Probing the role of the protonation state of a minor groove-linker histidine in Exd-Hox–DNA binding

Yibei Jiang,[1] Tsu-Pei Chiu,[1] Raktim Mitra,[1] and Remo Rohs[1,2,3,4,*]

[1]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, California; [2]Department of Chemistry, University of Southern California, Los Angeles, California; [3]Department of Physics and Astronomy, University of Southern California, Los Angeles, California; and [4]Thomas Lord Department of Computer Science, University of Southern California, Los Angeles, California

ABSTRACT   DNA recognition and targeting by transcription factors (TFs) through specific binding are fundamental in biological processes. Furthermore, the histidine protonation state at the TF-DNA binding interface can significantly influence the binding mechanism of TF-DNA complexes. Nevertheless, the role of histidine in TF-DNA complexes remains underexplored. Here, we employed all-atom molecular dynamics simulations using AlphaFold2-modeled complexes based on previously solved co-crystal structures to probe the role of the His-12 residue in the Extradenticle (Exd)-Sex combs reduced (Scr)-DNA complex when binding to Scr and Ultrabithorax (Ubx) target sites. Our results demonstrate that the protonation state of histidine notably affected the DNA minor-groove width profile and binding free energy. Examining flanking sequences of various binding affinities derived from SELEX-seq experiments, we analyzed the relationship between binding affinity and specificity. We uncovered how histidine protonation leads to increased binding affinity but can lower specificity. Our findings provide new mechanistic insights into the role of histidine in modulating TF-DNA binding.

SIGNIFICANCE   This study reveals the molecular mechanism for the significance of a histidine residue in the Scr Hox protein. Our findings indicate that His-12 plays a crucial role in stabilizing the linker region of the Hox protein with the Scr core motif by regulating the minor groove shape profile, which is essential for Hox specificity. The uniqueness of His-12 lies in its ability to adopt two protonation states at biological pH. Our results demonstrate that the differences in behavior between the two protonation states of His-12 not only explain its preference toward different core motifs but also provide insights into its preference for binding of high- and low-affinity target sequences.

INTRODUCTION

In the crowded genomic environment, transcription factors (TFs) navigate to their target DNA-binding sites and orchestrate gene expression. Because of this important role, understanding the mechanisms of TF binding has been an important research area (1–3). TFs from the same family can adopt different binding mechanisms to bind to their specific DNA targets (4). Binding affinity is influenced not only by the sequence preferences of the core binding motif (5) but also by the structural elements of the protein (6), DNA shape (7,8), electrostatics (9,10), and genomic context (6,11). In addition to the core motifs, the flanking sequences surrounding the core binding site have recently received increased attention (12–15).

Hox (homeobox) proteins, encoded by Hox genes, play crucial roles in animal body development, such as specifying embryo segmentation across species (16). Hox proteins regulate diverse types of target genes, including genes specific for a particular Hox paralog as well as genes spanning multiple Hox paralogs. This divergence hints at the existence of both paralog-specific and less-specific Hox-binding sites (17). Nonetheless, the capacity for Hox specificity is challenged by the fact that these proteins all bind to highly similar DNA sequences (18). Distinctive TF binding mechanisms are used by different Hox homologs to bind to the core motifs with their preferred DNA shape (8,19). Our study focuses on one anterior Hox protein in *Drosophila*, Sex combs reduced (Scr), due to available structures of its binding to different DNA targets (19). When this Hox protein binds to either the Scr-preferred or the Ultrabithorax (Ubx)-preferred sequence, a double- or single-minimum shape, respectively, emerges in the minor groove width (MGW) (19). This shape preference can be

explained by contacts between two key residues of Scr, arginine 3 (Arg3) and histidine-12 (His-12), with DNA in the core motif region. These residues facilitate the shape readout (7) of DNA by Scr to differentially bind to specific DNA sequences in the genome (8,19). Here, we analyze how these residues impact DNA MGW.

