidine tag was removed by digestion with TEV protease for 15 h at 4 °C, and the protein was further purified by size exclusion chromatography on a Superdex 30 column (GE Healthcare). The PAX3 HD protein was concentrated to 14 mg/mL by ultracentrifugation.

Crystallization of the Protein−DNA Complex. Synthetic DNA oligonucleotides with the sequences 5′-ACATAATC-GATTAC-3′ and 5′-TGTAATCGATTATG-3′ were purchased from Genemed Synthesis and purified by HPLC, resuspended in 5 mM HEPES (pH 7.5), 50 mM NaCl, and 0.5 mM DTT, mixed at an equimolar ratio, and annealed by being heated to 95 °C for 10 min and slowly cooled to 4 °C. The annealed DNA was precipitated with 70% ethanol and resuspended in water to a final concentration of 14 mg/mL as judged by the absorbance at 260 nm. Equimolar amounts of the double-stranded DNA and PAX3 HD protein were mixed at 4 °C for 30 min, and the resulting complex was crystallized in 19−22% polyethylene glycol 8000, 100 mM ammonium sulfate, 20 mM $MgCl_2$, 15% 2-methyl-2,4-pentanediol, and 50 mM HEPES (pH 7.6) at 20 °C by the sitting drop vapor diffusion method. Crystals were flash-frozen in a liquid nitrogen stream directly from the mother liquor. Native data were collected on beamline X12B at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory (Upton, NY). The data were processed and scaled using HKL2000 (*30*). The crystals belong to space group $C222_1$ with the following unit cell dimensions: a = 43.3 Å, b = 62.2 Å, and c = 92.6 Å.

Structure Determination and Refinement. The structure was determined by molecular replacement with MOLREP (*31*). The protein atoms of the *Drosophila* paired HD(S50Q) crystal structure (Protein Data Bank entry 1FJL), which shares 85% sequence identity with PAX3 HD, were used as the search model. The coordinates were subjected to several cycles of simulated annealing in CNS (*32*), followed by automated model building and location of water molecules with ARP/wARP (*33*) in combination with manual inspection of σ_A-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps using COOT (*34*). The asymmetric unit contains one HD molecule bound to a TAAT motif, and the complete complex is generated by a crystallographic 2-fold rotation about the a axis through the center of the P2 site. The DNA stacks in two orientations, one "up" and one "down" with respect to the c axis. The σ_A-weighted $F_o - F_c$ map clearly indicated the presence of mixed base pairs outside the palindromic region and the pseudo-two-fold axis of symmetry. An occupancy of 0.5 was assigned to each twin during refinement. With the exception of Ala60, all HD residues are clearly visible in the electron density map. The final model was refined isotropically with TLS parameters in REFMAC5 (*35*) and was validated using MolProbity (*36*). The structure determination and refinement statistics are listed in Table 1.

Table 1: Structure Determination and Refinement Statistics

resolution range (Å)	30−1.95 (2.02−1.95)[a]
wavelength (Å)	1.000
no. of observed reflections	73662
no. of unique reflections	9726
completeness (%)	99.7 (100)[a]
redundancy	7.6 (6.7)[a]
R_{sym} (%)[b]	6.4 (58.4)[a]
overall $\langle I/\sigma(I)\rangle$	27.1 (3.0)[a]
R_{cryst} (%)[c]	17.2
R_{free} (%)[d]	21.3
Ramachandran plot	
most favored (%)	100.0
additionally allowed (%)	0.0
generously allowed (%)	0.0
disallowed (%)	0.0
bond lengths[e] (Å)	0.01
bond angles[e] (deg)	1.53

[a] Values in parentheses are for the highest-resolution shell. [b] $R_{sym} = \sum|(I - \langle I\rangle)|/\sum(I)$, where I is the observed integrated intensity, $\langle I\rangle$ is the average integrated intensity obtained from multiple measurements, and the summation is over all observed reflections. [c] $R_{cryst} = \sum||F_o| - k|F_c||/\sum|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. [d] R_{free} is calculated as R_{cryst} using 5% of the reflections chosen randomly and omitted from the refinement calculations. [e] Bond lengths and angles are root-mean-square deviations from ideal values.

