Published in final edited form as:
Biochemistry. 2012 August 14; 51(32): 6312–6319. doi:10.1021/bi300849c.

Crystal Structure of the Human NKX2.5 Homeodomain in Complex with DNA Target

Lagnajeet Pradhan[†], **Caroli Genis**[‡], **Peyton Scone**[‡], **Ellen O. Weinberg**[§], **Hideko Kasahara**[⊥], and **Hyun-Joo Nam**[†,‡,*]

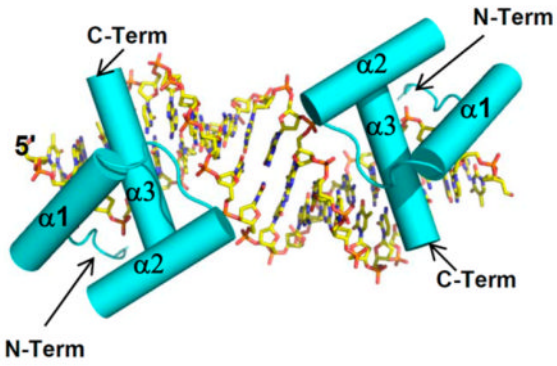
[†]Department of Bioengineering, University of Texas at Dallas, Richardson, Texas 75080, United States

[‡]Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610, United States

[⊥]Department of Physiology and Functional Genomics, University of Florida, Gainesville, Florida 32610, United States

[§]Cardiovascular Research, Boston University Medical Center, Boston, Massachusetts 02118, United States

Abstract



NKX2.5 is a homeodomain containing transcription factor regulating cardiac formation and function, and its mutations are linked to congenital heart disease. Here we provide the first report of the crystal structure of the NKX2.5 homeodomain in complex with double-stranded DNA of its endogenous target, locating within the proximal promoter –242 site of the atrial natriuretic factor gene. The crystal structure, determined at 1.8 Å resolution, demonstrates that NKX2.5 homeodomains occupy both DNA binding sites separated by five nucleotides without physical interaction between themselves. The two homeodomains show identical conformation despite the differences in the DNA sequences they bind, and no significant bending of the DNA was

[*]Corresponding Author: Address: University of Texas at Dallas, 800 W Campbell Road, RL10, Richardson, TX 75080. Telephone: (972) 883-5786. hnam@utdallas.edu.

Accession Codes

Atomic coordinates and the structural factors are deposited in the Protein Data Bank as entry 3RKQ.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

The authors declare no competing financial interest.

observed. Tyr54, absolutely conserved in NK2 family proteins, mediates sequence-specific interaction with the TAAG motif. This high resolution crystal structure of NKX2.5 protein provides a detailed picture of protein and DNA interactions, which allows us to predict DNA binding of mutants identified in human patients.

Congenital heart diseases (CHD) are the most common types of birth defects affecting approximately 40 000 infants in the United States each year. Multiple factors, both genetic and environmental, are implicated in its pathology (reviewed in ref 1). Molecular and genetic analyses revealed that mutations in transcription factors such as Nkx2.5, Tbx5, and Gata4 are linked to CHDs (refs 2–4; these studies are reviewed in ref 5). Mutations in the cardiac transcription factor, NKX2.5, in particular, are identified in one to a few percent of CHD patients and have been associated with atrial and ventricular septal defects and atrioventricular (AV) conduction block.[6]

NKX2.5 was first discovered as a homologue of *tinman*, a *Drosophila* cardiac transcription factor. It shows high expression in early heart progenitor cells and is involved in cardiac development and maturation (reviewed in ref 7). While heterozygous mutations of NKX2.5 are linked to CHDs, loss of NKX2.5 results in embryonic lethality in murine models.[8,9] Expression of NKX2.5 persists through adulthood, and conditional knockout experiments showed that, in addition to embryonic expression required for normal heart development, NKX2.5 is needed for maintenance of the conduction system in perinatal mice.[10,11]

The NKX2.5 protein is a member of the NK2-family of transcription factors and contains a homeodomain (HD), an evolutionarily conserved protein that recognizes and binds to a specific DNA sequence.[12] In addition to the HD, NKX2.5 (324 amino acids long in human) contains N- and C-terminal regulatory domains (Figure 1A). The HD is centrally located at amino acid positions 138–197 and is involved in nuclear translocation and interaction with other transcription factors as well as DNA binding.[3,13–17] In this report, the amino acid residues of HD are numbered based on the conventional HD numbering scheme by which residues 138–197 of NKX2.5 correspond to 1–60 of the HD (Figure 1B). Of the 20 missense point mutations in NKX2.5 identified to date, 9 are located in the 60 amino acid long HD,[3,18–21] underscoring its functional importance (Figure 1B).

The HD of NKX2.5 protein is highly homologous to that of other NK class protein members (55–90% homology) and binds to a consensus DNA motif, AAGTG (Figure 1A,B). Recent studies using high-throughput microarray platforms extended the recognition profile of NKX2.5 to include TT at its 5′ terminal (TTAAGTG).[22] A well-known NKX2.5 downstream target gene, atrial natriuretic factor (ANF), contains several potential NKX2.5 binding sites in its promoter region including –242 from the transcription start site (ANF-242). The ANF-242 region contains two palindromic NKX2.5 binding motifs separated by five nucleotides (Figure 1C).[14] Occupation of both binding sites by NKX2.5 *in vitro* has been previously reported, and NKX2.5 functions as a homodimer at this region. Homodimeric interaction of NKX2.5 was shown to be mediated through its C-terminal regulatory domain.[16] Using chromatin-immunoprecipitation-based techniques, the proximal promoter region, which includes ANF-242, has been confirmed as an *in vivo* binding site of NKX2.5.[23]

To understand the molecular basis of the pathological mechanisms of the NKX2.5 mutations manifested in CHDs, we have determined the three-dimensional structure of the NKX2.5 HD in complex with the double-stranded DNA fragments of ANF-242. Among various NK2 class proteins, the NMR structures of *Drosophila vnd*/NK2 HD and rat NKX2.1 HD are the only three-dimensional structures determined to date,[24,25] and currently no crystal structure of NK family protein is available. The high resolution (1.8 Å) structure of NKX2.5 HD

reported here provides a detailed picture of protein and DNA interactions. In addition, disruptions of DNA binding in the CHD related mutants were analyzed using structural models of mutants built based on the crystal structure of the NKX2.5 HD.