Histidine is an amino acid with a pK_a of around 6.0 (20), which allows this residue to assume two protonation states, neutral or protonated, under biological conditions. When basic amino acids, such as histidine, are surrounded by a charged environment as present in nucleic acids, local pH can alter their charged state (21–26). In neutral histidine, the Nδ1 atom can act as a hydrogen-bond acceptor. In the protonated state, both the Nδ1 and Nϵ2 atoms of histidine can act as hydrogen bond donors (20). It is crucial to consider both protonation states because they can affect the hydrogen-bonding patterns and electrostatics within protein-DNA interactions. Given its role in regulating the binding specificity of protein-DNA interactions, histidine is indispensable in TF-DNA recognition (27). Using different hydrogen-bonding capabilities and electrostatic potentials between two protonation states, histidine affects the stability of protein-ligand interactions (28) for drug design (29), target identification (30), and drug delivery (31). Additionally, varying the histidine protonation state can induce structural changes in proteins and result in protein pathogenic mutations (28). In the Extradenticle (Exd)-Hox system, positively charged Arg3 and thymine surround His-12. The His-12 residue, in turn, forms crucial contacts with DNA and, together with Arg3, modulates the Hox-DNA binding affinity (19).

Although experiments exist to identify the protonation states of protein residues (32), as yet no experimental data on the His-12 protonation states of the Exd-Scr–DNA complex are available. Moreover, detailed mechanistic studies of the Hox-DNA complex that explain the role of His-12 in binding are lacking. To address these gaps, we used molecular dynamics (MD) simulations to answer how histidine protonation states can affect the binding specificity of Scr protein binding to the Scr-preferred site (PDB: 2R5Z) and the Ubx-preferred site (PDB: 2R5Y). Utilizing data generated by systematic evolution of ligands by exponential enrichment followed by high-throughput sequencing (SELEX-seq) (33), we investigated how histidine protonation impacts binding to the DNA core motif for DNA flanking sequences of various affinities. Our mechanistic findings can provide additional insights into the binding mechanisms of Hox family proteins.

MATERIALS AND METHODS

Sequence selection for molecular dynamics simulations

We simulated the DNA sequence observed in the co-crystal structure of the Exd-Scr complex with the Scr-preferred core binding motif (PDB:

2R5Z) and the Ubx-preferred core binding motif (PDB: 2R5Y). These co-crystal structures share the same protein heterodimer but have different underlying DNA sequences. The core binding motifs are the Scr-preferred core, $A_1G_2A_3T_4T_5A_6A_7T_8$, and the Ubx-preferred core, $T_1G_2A_3T_4T_5$-$T_6A_7T_8$. Next, we performed mutation simulations on the Arg3 and His-12 residues in both structures, using sequences from the co-crystal structures, to test the influence of His-12 protonation states. Single-residue mutations of the Scr protein were generated using PyMol (34). Finally, to assess the influence of His-12 on flanking sequences, we fixed the 5′ end flanking sequences and selected the 3′ end flanking sequences (because His-12 is closer to the 3′ end). We selected sequences with high, medium-high, medium-low, low, and very-low binding affinities using data from SELEX-seq experiments (35), as described in supporting material section S-I and Fig. S1.

Simulation protocol

To fill in missing residues that are not present in the co-crystal structure, such as those in the N-terminal linker region of the Scr protein and missing residues in the Exd protein, we used AlphaFold2 (36) to build the initial structures of the Exd and Scr homeodomains. We then aligned them to the co-crystal structure of the Exd-Scr heterodimer bound to the Scr-preferred site (PDB: 2R5Z). The resulting protein structures aligned well with the co-crystal structure, with a root-mean-square deviation (RMSD) of 2.53 Å for all the protein heavy atoms and 1.09 Å for the protein heavy atoms of the homeodomains. The same Exd-Scr heterodimer is used for the system with the Ubx-preferred site, with an RMSD of 3.22 Å for all the protein heavy atoms when aligned to the co-crystal structure (PDB: 2R5Y) and 1.21 Å for the protein heavy atoms of the homeodomains.