RESULTS AND DISCUSSION

Structure Determination. We expressed the human PAX3 HD (Figure 1A) in *E. coli* cells and purified it to homogeneity, using a combination of affinity and size exclusion chromatography. The protein was mixed stoichiometrically with a P2 palindromic DNA (Figure 1B), and the resulting nucleoprotein complex was crystallized using the sitting drop vapor diffusion method. The structure was determined by molecular replacement using the protein atoms of the paired HD(S50Q)−DNA crystal structure (Protein Data Bank entry 1FJL) as the search model in space group $P2_1$. Inspection of the $F_o - F_c$ electron density maps showed the presence of mixed base pairs outside the palindromic region, indicating a pseudo-two-fold axis of symmetry through the center of the P2 site and perpendicular to the helical axis. The data were reprocessed in the higher symmetry space group $C222_1$ and refined to an R_{cryst} of 17.2% and an R_{free} of 21.3% (Table 1).

Overall Structure of the Nucleoprotein Complex. Two HDs bind to DNA in a head-to-head arrangement (Figure 1C). The DNA molecules stack end-to-end, forming a pseudo-continuous helix which runs parallel to the crystallographic c axis. The PAX3 HD folds into a globular structure very similar to that of other HDs (*12*−*25*), comprising an extended N-terminal arm (residues 1−9) that does not conform to any secondary structure, and three α-helices, α1 (residues 10−22), α2 (residues 28−38), and recognition helix α3 (residues 42−58) connected by short loops (Figure 1A,C). Helices α1 and α2 are arranged in an antiparallel fashion relative to each other and perpendicular to α3, which fits into the major groove. The HD fold is stabilized by hydrophobic interactions involving Phe8, Leu13, Leu16, Phe20, Pro26, Leu34, Ala38, Leu40, Val45, Trp48, Phe49, and Arg52. Hydrogen bonding between the side chain pairs