MATERIALS AND METHODS

Protein Expression, Purification, and Crystallization of NKX2.5HD/DNA

Construction of the *E. coli* expression plasmid and purification of NKX2.5 HD was previously described.[26] Briefly, an NKX2.5 HD construct expressing residues 137–197 with oxidizable Cys193 (56 of HD) substituted with Ser was generated. The NKX2.5 HD (C56S) protein was expressed as a maltose-binding protein (MBP) fusion protein using pMAL-c2 expression vector (New England Biolabs) and was first purified by affinity chromatography using an amylose-agarose column (New England Biolabs). The MBP-tag was removed by TEV protease, and the protein was further purified using a cation exchange column. Purified NKX2.5 HD protein was mixed with the 19 base-pair long DNA oligonucleotide (TGAAGTGGGGGCCACTTGA/TCAAGTGGCCCCCACTTCA), and the protein/DNA complex was purified using size exclusion chromatography. The protein/DNA complex solution was concentrated to the final protein concentration of 10 mg/mL.

Among several forms of NKX2.5 HD/DNA crystals, the best diffracting crystals were grown by vapor diffusion in hanging drops containing 1 μL of 50 mM Tris pH 7.0, 5 mM $MgCl_2$, 15% polyethylene glycol monomethylether 550, and 1 μL of NKX2.5 HD(C56S)/DNA solution.

X-ray Data Collection

For data collection, crystals were flash-frozen in cryosolution containing 50 mM Tris pH 7.0, 5 mM $MgCl_2$, 15% polyethylene glycol monomethyl ether 550, and 30% 2-methyl-2,4-pentanediol. The X-ray diffraction data were collected at various synchrotron sources including the 22-ID beamline of the South East Regional Collaborative Access Team (SER-CAT) facility at the Advanced Photon Source, Argonne National Laboratory, the A-1 beamline of Cornell High Energy Synchrotron Source at Cornell University, and the X-29 beamline of National Synchrotron Light Source at Brookhaven National Laboratory. The data were indexed and processed using HKL2000.[27]

Structure Determination

The long, rod-shaped crystals of the NKX2.5 HD (C56S)/DNA complex diffracted to 1.8 Å resolution at APS 22-ID beamline in Argonne National Laboratory synchrotron sources. The crystal belongs to the hexagonal space group $P6_5$ with high merohedral twinning fractions (0.47). On the basis of the unit cell dimensions and the molecular mass of the NKX2.5 HD(C56S)/DNA complex, two HD domain and a 19-bp double-stranded DNA were expected in an asymmetric unit with a solvent content of 58% (V_M = 2.61 $Å^3$ Da^{-1}).[28] The structure was determined by molecular replacement using the program, Phaser[29] and the crystal structure of the MSX-1 transcription factor HD/DNA complex (PDB ID 1IG7)[30] as a search model. The DNA in the MSX-1 model was trimmed to 8 base pairs to exclude any flexible regions in the structure. The initial search revealed two unambiguous solutions for the HD/DNA complexes. When the model was positioned in a unit cell, a pair of DNA molecules related by noncrystallographic symmetry formed a continuous helix. Subsequent iterative rounds of structural refinement were carried out using TWIN options of rigid body, simulated annealing, conjugated-gradient energy minimization, and B-factor refinement of the Crystallography and NMR System program suite (CNS),[31] followed by manual rebuilding using the program COOT.[32] The final model was validated using Molprobity.[33]

Building Models for the CHD Related NKX2.5 HD Mutants

A total of five NKX2.5 HD mutants linked to CHDs, Q50H, N51K, R52G, R53H, and Y54C were modeled based on NKX2.5 HD (C56S)/DNA structure using the program *interface model*.[34] The wild type structure was energy-minimized using the model minimization option of the CNS program.[31] Mutations were introduced to the energy-minimized model, and the lowest energy conformational rotamers of residues near the mutated residues were selected using the Monte Carlo procedure described in the previous report.[34]

Calculation of Intraprotein and Protein–DNA Interaction Energies for CHD-Related Mutants

Free energies for intraprotein and protein–DNA interaction of the NKX2.5/DNA complex were calculated using the NKX2.5 HD model and mutant models built as described in the previous section. The program, *interface model*, was used for the calculation.[34] The energy was computed using the all-atom energy function based on the AMBER param98 force-field[35] combined with additional hydrogen bond terms.[34]

RESULTS AND DISCUSSION

Overview of NKX2.5HD/ANF-242 Complex

The crystal structure of NKX2.5 HD C56S/ANF-242 DNA complex was solved at 1.8 Å resolution. In order to prevent oxidization of the protein, Cys56 to Ser (C56S) mutation was introduced, which will not affect binding, and the mutant protein showed stable folding characteristics.[36] Data collection and final refinement statistics are summarized in Table 1.

The model includes 19 base-pair ANF-242 DNA and two molecules of NKX2.5 HD comprising residues 137–195, which correspond to 1–57 in HD numbering with an additional N-terminal amino acid (Figures 2A and 1B). Two molecules of NKX2.5 HD show an identical structure with an overall rmsd of Cα atoms of 0.22 Å. The HD adopts a canonical HD fold formed by three α-helices and the N-terminal arm (residues 1–9), located upstream of helix α1 (Figure 2A). Two N-terminal helices, α1 (residue 10–22) and α2 (residues 28–38), run antiparallel to each other and position α3 (residues 42–58) to the major groove of the DNA.