To construct the DNA used in the simulations of sequences from SELEX-seq experiments, we generated DNA structures for the flanking sequences with GC caps as shown in Table S1 using 3DNA (37). The DNA in the co-crystal structure (PDB: 2R5Z) was trimmed to only contain the core motif sequence. The 3DNA sequences were designed to contain extra two basepairs that match the ends of the core motif sequence. These short DNA flanks were then aligned and appended to the ends of the core motif sequence in the co-crystal structure using the matching bases. The geometry of the constructed DNA was refined and minimized using PHENIX (38) with the *geometry_minimization* program. This was done to maintain the same DNA shape in the core binding domain as the DNA from the co-crystal structure. The same procedure was used for the system with the Ubx-preferred site.

All simulations and subsequent analyses were performed using the GROMACS 2020.3 software package on a GPU (NVIDIA Tesla A40/ 100) with the AMBER ff14SB (39) force field for the protein and the Parmbsc1 (40) force field for the DNA. Protonation was added by using the GROMACS 2020.3 *pdb2gmx* program. All complexes were solvated by using the explicit TIP3P water model. The negative net charge of the Exd-Hox–DNA complex was neutralized by adding Na$^+$ and Cl$^-$ counterions to reach a final NaCl concentration of 150 mM, which approximates the physiological concentration. The GROMACS 2020.3 *genion* program was used to distribute these counterions throughout the box. First, all systems were subjected to 2000 steps of steepest descent energy minimization to distribute solvent molecules around the solute. Next, we performed three rounds of NVT equilibration (constant number of particles, volume, and temperature) at 10 ps each to gradually heat the system to 300 K, followed by one round of NPT equilibration (constant number of particles, pressure, and temperature) at 700 ps to equilibrate the pressure with v-rescale thermostat and Parrinello-Rahman barostat: coupling constants $\tau_T = 0.1$ ps and $\tau_P = 1$ ps, reference temperature $T_{ref} = 300$ K, reference pressure $P_{ref} = 1$ bar. The production simulations for all systems were run for 300 ns in the isobaric-isothermal ensemble ($P = 1$ bar, $T = 300$ K). The integration time step of 2 fs was used for all calculations. The Verlet cutoff scheme was used for all calculations. Long-range electrostatic interactions were computed using the particle-mesh Ewald method (41) with a 12 Å

cutoff. Nonbonded van der Waals interactions were calculated with a 12 Å cutoff. The LINCS (42) algorithm was employed to constrain all bonds.

Simulation convergence was assessed by counting structural clusters as a function of time using the Gromos algorithm (43,44) implemented in GROMACS (*cluster* package) (Fig. S2). Simulations show convergence after around 150 ns (supporting material section S-II).

We ran three replica simulations each for the wild type (WT), R3A, and H-12A, using independent random seeds. The behavior discussed here is consistent among the three replicas. The individual replica results can be found in the provided Figshare link.

Biophysical calculations

His-12 protonation state prediction

We selected the PROPKA 3 (45,46), H++ (47), and DelPhiPKa (48) methods for pK_a prediction and reported calculated results for both the Scr- and Ubx-preferred sites in supplemental tables in the provided Figshare link. We calculated the His-12 pK_a for both the protein-only system and the protein-DNA complex to examine the effect of DNA on the His-12 protonation state.

Electrostatic surface potential calculation

Visualization of electrostatic surface potentials was generated with APBS (49) using the PyMol-2.4.0 APBS plugin. Hydrogen atoms were added to the protein-DNA structures using PDB2PQR (50). The electrostatic potential calculations were performed by solving the nonlinear Poisson-Boltzmann equation at a temperature of 300 K, solute and solvent dielectric constants of 4 and 80, respectively, and ion concentration of 150 mM. The exclusion radius was set to 2 Å.

Hydrogen bond calculation

Hydrogen bonds were calculated with the GROMACS 2020.3 *hbond* package. Hydrogen bonds were defined as a distance cutoff of 3.5 Å and an angle cutoff of 120° between the donor and acceptor. All hydrogen bonds were analyzed for the last 150 ns of simulations.