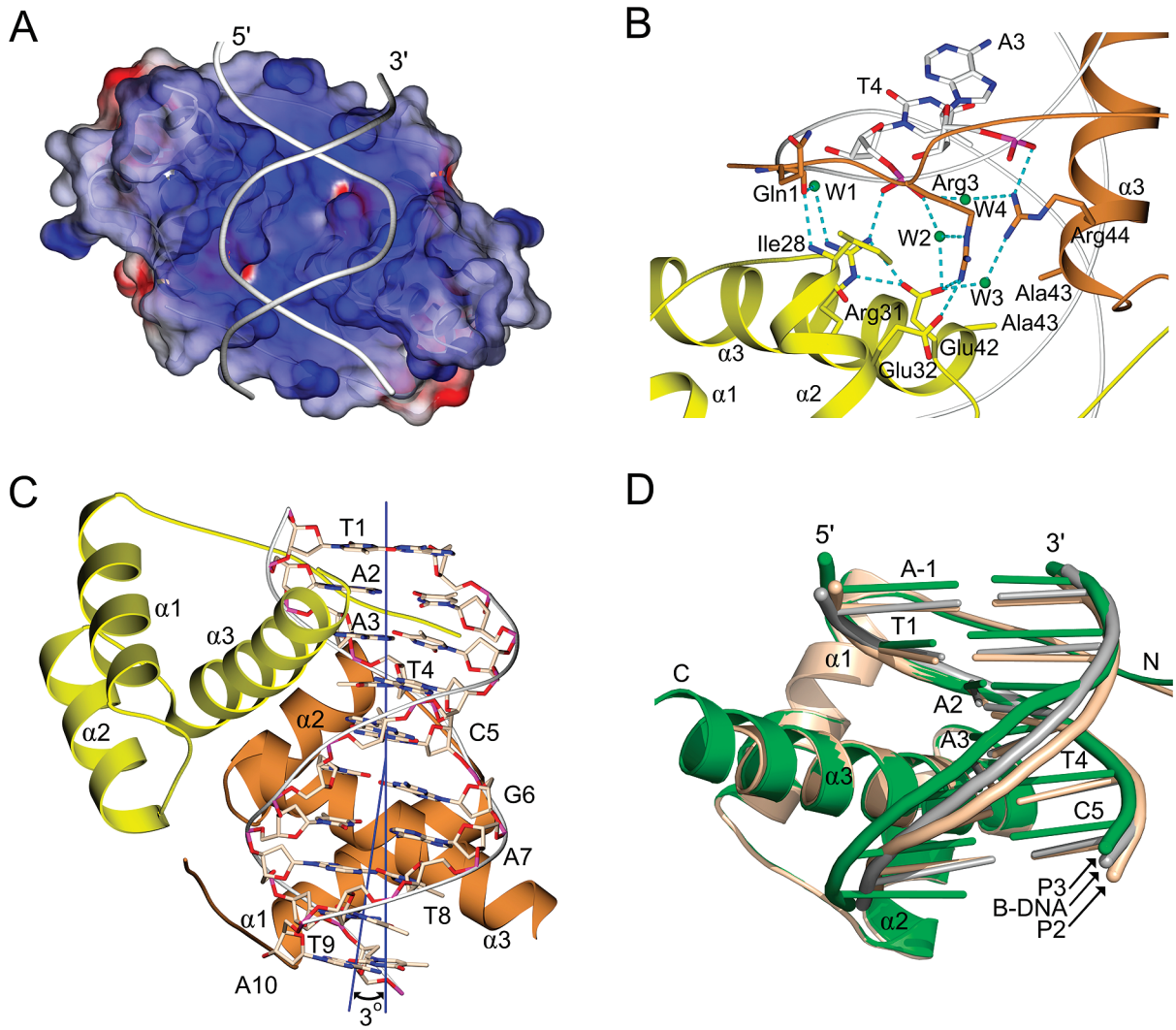

FIGURE 2: (A) Electrostatic potential of the HD dimer bound to DNA. Electrostatic potentials were calculated with GRASP (*40*) and are colored red (acidic, $-10k_BT$), white (neutral, $0k_BT$), and blue (basic, $10k_BT$). (B) Stick model of the HD dimer interface. One molecule is colored yellow and the other brown. Water molecules are depicted as green spheres, and hydrogen bonds are depicted as dashed cyan lines. Carbon atoms of the DNA are colored white. (C) Illustration of the bend imposed on the palindromic DNA region due to protein binding. The bend angle was calculated using CURVES (*41*). (D) Least-squares superposition of the protein atoms of paired HD(S50Q) bound to P3 (Protein Data Bank entry 1FJL) with the protein atoms of the wild-type PAX3 HD bound to P2. Ideal B-form DNA having the same sequence as the P2 element was generated using 3DNA (*42*). This figure was made using PyMOL (www.pymol.org).

of Glu15 and Arg37, Glu17 and Arg52, Glu17 and Arg55, and Glu42 and Arg31 also contribute to the stability of the fold.

Dimer Interface. The HD dimer forms a continuous surface that interacts with the DNA, burying ~900 Å^2 of area between the protein chains and ~1800 Å^2 between the protein and DNA (Figure 2A). The N-terminal arm of one HD interacts with residues in the N-terminal regions of α2 and α3 of the other molecule of the dimer. The carbonyl oxygen of Gln1 hydrogen bonds directly with the amide nitrogen of Ile28 and through a water molecule with the $N^{\eta 1}$ atom of Arg31 (Figure 2B). The side chain of the conserved Arg3 makes hydrogen bonds to Glu32 and Glu42 and hydrophobic contacts with Ile28 of the neighboring HD. The dimeric interface is further stabilized by van der Waals interactions between the Ala43 methyl groups at the N-terminal regions of the juxtaposed α3 helices.

Structure of the DNA. The DNA adopts a standard B-DNA helix with a mean rise of 3.4 Å per base pair and a mean helical twist of 37.2°. An overall bend of 3° is caused by protein binding (Figure 2C). Most of the HD−DNA structures described to date show a bend of 10−13° in the DNA, whereas in the MATa1/MATα2−DNA complex, there is a 60° bend (*14*). Notably, the paired HD(S50Q) induces a bend of 21° on the P3 DNA (*15*), which is considerably larger than that seen in the structure presented here (Figure 2D). The dramatic differences in the DNA bend by the wild-type PAX3 and paired(S50Q) HDs most likely account for the differences in the electrophoretic mobilities of the two nucleoprotein complexes (*26*).