The NK2 family HD is well folded even in the absence of DNA[36,37] probably because of a well-formed hydrophobic core composed of Phe8 in the N-terminus; Leu16 and Phe20 from α1; Leu26 from α1-α2 loop; Leu34 and Leu38 from α2; Leu40 from α2-α3 loop; Val45, Trp48, Phe49 and Arg52 from α3 (Figure 2B). All helices are further stabilized by "N-capping" similar to that previously observed in the NMR structure of *vnd*/NK2.[24] In α1, the O_γ of Ser9 and backbone amide of Gln12, and O_ε of Gln12 and backbone amide of Ser9 make hydrogen bonds thus reducing the helix dipole. Similar double hydrogen bonds were observed in α3 between O_γ and the backbone amide of Thr41 and the backbone amide and O_ε Gln44, respectively. The O_γ of Ser27 at the N-terminus of α2, on the other hand, does not make a hydrogen bond with the backbone amide of Glu30, but O_ε of Gln30 hydrogen bonds with the backbone amide of Ser27. Ser27 was shown to be phosphorylated by casein kinase II *in vivo*,[13] and its O_γ is readily accessible from the surface for such a modification. N-capping plays an important role in stabilizing the HD, and the residues involved in N-capping are highly conserved throughout NK2 family proteins (Figure 1B).

The ANF-242 DNA structure shows three distinct conformational segments (Figure 1C). The two NK binding DNA motifs (TGAAGTG/TCAAGAG) show a primarily B-form structure with values for mean rise of 3.3–3.4 Å and twist of 34–35° per base pair. The five base-pair spacer region shows an intermediate of A- and B-form DNA with a mean rise of

3.1 Å and twist of 33 Å. Both NK binding motifs show overall bending of 3°, considerably smaller than bend angles observed in other HD-DNA structures.[38,39]

The ANF-242 DNA contains two NK binding motifs, (TGAAGTG/TCAAGAG), in a palindromic arrangement separated by a five-base pair long spacer region (Figure 1C). NKX2.5 HDs occupy the NK motifs on the same face of the DNA with 180° rotational symmetry as was seen in Figure 2A. NKX2.5 shows synergistic and cooperative binding as a homodimeric complex, and homodimeric interactions were shown to be mediated through the C-terminal domain.[16] Our structure lacking the C-terminal domain shows no physical interactions between the two HDs. Previous experiments with *Tinman*, an ortholog of NKX2.5 in *Drosophila*, showed that the homodimeric protein functions optimally when a six base pair long spacer is placed between the two NK binding motifs. Shorter (3 base pairs) or longer (10 base pairs) spacers reduced transcriptional activation by *Tinman*.[40] NKX2.5 HDs bound at the palindromic sites are positioned at the same side on a helical DNA with a half a helical turn spacer (5–6 base pairs), and this spatial arrangement may be important for homodimeric interactions mediated by the C-terminal domain.

In addition to the homodimeric interaction, NKX2.5 is involved in combinatorial interactions with other core cardiac transcription factors to regulate heart development. The HD domain is involved in interactions with TBX5, one of the master cardiac transcription factors.[15,16] Since the interprotein interaction can vary among NK2 family members, less conserved residues positioned at the surface of the protein function are expected to serve as sites for such interactions. Residues Gln10, Tyr14, Lys21, Gln33, Val37, and Lys39 belong to this category.

NKX2.5 HD-DNA Interactions

Interactions between the DNA and HD are mediated through residues from three regions of the HD: N-terminal extension, the loop connecting α1 and α2, and helix α3 (Figure 3A,B). Arg5, Val6, and Phe8 of the N-terminal extension interact with the DNA through the minor groove (Figure 3C). Among them, only Arg5 makes contacts with the DNA base. It makes hydrogen bonds to T1 and G2 bases and interacts with A19′ and C18′ bases through water molecules. The Arg5 of the second molecule in the unit cell also interacts with T1′ and C2′ bases. Therefore, the first two bases of the NK element (TG/C/TAAGTG) are recognized by Arg5. Arg5 is highly conserved among HD members (Figure 1B), and in other HDs such as Antp, en, and oct1, the conserved Arg5 interacts with various bases at these positions: CT, GT, GT, and CA, respectively.[41–46] The flexible side chain of Arg5 may accommodate various conformations of DNA and may not be a determining factor for DNA sequence specificity. Regardless, high-throughput microarray studies report that TT is preferred at this position by NKX2.5.[22]

Backbone amide and carboxyl groups of Val6 and Phe8 interact with DNA backbone phosphate (A3 and A4) through water molecules. Lys3, which was shown to contact the DNA in the *vnd*/NK2 NMR structure, does not participate in DNA interaction.[24] The N-terminus of NKX2.5 HD is involved in crystallographic contact in our structure, and the Arg1, Arg2, and Lys3 make extensive contacts with neighboring DNA.

Tyr25, Leu26 of α1-α2 loop and Arg31 of α2 interact with DNA backbone phosphate groups. The hydroxyl group of Tyr25 makes a hydrogen bond with the O2P atom of C13′, and the backbone amide group of Leu26 with O1P of C12′ by a water mediated interaction (Figure 3A,B). The N_ε of Arg31 interacts with the C11′ O2P atom through a salt bridge. Tyr25 and Arg31 serve as interaction points at the 3′ side of the NK2 recognition motif, and both residues are highly conserved throughout the HD proteins.

The majority of the HD-DNA interactions are mediated by helix α3 through residues Gln44, Lys46, Ile47, Gln50, Asn51, Arg53, Tyr54, Lys55, and Lys57. Figure 3D shows residues on the N-terminal side of the helix α3 (amino acid residues 44–49). The ε-amide group of the Gln44 makes a hydrogen bond with the phosphate of A4, and its main chain carbonyl group also makes water-mediated hydrogen bonds with the A3 and A4 phosphates. Lys46 interacts with phosphate groups of C11′ and C12′ through the ε-amide group and the main chain carbonyl group, respectively. The side chain of Trp48 folds back into the hydrophobic core but $N_{\varepsilon}1$ of the indole ring participates in a water-mediated hydrogen bond with the A3 phosphate group. Ile47 does not make hydrogen bonds but its side chain is within van der Waals interaction range of the A3 base.

Interactions between the DNA and the C-terminal segment of helix α (amino acid residues 50–57) is shown in Figure 3E. Residues Gln50, Arg53, and Lys57 are involved in interactions with backbone phosphate groups: Gln50 to C12′ and C13′ through the terminal and the main chain carbonyl group, respectively; Arg53 also to C12′ and C13′ through its guanidine group; Lys57 to A14 using the ε-amino group. Gln50, in addition, makes van der Waals contacts with the C13′ base. Asn51 and Tyr54 make extensive interactions with DNA bases. The δ-carbonyl group of Asn51 hydrogen-bonds with the N6 atom of A4 and also interacts with the C15′ N4 atom through a water molecule. The δ-amide group of Asn51 interacts with the A4 base (N7 atom) and a phosphate group through a water-mediated hydrogen bond (Figure 3E).