DNA shape calculation

DNA MGW values were calculated with Curves 5.3 (51) using MD snapshots obtained every 100 ps. MGW calculations were plotted for the last 150 ns of simulations.

Root-mean-square fluctuation calculation

The root-mean-square fluctuation (RMSF) values were calculated using the GROMACS 2020.3 *rmsf* program. The RMSF was computed for the C_α atoms of the protein residues for the last 150 ns of the trajectory. A least-squares superposition to the equilibrated structure (after NVT and NPT) was performed before computing RMSF.

Molecular mechanics/Poisson-Boltzmann surface area calculation

For each mutation simulation, a single trajectory approach was used for molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) calculation with g_*mmpbsa* (49,52). The energy components ΔE_{MM}, ΔG_{Polar}, and ΔG_{Apolar} of the protein-DNA complexes were calculated for 200 snapshots extracted every 500 ps from the production trajectories from 150 to 300 ns. ΔE_{MM} was calculated from van der Waals and electrostatics interactions based on molecular mechanics force-field parameters (52). Dielectric constants 4 and 6 were examined for this protein-DNA system based on (53,54). To calculate ΔG_{Polar}, a box was generated using the extreme coordinates of the molecular complex in each dimension. The box was then expanded in each dimension by 3-fold to obtain a coarse-grid box (cfac = 3). A finer grid box was then placed within the coarse grid box ex-

tending 30 Å (fadd = 30) from the complex's extreme coordinates in each direction. The ionic strength was set to 150 mM NaCl with radii of 0.95 Å for sodium ions and 1.81 Å for chloride ions. The solvent dielectric constant was 80. The solvent radius was 1.4 Å. The temperature was set to 300 K. The internal dielectric constant was set to the same value as the dielectric constant when calculating ΔE_{MM}. The value for the vacuum (vdie) dielectric constants was set to 1. The nonlinear PB equation was solved using APBS. The solvent-accessible surface area (SASA) nonpolar model was used to calculate ΔG_{Apolar} where the probe radius for SASA was 1.4 Å, and the surface tension (gamma) was set to 0.0226778 kJ/(mol A^2).

RESULTS

Protonated His-12 enhances and stabilizes minor groove narrowing

We assessed the His-12 protonation state by predicting the pK_a. In protein-only systems, the pK_a values of His-12 were calculated to be 6.04, 4.45, and 5.89 using PROPKA 3, H++, and DelPhiPKa (see materials and methods), showing that histidine was predominantly neutral. For protein-DNA systems with the Scr-preferred site, predicted pK_a values increased to 7.08, 9.41, and 6.92. Despite differences in predicted pK_a values, His-12 proton uptake from DNA binding was consistently observed across all three methods, also observed in other studies (22–26). This result underscores the importance of simulating both histidine protonation states.

Running an atomistic MD simulation of the 21-bp DNA sequence (as in the co-crystal structure) in complex with the Exd-Scr heterodimer (Fig. 1 A), we plotted DNA shape for the Scr-preferred core sequence for both neutral and protonated His-12 states (Figs. 1 B and S3 A). Our MD simulations essentially confirmed the double-minima "W" shape profile for MGW of the Scr-preferred site (19). Protonation of His-12 resulted in pronounced narrowing of the MGW at the A_7 nucleotide, reflecting its impact on the second minimum. Additionally, His-12 protonation reduced the fluctuation of the Scr protein (Fig. S4 and supporting material section S-IV), showing that His-12 protonation enhances and stabilizes minor groove narrowing.