Protein−DNA Recognition. HD residues Gln1, Arg2, Arg3, Arg5, Thr6, Tyr25, Pro26, Arg31, Glu42, Arg44, Val47, Trp48, Ser50, Asn51, Arg53, and Arg57 mediate the interactions with DNA (Figure 3A). The N-terminal arm of the HD inserts into the minor groove and forms base-specific interactions with the DNA. Gln1 forms a water-mediated hydrogen bond with the T11′ phosphate. The guanido group of Arg2 forms an electrostatic interaction with T9′, a direct hydrogen bond to A3, and water-mediated hydrogen bonds to A2, A3, T4, T8′, and T9′ (Figure 3A,B). The N^{ε} and amide

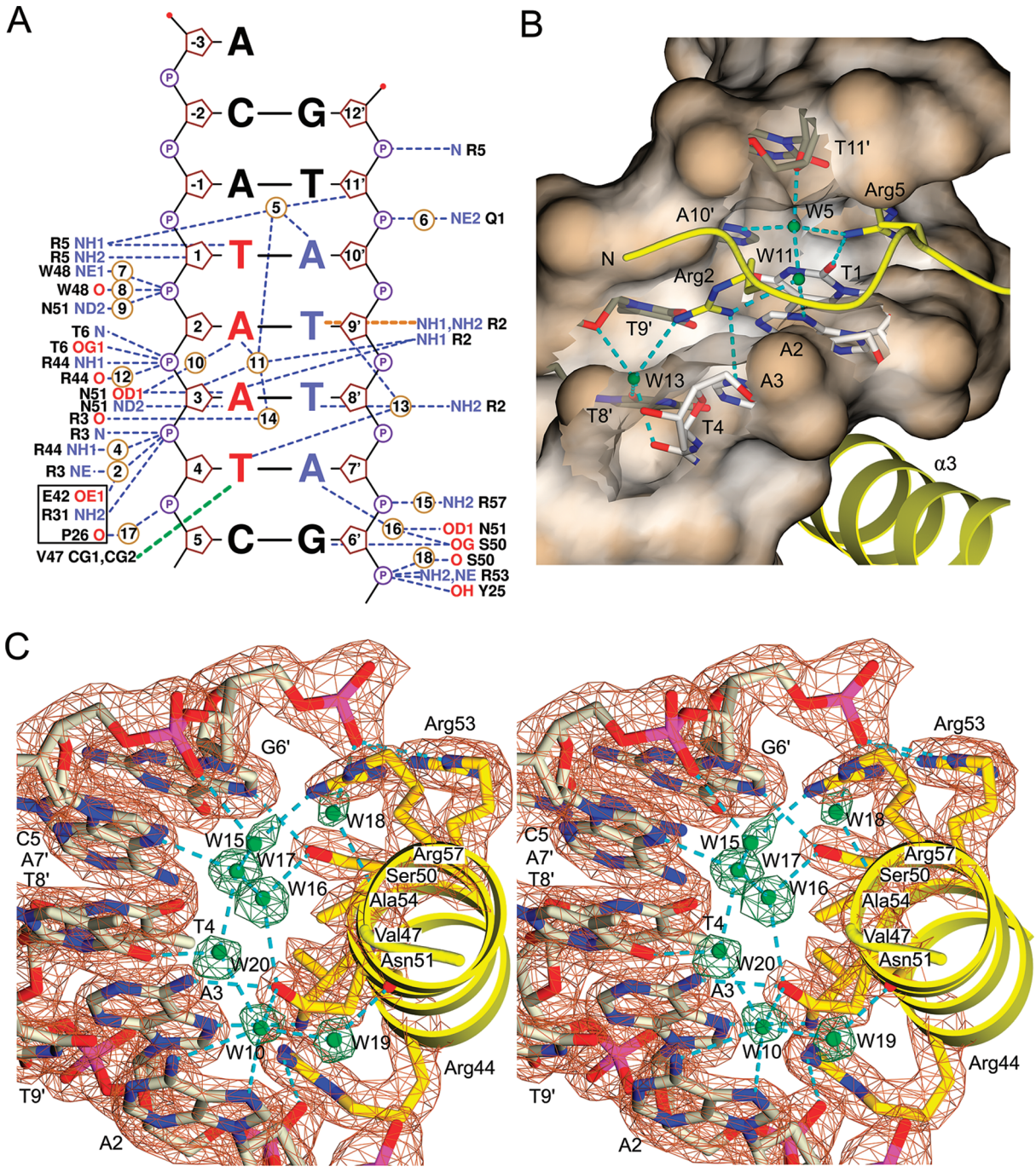

FIGURE 3: Structure of the PAX3 HD−DNA interface. (A) Schematic diagram of the HD−DNA contacts. Interactions involving amino acids from the homodimeric partner are contained within a black box. Hydrogen bonds, nonbonded electrostatic interactions, and hydrophobic interactions are depicted with dashed blue, orange, and green lines, respectively. Water molecules are depicted as numbered circles. (B) Direct and water-mediated interactions involving positively charged side chains in the N-terminal arm of the PAX3 HD (shown as a ribbon model) and bases in the minor groove of DNA (shown as a molecular surface). This figure was made with GRASP, POVSCRIPT (*43*), and POV-Ray. (C) Stereoview of interactions of helix α3 with the major groove. A σ_A-weighted $2F_o - F_c$ electron density map calculated at 1.95 Å and contoured at 1σ is superimposed on the DNA (brown), selected HD amino acid chains (brown), and water molecules (green). Atoms are colored as in Figure 1C. Hydrogen bonds are depicted as dashed cyan lines.

N atoms of Arg3 form water-mediated and direct hydrogen bonds to the T4 phosphate, respectively. The guanido group of Arg5 forms direct hydrogen bonds to the O2 atom of T1 and the O4′ atom of its deoxyribose, and water-mediated hydrogen bonds to the N3 atom of A10′ and the O4′ atom of the T11′ sugar (Figure 3A,B). The amide N and $O^{\gamma 1}$ atoms of Thr6 hydrogen bond with the sugar O3′ and backbone O2P atoms of A2 and A3, respectively. The hydroxyl group of Tyr25 forms a hydrogen bond with the O1P atom of G6′. The side chains of Arg31 and Glu42 from the homodimeric partner form direct and water-mediated hydrogen bonds to the T4 phosphate, respectively, whereas the carbonyl oxygen of Pro26 forms a water-mediated hydrogen bond to the C5 phosphate.