Tyr54 is uniquely conserved in all NK2 family proteins but not in other HDs. It is not involved in hydrogen bonding but within van der Waals interaction ranges of the DNA. It bonds with the bases of A14′ and C15′. Figure 3F,G shows Tyr54 and interacting DNA residues. Tight shape complementarity can be observed between the Tyr54 side chain and the bases of A14′ and C15′. The DNA residues, A14′ and C15′, are complementary to the fifth and sixth residues of the NK2 element (TG/C/TAAGTG). The majority of other HD proteins contain Ala or Met at this position and recognize DNA motifs containing TAAT.[41–48] NK2 family proteins uniquely show high affinity to the TAAGT sequence, and mutation to Met at this position reduces affinity by 10-fold.[49] This shows that Tyr54, which is invariant in NK2 family proteins, is responsible for the NK2-specific DNA sequence recognition.

The DNA binding interaction described here is in agreement with the NMR structure of *vnd*/NK2 HD in complex with DNA.[24,50] In the NMR structural analyses, Lys3, Arg5, Ile47, Gln50, Asn51, and Tyr54 were shown to be involved in DNA base contact, and in our crystal structure, these residues also make DNA base contacts except Lys3, which is involved in crystallographic interaction with neighboring DNA.[24,50] In addition, the long-lived water molecules observed in NK2/DNA complex structures are also present in our structure which includes W1, W2, W3, W7, W9, and W13 shown in Figure 3C–E.[51] All the residues making DNA contacts are well conserved in NK2 family proteins and their mutations result in pathological phenotypes (Figure 1B).

Congenital Heart Disease Related Mutations in NKX2.5 HD

Missense mutations of NKX2.5 HD linked to cardiac pathology are summarized in Figure 1B (reviewed in ref 52). To understand the effect of these mutations, free-energy differences of several selected mutants from the wild type were calculated and compared.[34] The mutant models were generated based on the NKX2.5 HD/ANF-242 crystal structure, energy minimized, and used for the calculation. Free energies of intraprotein interaction and HD-DNA interaction were computed and the relative ΔG from the wild type was calculated. Solvent molecules were not included in the calculation for simplicity. Table 2 shows varying degrees of ΔG in the selected mutants.

The crystal structure shows that the side chains of Arg5, Gln50, Asn51, Arg53, and Tyr54 make direct contacts with the DNA (Figure 3C–G), and mutations in these residues such as R5C, Q50H, N51K, R53H, R53C, and Y54C are likely to affect DNA binding. Among them, ΔG of N51K, R53H, and Y54C but not Q50H showed a decrease in DNA binding (increase in ΔG) (Table 2). Q50H shows basically the same binding as the wild type. In the mutant model, the hydrogen bond between the N_{ε} of Gln50 and O6 of G5 in wild type protein is replaced with an equivalent hydrogen bond between N_{ε} of His50 and the phosphate group of C12′. The more drastic energy change was observed with intraprotein interaction than in DNA binding with N51K, R53H, and Q50H mutants. In the Q50H mutant, unlike DNA binding, folding is negatively affected by the mutation, resulting in an increase in overall ΔG (Table 2). The Asn51 and Arg53 are involved in DNA binding, but their mutations also reduce intraprotein interactions (Table 2). The energy-minimized model of N51K shows that, to accommodate a longer Lys at position 51, the side chain of Lys55 assumes an alternate conformation. Compared to the wild type, the protein ends up losing two intraprotein hydrogen bonds in addition to several van der Waals interactions (Figure 4A). Therefore, the bulk of energy loss in N51K stems from reduced intraprotein interaction. R53H shows a similar trend. The side chain of Arg53 makes intraprotein interactions, which include a salt bridge with the Arg24 carbonyl group and amino–aromatic interactions with Tyr25, which are lost in the R53H mutant (Figure 4B). Disruption in intraprotein interactions causes an increase in free energy in these mutants compared to the wild type.

Members of the next group of mutants such as L34P, T41M, W47L, and R52G do not directly contact DNA. These residues are, however, important in stabilizing the protein structure, and thus their mutations significantly change overall energy. Thr41 participates in "N-capping" of helix $\alpha 3$, and T41M is expected to destabilize the helix. Leu34, Trp48, and Arg52 make up parts of the hydrophobic core and their mutation will destabilize the whole HD, which is detrimental to its function. The effect of core-destabilizing mutants seems more serious than mutations that affect only DNA binding. R52G mutations show the highest overall energy difference among the mutants selected for energy calculation (Table 2). In wild type *vnd*/NK2, this position is occupied by His. Using the circular dichroism technique, H52R mutation in *vnd*/NK2 was shown to increase thermal stability of the protein.[49] This demonstrates that Arg52 functions to stabilize the protein.

DNA binding of the mutant NKX2.5 proteins, R5C, L34P, T41M, Q50H, N51K, R52G, R53H, and Y54C, has been previously analyzed using electrophoretic mobility shift assay (EMSA).[21,53] All the mutants showed a range of reduction in DNA binding. R5C, T41M, Q50H, and Y54C showed ~10–10^2 fold reduction compared to the wild type, N51G, ~10^3 fold reduction, and for L34P, R52G and R53H, DNA binding was not detectable.[21,53] These results are in agreement with our ΔG calculation data and structure based analyses.

The energy difference between the wild type and a mutant can be expressed as $\Delta G_M - \Delta G_W = -RT \ln (K^{eq}{}_M / K^{eq}{}_W)$ where ΔG_M is free energy of mutant, ΔG_W, free energy of wild type, R, the ideal gas constant, T, absolute temperature, and K_{eq}, equilibrium constant, respectively. On the basis of this, ~1–3.5 kcal M^{-1} differences in the energy translates to ~10–10^2 fold differences in K^{eq} (R = 1.99 cal K^{-1} M^{-1}, K = 310 K). The EMSA results show that Q50H and Y54C with 1.1 and 3.49 kcal $M^{-1}$of ΔG differences also show ~10–10^2 fold decrease in binding efficiency. On the other hand, the R52G and R53H mutants show large free energy changes by the energy calculation (Table 2), and correspondingly the EMSA results put R52G and R53H among the poorest DNA binders.[21,53]

The homeodomain is an evolutionarily conserved DNA binding domain. Through evolution, this compact protein is optimized for its function, and each residue is strategically placed for DNA binding and multiple layers of regulation. Many residues involved in DNA binding

also tend to be involved in conformational stabilization, and a mutation in these residues can be detrimental not only for DNA binding but also for protein stability as was seen with pathological NKX2.5 mutants (Table 2).

CONCLUSIONS

We report the first high-resolution crystal structure of NK2 family HD bound to an *in vivo* target DNA. The structure provides the atomic basis for specificity of protein and DNA interactions. As in previously reported HD structures, the N-terminal extension and helix $\alpha 3$ make the majority of the DNA–protein interactions, but the NK2-specific sequence recognition seems mediated by conserved Tyr54, which showed tight shape complementarity with the DNA sequence. Although two HDs were bound to both palindromic sites, physical interactions between the two HDs were not observed. The C-terminal domain, absent in our structure, was shown to be important for homodimeric interactions, and the spatial organizations of the two HDs bound to the DNA may enable such dimeric interactions. Furthermore, this structure provides the structural foundation that allows us to predict disruptions in protein–DNA binding caused by pathological mutations identified in human CHD patients.

Acknowledgments

Funding

This work was supported by grants to H.-J. N. from American Heart Association (0535161B) and to H.K. from National Institutes of Health (HL081577). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38, and CHESS is supported by NSF Grant DMR-0225180 and NIH/NCRR Grant RR-01646.

We thank Drs. Mavis Agbandje-McKenna, and L. Govindasamy for their advice and help, and the staff at the SER-CAT 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory, the staff at CHESS, and the staff at the Brookhaven National Laboratory for assistance during X-ray data collection.

ABBREVIATIONS USED

A	adenine
ANF	atrial natriuretic factor
C	cytosine
CHD	congenital heart disease
EMSA	electrophoretic mobility shift assay
G	guanine
HD	homeodomain
T	thymidine

References

1. Mitchell ME, Sander TL, Klinkner DB, Tomita-Mitchell A. The molecular basis of congenital heart disease. Semin Thorac Cardiovasc Surg. 2007; 19:228–237. [PubMed: 17983950]
2. Garg V, Kathiriya I, Barnes R, Schluterman M, King I, Butler C, Rothrock C, Eapen R, Hirayama-Yamada K, Joo K, Matsuoka R, Cohen J, Srivastava D. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature. 2003; 424:443–447. [PubMed: 12845333]

3. Schott J, Benson D, Basson C, Pease W, Silberbach G, Moak J, Maron B, Seidman C, Seidman J. Congenital heart disease caused by mutations in the transcription factor NKX2–5. Science. 1998; 281:108–111. [PubMed: 9651244]
4. Basson CT, Huang T, Lin RC, Bachinsky DR, Weremowicz S, Vaglio A, Bruzzone R, Quadrelli R, Lerone M, Romeo G, Silengo M, Pereira A, Krieger J, Mesquita SF, Kamisago M, Morton CC, Pierpont ME, Muller CW, Seidman JG, Seidman CE. Different TBX5 interactions in heart and limb defined by Holt-Oram syndrome mutations. Proc Natl Acad Sci U S A. 1999; 96:2919–2924. [PubMed: 10077612]
5. Nemer M. Genetic insights into normal and abnormal heart development. Cardiovasc Pathol. 2008; 17:48–54. [PubMed: 18160060]
6. Harvey R, Lai D, Elliott D, Biben C, Solloway M, Prall O, Stennard F, Schindeler A, Groves N, Lavulo L, Hyun C, Yeoh T, Costa M, Furtado M, Kirk E. Homeodomain factor Nkx2–5 in heart development and disease. Cold Spring Harbor Symp Quant Biol. 2002; 67:107–114. [PubMed: 12858530]
7. Akazawa H, Komuro I. Cardiac transcription factor Csx/Nkx2–5: Its role in cardiac development and diseases. Pharmacol Ther. 2005; 107:252–268. [PubMed: 15925411]
8. Lyons I, Parsons LM, Hartley L, Li R, Andrews JE, Robb L, Harvey RP. Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2–5. Genes Dev. 1995; 9:1654–1666. [PubMed: 7628699]
9. Tanaka M, Chen Z, Bartunkova S, Yamasaki N, Izumo S. The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development. Development. 1999; 126:1269–1280. [PubMed: 10021345]
10. Takeda M, Briggs LE, Wakimoto H, Marks MH, Warren SA, Lu JT, Weinberg EO, Robertson KD, Chien KR, Kasahara H. Slow progressive conduction and contraction defects in loss of Nkx2–5 mice after cardiomyocyte terminal differentiation. Lab Invest. 2009; 89:983–993. [PubMed: 19546853]
11. Briggs LE, Takeda M, Cuadra AE, Wakimoto H, Marks MH, Walker AJ, Seki T, Oh SP, Lu JT, Sumners C, Raizada MK, Horikoshi N, Weinberg EO, Yasui K, Ikeda Y, Chien KR, Kasahara H. Perinatal loss of Nkx2–5 results in rapid conduction and contraction defects. Circ Res. 2008; 103:580–590. [PubMed: 18689573]
12. Gehring WJ, Affolter M, Burglin T. Homeodomain proteins. Annu Rev Biochem. 1994; 63:487–526. [PubMed: 7979246]
13. Kasahara H, Izumo S. Identification of the in vivo casein kinase II phosphorylation site within the homeodomain of the cardiac tisue-specifying homeobox gene product Csx/Nkx2.5. Mol Cell Biol. 1999; 19:526–536. [PubMed: 9858576]
14. Durocher D, Charron F, Warren R, Schwartz R, Nemer M. The cardiac transcription factors Nkx2–5 and GATA-4 are mutual cofactors. EMBO J. 1997; 16:5687–5696. [PubMed: 9312027]
15. Hiroi Y, Kudoh S, Monzen K, Ikeda Y, Yazaki Y, Nagai R, Komuro I. Tbx5 associates with Nkx2–5 and synergistically promotes cardiomyocyte differentiation. Nat Genet. 2001; 28:276–280. [PubMed: 11431700]
16. Kasahara H, Usheva A, Ueyama T, Aoki H, Horikoshi N, Izumo S. Characterization of homo- and heterodimerization of cardiac Csx/Nkx2.5 homeoprotein. J Biol Chem. 2001; 276:4570–4580. [PubMed: 11042197]
17. Lee Y, Shioi T, Kasahara H, Jobe SM, Wiese RJ, Markham BE, Izumo S. The cardiac tissue-restricted homeobox protein Csx/Nkx2.5 physically associates with the zinc finger protein GATA4 and cooperatively activates atrial natriuretic factor gene expression. Mol Cell Biol. 1998; 18:3120–3129. [PubMed: 9584153]
18. Benson D, Silberbach G, Kavanaugh-McHugh A, Cottrill C, Zhang Y, Riggs S, Smalls O, Johnson M, Watson M, Seidman J, Seidman C, Plowden J, Kugler J. Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. J Clin Invest. 1999; 104:1567–1573. [PubMed: 10587520]
19. Goldmuntz E, Geiger E, Benson D. NKX2.5 mutations in patients with tetralogy of fallot. Circulation. 2001; 104:2565–2568. [PubMed: 11714651]