In the recognition helix, the guanido group of Arg44 makes a direct hydrogen bond to the O1P atom of A3 and a water-mediated hydrogen bond to the O1P atom of T4, whereas the carbonyl oxygen of Arg44 makes a water-mediated

hydrogen bond to the O1P atom of A3. The isopropyl group of Val47 makes a conserved hydrophobic interaction with the methyl group of T4. The $N^{\epsilon 1}$ and carbonyl oxygen atoms of Trp48 make water-mediated hydrogen bonds to the A2 phosphate, whereas the N^{ε} and $N^{\eta 2}$ atoms of Arg53 hydrogen bond to the O2P atom of G6′. Importantly, the O^{γ} atom of Ser50 makes a direct hydrogen bond to the N^7 atom of G6′ and a water-mediated hydrogen bond to the N^6 atom of A7′ (Figure 3A,C), thereby providing the structural basis for selection of the sequence CG 3′ to the TAAT core by the PAX3 HD dimer. By contrast, in the crystal structure of the paired HD(S50Q), the side chain of Q50 makes a hydrophobic contact with the methyl group of the T4 base at the center of the P3 site, thereby stabilizing the interaction with the DNA (*15*). In the structure presented here, the S50Q mutation would introduce a clash between the N^7 atom of C5′ and the hydrophobic portion of the Gln50 side chain that would interfere with DNA binding, as described previously (*26*). The invariant Asn51 makes several interactions with the DNA. The $O^{\delta 1}$ and $N^{\delta 2}$ atoms of Asn51 make two strictly conserved hydrogen bonds to the N6 and N7 atoms of A3, respectively. The $O^{\delta 1}$ atom also makes water-mediated hydrogen bonds to the N6 atoms of A2 and A7′, whereas the $N^{\delta 2}$ atom makes a water-mediated hydrogen bond to the O1P atom of A2, differing from the paired HD(S50Q) whose Asn51 interacts with T4 via its amide oxygen and a water molecule (*15*).

In contrast to other HDs that have a large hydropobic residue (methionine or isoleucine) at position 54, the PAX, paired, engrailed, and Msx-1 HDs have an alanine, which creates a cavity between the protein and the DNA. This cavity is occupied by a largely conserved network of water molecules. In the structure presented here, the carbonyl oxygen of Asn51 interacts with A2 via two water molecules that are conserved in the paired(S50Q) (*15*) and Msx-1 (*21*) HD−DNA complexes. In the PAX3 and paired(S50Q) HDs, the side chain of Arg57 replaces a water molecule found in Msx-1 and fulfills the hydrogen bonding pattern of the network (Figure 3C). The $N^{\eta 2}$ atom of Arg57 also hydrogen bonds with A7′ through two water molecules that have equivalents in the paired(S50Q) HD−DNA complex. This flexible water network accommodates the changes in the DNA geometry imposed by protein binding.

Disease-Associated Mutations in the PAX3 HD. Missense mutations in the HD that cause Waardenburg syndrome include R5Q, F20S, Y25S, V47F, W48C, F49I, R52C, R53C, R53G, and R53H (*3*−*9*). All involve substitutions of highly conserved amino acids, and most of them are located in the recognition helix (Figure 1A). On the basis of the PAX3 HD−DNA atomic model, these mutations can be grouped into those that could be predicted to have detrimental effects on the stability of the HD fold (F20S, W48C, F49I, and R52C) and those that could affect the interaction of the HD with DNA (R5Q, Y25S, V47F, R53C, R53G, and R53H) (Figure 4).

The phenyl group of the conserved Phe20 protrudes into a hydrophobic pocket composed of the His24, Trp48, Phe49, Trp56, and Arg52 side chains and has a structural role in the protein core. The F20S mutation introduces a small hydrophilic side chain, which could likely disrupt the hydrophobic pocket and destabilize the protein fold. A similar effect on the stability of the fold would be predicted for the

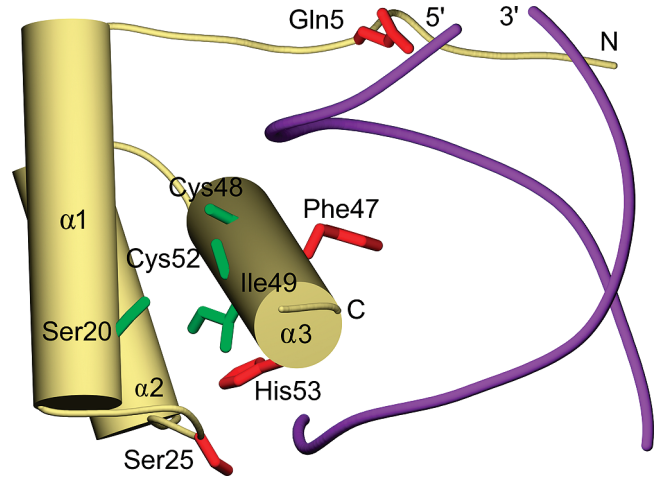

FIGURE 4: Mapping of Waardenburg syndrome-associated missense mutations on the PAX3 HD structure. The side chains of Ser20, Cys48, Ile49, and Cys52 that could destabilize the hydrophobic core are shown as green ball-and-stick models. The side chains of Gln5, Ser25, Phe47, and His53 that could affect the PAX3 HD−DNA interaction are shown as red ball-and-stick models. This figure was made using POVSCRIPT and POV-Ray.

substitution of the bulky indole ring of Trp48 by the smaller and hydrophilic sulfhydryl group in the W48C mutation. Likewise, replacement of the phenyl ring of Phe49 with the smaller side chain of isoleucine in the F49I mutation would create a void that could affect the packing of the protein interior. It is noteworthy that the short isopropyl side chain of Val49 in the MATα2 HD is compensated by the large indole group of Trp19 (*12, 14*). In contrast, the methyl group of Ala19 in the PAX3 HD cannot fill the void in the hydrophobic pocket generated by the F49I mutation. Similarly, the aliphatic portion of the Arg52 side chain makes hydrophobic contacts with Trp48 and Phe20 that contribute to the stability of the protein interior, and its replacement with the sulfhydryl group in R52C would be predicted to destabilize the fold. The guanido group of Arg52 also forms a hydrogen bond with the side chain of Glu17, and the loss of this hydrogen bond would have an additional destabilizing effect on the HD structure.