20. Gutierrez-Roelens I, Sluysmans T, Gewillig M, Devriendt K, Vikkula M. Progressive AV-block and anomalous venous return among cardiac anomalies associated with two novel missense mutations in the CSX/NKX2–5 gene. Hum Mutat. 2002; 20:75–76. [PubMed: 12112663]
21. Kasahara H, Benson DW. Biochemical analyses of eight NKX2.5 homeodomain missense mutations causing atrioventricular block and cardiac anomalies. Cardiovasc Res. 2004; 64:40–51. [PubMed: 15364612]
22. Keles S, Warren CL, Carlson CD, Ansari AZ. CSI-Tree: a regression tree approach for modeling binding properties of DNA-binding molecules based on cognate site identification (CSI) data. Nucleic Acids Res. 2008; 36:3171–3184. [PubMed: 18411210]
23. Warren SA, Terada R, Briggs LE, Cole-Jeffrey CT, Chien WM, Seki T, Weinberg EO, Yang TP, Chin MT, Bungert J, Kasahara H. Differential role of Nkx2–5 in activation of the atrial natriuretic factor gene in the developing versus failing heart. Mol Cell Biol. 2011; 31:4633–4645. [PubMed: 21930795]
24. Gruschus J, Tsao D, Wang L, Nirenberg M, Ferretti J. The three-dimensional structure of the vnd/NK-2 homeodomain- DNA complex by NMR spectroscopy. J Mol Biol. 1999; 289:529–545. [PubMed: 10356327]
25. Esposito G, Fogolari F, Damante G, Formisano S, Tell G, Leonardi A, Di Lauro R, Viglino P. Analysis of the solution structure of the homeodomain of rat thyroid transcription factor 1 by 1H-NMR spectroscopy and restrained molecular mechanics. Eur J Biochem. 1996; 241:101–113. [PubMed: 8898894]
26. Genis C, Scone P, Kasahara H, Nam HJ. Crystallization and preliminary X-ray analysis of the NKX2.5 homeodomain in complex with DNA. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2008; 64:1079–1082.
27. Otwinowsky, Z.; Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. In: Carter, JCW.; Sweet, RM., editors. Methods in Enzymology. 1997. p. 307-326.
28. Matthews BW. Solvent content of protein crystals. J Mol Biol. 1968; 33:491–497. [PubMed: 5700707]
29. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
30. Hovde S, Abate-Shen C, Geiger JH. Crystal structure of the Msx-1 homeodomain/DNA complex. Biochemistry. 2001; 40:12013–12021. [PubMed: 11580277]
31. Brünger A, Adams P, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. 1998; D54:905–921.
32. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
33. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010; 66:12–21. [PubMed: 20057044]
34. Siggers TW, Honig B. Structure-based prediction of C2H2 zinc-finger binding specificity: sensitivity to docking geometry. Nucleic Acids Res. 2007; 35:1085–1097. [PubMed: 17264128]
35. Cheatham TE 3rd, Cieplak P, Kollman PA. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. J Biomol Struct Dyn. 1999; 16:845–862. [PubMed: 10217454]
36. Fodor E, Mack JW, Maeng JS, Ju JH, Lee HS, Gruschus JM, Ferretti JA, Ginsburg A. Cardiac-specific Nkx2.5 homeodomain: conformational stability and specific DNA binding of Nkx2.5(C56S). Biochemistry. 2005; 44:12480–12490. [PubMed: 16156660]
37. Tsao DH, Gruschus JM, Wang LH, Nirenberg M, Ferretti JA. The three-dimensional solution structure of the NK-2 homeodomain from Drosophila. J Mol Biol. 1995; 251:297–307. [PubMed: 7643404]
38. Li T, Stark MR, Johnson AD, Wolberger C. Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. Science. 1995; 270:262–269. [PubMed: 7569974]