The R5Q mutation replaces the guanido group of Arg5, which makes extensive electrostatic and hydrogen bonding interactions in the minor groove, with a shorter side chain that would fail to make similar contacts, resulting in a lower binding affinity for DNA. Likewise, the Y25S mutation would abolish the conserved hydrogen bond between the hydroxyl group of the tyrosine and the sugar−phosphate backbone. Replacement of the Val47 isopropyl group with the bulkier phenyl ring in the V47F mutation would result in steric hindrance with T4 and abolish DNA binding, as described previously (*37*). Mutation of Arg53 to glycine, histidine, or cysteine would eliminate the hydrogen bonds from the $N^{\eta 2}$ and N^{ε} atoms to G6 and lead to destabilization of a highly conserved interaction with DNA. Indeed, the R53G mutation completely abrogates DNA binding (*37*).

Conclusions. This work describes the first high-resolution structure of a wild-type HD from a PAX family member bound to a P2 DNA element and elucidates the structural basis for specificity and affinity of this interaction at the atomic level. The PAX3 HD binds to the P2 site without significant bending of the DNA. The Ser50 residues of the

dimeric HDs interact with G6 and G6′, thereby providing the structural basis for the preference of a CG sequence between the inverted TAAT motifs. Most of the PAX3 HD−DNA interactions involve intervening water molecules, suggesting that significant conformational changes in the DNA can be accommodated while still maintaining the base-specific interactions. Importantly, this structure provides a mechanistic explanation of how Waardenburg syndrome-associated missense mutations lead to destabilization of the HD fold or affect its interaction with DNA.

ACKNOWLEDGMENT

We express our deep appreciation to Mortimer B. Zuckerman for his generous support that made this work possible. We also thank Daniel Panne and Aneel Aggarwal for useful discussions and the staff at the NSLS for assistance during data collection.

REFERENCES

1. Buckingham, M., and Relaix, F. (2007) The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annu. Rev. Cell Dev. Biol. 23*, 645–673.
2. Lang, D., Powell, S. K., Plummer, R. S., Young, K. P., and Ruggeri, B. A. (2007) PAX genes: Roles in development, pathophysiology, and cancer. *Biochem. Pharmacol. 73*, 1–14.
3. Tassabehji, M., Read, A. P., Newton, V. E., Harris, R., Balling, R., Gruss, P., and Strachan, T. (1992) Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature 355*, 635–636.
4. Baldwin, C. T., Hoth, C. F., Amos, J. A., da-Silva, E. O., and Milunsky, A. (1992) An exonic mutation in the HuP2 paired domain gene causes Waardenburg's syndrome. *Nature 355*, 637–638.
5. Tassabehji, M., Newton, V. E., Liu, X. Z., Brady, A., Donnai, D., Krajewska-Walasek, M., Murday, V., Norman, A., Obersztyn, E., Reardon, W., et al. (1995) The mutational spectrum in Waardenburg syndrome. *Hum. Mol. Genet. 4*, 2131–2137.
6. Lalwani, A. K., Brister, J. R., Fex, J., Grundfast, K. M., Ploplis, B., San Agustin, T. B., and Wilcox, E. R. (1995) Further elucidation of the genomic structure of PAX3, and identification of two different point mutations within the PAX3 homeobox that cause Waardenburg syndrome type 1 in two families. *Am. J. Hum. Genet. 56*, 75–83.
7. DeStefano, A. L., Cupples, L. A., Arnos, K. S., Asher, J. H., Baldwin, C. T., Blanton, S., Carey, M. L., da Silva, E. O., Friedman, T. B., Greenberg, J., et al. (1998) Correlation between Waardenburg syndrome phenotype and genotype in a population of individuals with identified PAX3 mutations. *Hum. Genet. 102*, 499–506.
8. Nakamura, M., Ishikawa, O., and Tokura, Y. (2008) A novel missense mutation in the PAX3 gene in a case of Waardenburg syndrome type I. *J. Eur. Acad. Dermatol. Venereol.* (in press).
9. Kozawa, M., Kondo, H., Tahira, T., Hayashi, K., and Uchio, E. (2008) Novel mutation in PAX3 gene in Waardenburg syndrome accompanied by unilateral macular degeneration. *Eye* (in press).
10. Barr, F. G. (2001) Gene fusions involving PAX and FOX family members in alveolar rhabdomyosarcoma. *Oncogene 20*, 5736–5746.
11. Gehring, W. J., Affolter, M., and Bürglin, T. (1994) Homeodomain proteins. *Annu. Rev. Biochem. 63*, 487–526.
12. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell 67*, 517–528.
13. Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W., and Wüthrich, K. (1993) Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain−DNA complex. *J. Mol. Biol. 234*, 1084–1093.
14. 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.
15. 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.
16. 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.
17. 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.
18. Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999) Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature 397*, 714–719.
19. 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.
20. Gruschus, J. M., Tsao, D. H., Wang, L. H., Nirenberg, M., and Ferretti, J. A. (1999) The three-dimensional structure of the vnd/NK-2 homeodomain-DNA complex by NMR spectroscopy. *J. Mol. Biol. 289*, 529–545.
21. Hovde, S., Abate-Shen, C., and Geiger, J. H. (2001) Crystal structure of the Msx-1 homeodomain/DNA complex. *Biochemistry 40*, 12013–12021.
22. Chaney, B. A., Clark-Baldwin, K., Dave, V., Ma, J., and Rance, M. (2005) Solution structure of the K50 class homeodomain PITX2 bound to DNA and implications for mutations that cause Rieger syndrome. *Biochemistry 44*, 7497–7511.
23. Baird-Titus, J. M., Clark-Baldwin, K., Dave, V., Caperelli, C. A., Ma, J., and Rance, M. (2006) The solution structure of the native K50 Bicoid homeodomain bound to the consensus TAATCC DNA-binding site. *J. Mol. Biol. 356*, 1137–1151.
24. Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A., Jacob, V., Aggarwal, A. K., Honig, B., and Mann, R. S. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell 131*, 530–543.
25. Longo, A., Guanga, G. P., and Rose, R. B. (2007) Structural basis for induced fit mechanisms in DNA recognition by the Pdx1 homeodomain. *Biochemistry 46*, 2948–2957.
26. Wilson, D., Sheng, G., Lecuit, T., Dostatni, N., and Desplan, C. (1993) Cooperative dimerization of paired class homeodomains on DNA. *Genes Dev. 7*, 2120–2134.
27. 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.
28. 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.
29. Hiremath, C. N., and Ladias, J. A. A. (1998) Expression and purification of recombinant hRPABC25, hRPABC17, and hRPABC14.4, three essential subunits of human RNA polymerases I, II, and III. *Protein Expression Purif. 13*, 198–204.
30. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307–326.
31. Vagin, A., and Teplyakov, A. (1997) MOLREP: An Automated Program for Molecular Replacement. *J. Appl. Crystallogr. 30*, 1022–1025.
32. 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, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D54*, 905–921.
33. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol. 6*, 458–463.
34. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
35. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240–255.
36. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res. 35*, W375–W383.
37. Fortin, A. S., Underhill, D. A., and Gros, P. (1997) Reciprocal effect of Waardenburg syndrome mutations on DNA binding by

the Pax-3 paired domain and homeodomain. *Hum. Mol. Genet. 6*, 1781–1790.

38. 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.
39. Esnouf, R. M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics Modell. 15*, 132–134.
40. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins 11*, 281–296.
41. 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.
42. Lu, X. J., and Olson, W. K. (2008) 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc. 3*, 1213–1227.
43. Fenn, T. D., Ringe, D., and Petsko, G. A. (2003) Povscript+: A program for model and data visualization using persistence of vision ray-tracing. *J. Appl. Crystallogr. 36*, 944–947.

BI802052Y