39. Xu W, Rould MA, Jun S, Desplan C, Pabo CO. Crystal structure of a paired domain-DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations. Cell. 1995; 80:639–650. [PubMed: 7867071]
40. Zaffran S, Frasch M. The homeodomain of Tinman mediates homo- and heterodimerization of NK proteins. Biochem Biophys Res Commun. 2005; 334:361–369. [PubMed: 16004970]
41. Fraenkel E, Pabo CO. Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex. Nat Struct Biol. 1998; 5:692–697. [PubMed: 9699632]
42. Billeter M, Qian YQ, Otting G, Muller M, Gehring W, Wuthrich K. Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. J Mol Biol. 1993; 234:1084–1093. [PubMed: 7903398]
43. Qian YQ, Billeter M, Otting G, Muller M, Gehring WJ, Wuthrich K. The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell. 1989; 59:573–580. [PubMed: 2572329]
44. Kissinger CR, Liu BS, Martin-Blanco E, Kornberg TB, Pabo CO. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. Cell. 1990; 63:579–590. [PubMed: 1977522]
45. Cepek KL, Chasman DI, Sharp PA. Sequence-specific DNA binding of the B-cell-specific coactivator OCA-B. Genes Dev. 1996; 10:2079–2088. [PubMed: 8769650]
46. Remenyi A, Tomilin A, Pohl E, Lins K, Philippsen A, Reinbold R, Scholer HR, Wilmanns M. Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. Mol Cell. 2001; 8:569–580. [PubMed: 11583619]
47. Jauch R, Ng CK, Saikatendu KS, Stevens RC, Kolatkar PR. Crystal structure and DNA binding of the homeodomain of the stem cell transcription factor Nanog. J Mol Biol. 2008; 376:758–770. [PubMed: 18177668]
48. Birrane G, Soni A, Ladias JA. Structural basis for DNA recognition by the human PAX3 homeodomain. Biochemistry. 2009; 48:1148–1155. [PubMed: 19199574]
49. Weiler S, Gruschus JM, Tsao DH, Yu L, Wang LH, Nirenberg M, Ferretti JA. Site-directed mutations in the vnd/NK-2 homeodomain. Basis of variations in structure and sequence-specific DNA binding. J Biol Chem. 1998; 273:10994–11000. [PubMed: 9556579]
50. Gruschus JM, Tsao DH, Wang LH, Nirenberg M, Ferretti JA. Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. Biochemistry. 1997; 36:5372–5380. [PubMed: 9154919]
51. Gruschus JM, Ferretti JA. Quantitative measurement of water diffusion lifetimes at a protein/DNA interface by NMR. J Biomol NMR. 2001; 20:111–126. [PubMed: 11495243]
52. Reamon-Buettner SM, Borlak J. NKX2–5: an update on this hypermutable homeodomain protein and its role in human congenital heart disease (CHD). Hum Mutat. 2010; 31:1185–1194. [PubMed: 20725931]
53. Kasahara H, Lee B, Schott JJ, Benson DW, Seidman JG, Seidman CE, Izumo S. Loss of function and inhibitory effects of human CSX/NKX2.5 homeoprotein mutations associated with congenital heart disease. J Clin Invest. 2000; 106:299–308. [PubMed: 10903346]

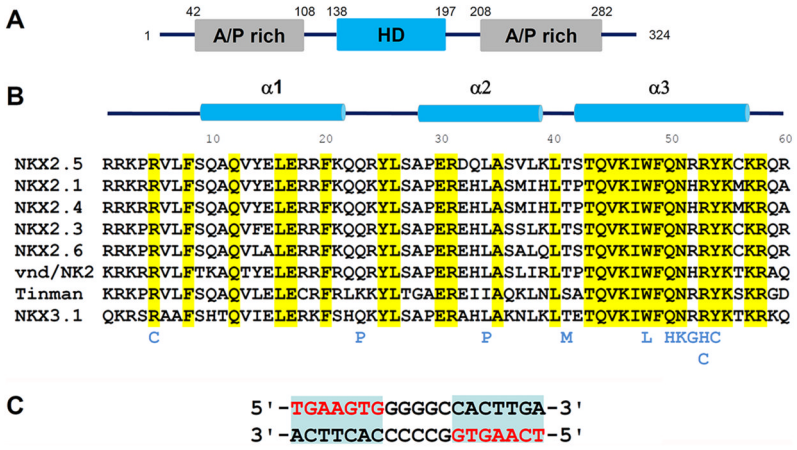


Figure 1.

The schematic diagram of NKX2.5, NK family HD sequence alignment and ANF-242 DNA sequence. (A) A schematic model of NKX2.5. Both N-terminal (1–137) and C-terminal (198–324) domains contain A/P rich sequences and HD (138–197) is centrally located. (B) Sequence alignment of NK family HD domains. The numbering is based on the conventional HD numbers where NKX2.5 138–197 corresponds to HD residue 1–60. Secondary structure elements are represented using cyan cylinders for α-helices and black lines for coils on top of the sequences. Missense mutations identified in NKX2.5 HD were shown below the sequences. (C) ANF-242 DNA sequence used for the structural study. Palindromic NK2 elements are shaded in cyan.

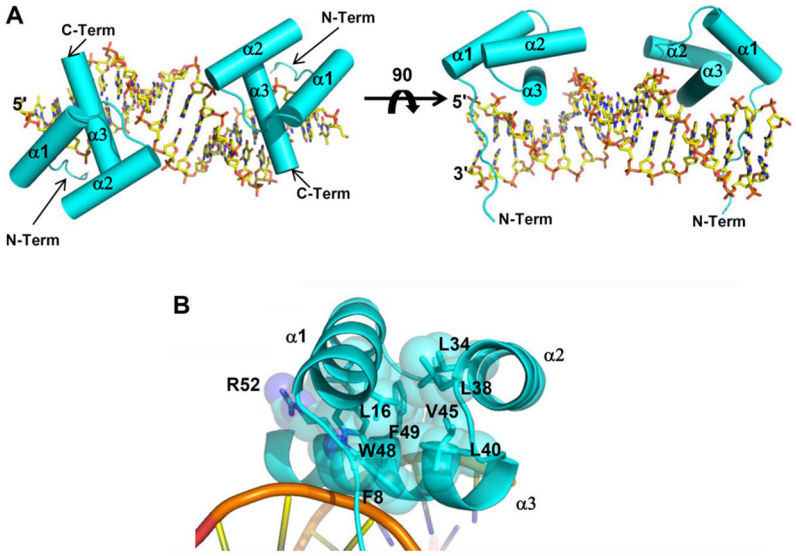


Figure 2.

Overall structure of NKX2.5 HD/ANF-242 complex. (A) Two different views are shown of the protein–DNA complex. The HD is shown in cyan, and the DNA is shown in stick representation with carbon in yellow, oxygen in red, nitrogen in blue, and phosphorus in orange. On the left diagram, both HDs are positioned on top of the DNA, and the view on the right was generated by rotating the molecule 90° along the horizontal axis. The helices and the N- and C-termini are labeled. (B) Side chains of the residues involved in formation of the hydrophobic core are shown in stick and van der Waals sphere representation. Only one HD is shown in this figure.

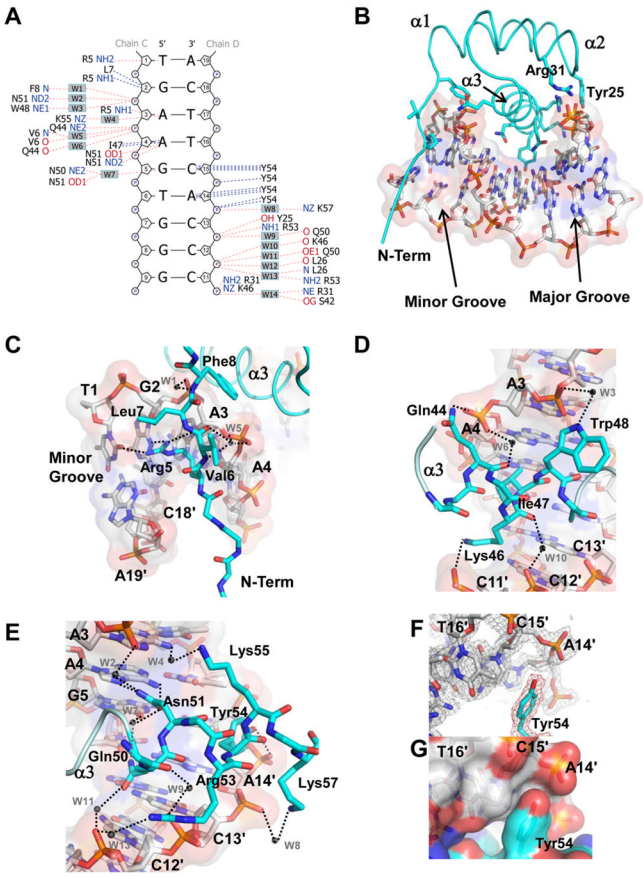


Figure 3.

NKX2.5 HD and ANF-242 DNA interaction. (A) A diagram of HD-DNA contact in a half site of ANF-242 DNA. Hydrogen bonds and van der Waals interactions are represented with solid red and dotted blue lines, respectively. Numbered water molecules are shown in cyan boxes. (B) Structure of HD monomer and the DNA. Side chains of amino acid residues involved in DNA binding are shown with carbon in cyan, oxygen in red, and nitrogen in blue. The DNA model is shown in stick representation. Three helices and the N-terminus of the HD, and the major and minor grooves of DNA are labeled. (C–E) Detailed view of HD-DNA interactions through the N-terminus (C), the N-terminal section of α3 (D) and the C-terminal section of α3 of the HD (E). DNA model is shown with carbon in white using stick and surface representation. The HD is shown in stick and ribbon representation with carbon in cyan. For all molecules, oxygen is shown in red, nitrogen in blue, and phosphorus in orange. Hydrogen bonds are depicted using dotted black lines. Water molecules are shown as gray spheres. (F, G) Tyr54 and interacting DNA region. A stick model of Tyr54 side chain is shown with $2F_o - F_c$ map (1 σ level) calculated at 1.8 Å resolution (F) and with van der Waals surface representation (G).

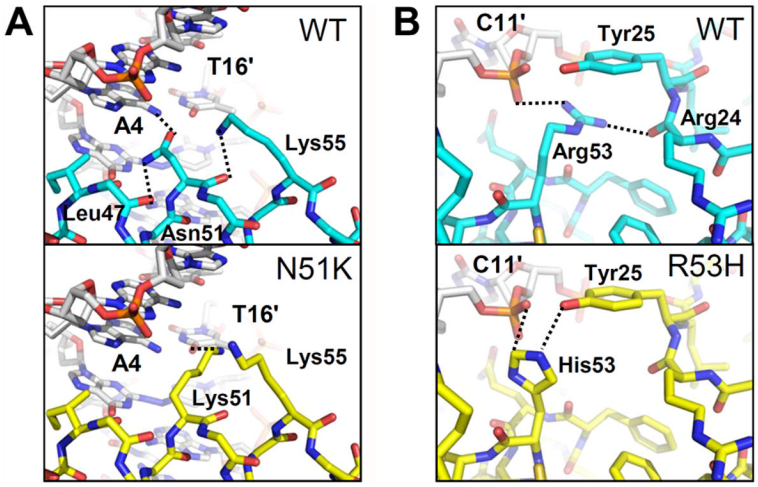


Figure 4.

Comparisons of wild type and HD mutants, N51K (A) and R53H (B), interactions. Energy minimized models of wild type, N51K, and R53H are shown in stick representation. Color scheme follows that of Figure 3 except in N51K and R53H mutants, carbon is shown in yellow. Dotted lines depict hydrogen bonds.

Table 1

Crystallographic Data and Refinement Statistics

Crystallographic Data	
space group	$P6_5$
cell parameters	$a = b = 71.45$ Å, $c = 94.33$ Å
X-ray source	22ID, APS
resolution	40–1.7 Å
observations (n)	330,129
unique reflections (completeness)	29,990 (99.4%)
R_{sym} [a]	0.065
Refinement Statistics of Current Model	
protein atoms (n)	999
resolution	40–1.8 Å
DNA atoms (n)	773
solvent atoms (n)	194
R_{cryst} (work[b]/free[c])	0.208/0.260
rmsd bond lengths[d]	0.006 Å
rmsd bond angles[d]	1.023°

[a] $R_{sym} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of an individual hkl, and $\langle I(hkl)\rangle$ is the mean intensity for all measured values of this reflection.

[b] $R_{work} = ||F_o| - |F_c|| / |F_o|$, where F_o and F_c represents the observed structure factor amplitudes and the structure factor amplitudes calculated from the atomic model, respectively.

[c] R_{free} was calculated with the 5% of randomly selected reflections excluded from the data set during refinement.

[d] rmsd, rootmean- square deviation.

Table 2

Calculated Relative Interaction Energy of NKX2.5 HD Mutants

mutations	calculated relative Δ*G* intra-protein (kcal/mol)	calculated relative Δ*G* protein–DNA interaction (kcal/mol)	calculated relative Δ*G* total (kcal/mol)
wild type	0	0	0
Q50H	3.52	−0.03	3.49
N51K	13.29	1.81	15.10
R52G	32.18	2.18	34.36
R53H	26.07	0.16	26.23
Y54C	−0.42	1.52	1.10