To further validate that protonation was responsible for the effect on MGW, we performed and analyzed protonation switch simulations. In brief, we extracted the last frame from both the neutral and protonated simulations before switching the His-12 protonation state. Next, we minimized and equilibrated the system before resuming the simulation. Our findings showed that changing the histidine protonation state led to corresponding MGW profile adjustments. Specifically, the initially protonated system with a more prominent second minimum transitioned into a somewhat neutral DNA shape profile with a wider groove (Fig. 1 C, *bottom*). By contrast, the initially neutral system with a less prominent second minimum was switched to a protonated system with a narrower groove (Fig. 1 C, *top*). Taken together, these findings further confirm that protonation enhances narrowing of the DNA minor groove. To investigate the mechanism
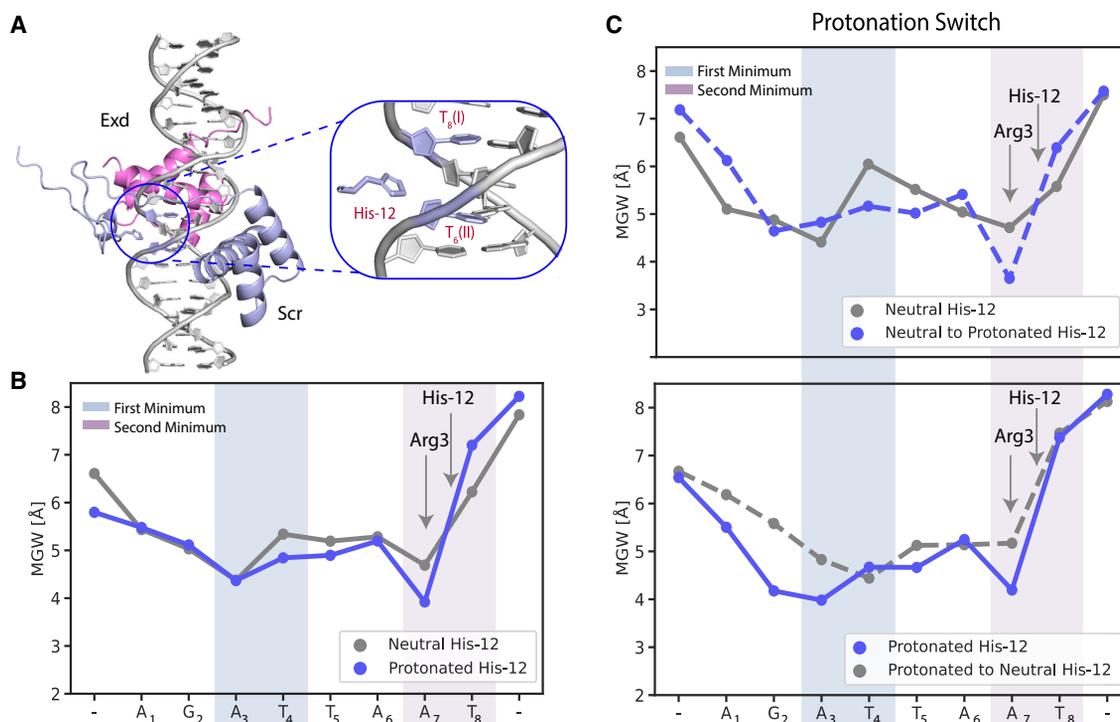
FIGURE 1 Effect of protonation on Exd-Hox–DNA system. DNA sequence was derived from the co-crystal structure (PDB: 2R5Z). (*A*) Exd-Hox–DNA system. Inset: magnification of linker region where His-12 inserts into DNA minor groove. I indicates DNA strand 1, and II indicates DNA strand 2. (*B*) DNA MGW shape profile. Blue and gray lines: systems with protonated and neutral His-12, respectively, averaged among three replicates. Blue and purple shaded regions: first and second minima, respectively. Relative position of His-12 is marked. (*C*) MGW profiles with protonation switch. Upper panel: after reporting the MGW profile for the neutral His-12 system (*gray solid line*) for replica 1, we switch to the protonated system and plot the MGW profile again (*blue dashed line*). Lower panel: after reporting the MGW profile for the protonated His-12 system (*blue solid line*) for replica 1, we switch to the neutral system and plot the MGW profile again (*gray dashed line*).

by which His-12 narrows the MGW, we calculated the distance between neighboring nucleotides (T_6 and T_8) to His-12 (Figs. 1 *A* and S5). We found that MGW was narrowed by synchronized pinching from both backbone strands (supporting material section S-V).

We further expanded our analysis to understand the influence of His-12 protonation on the stability of the whole protein-DNA complex based on its thermodynamics. The binding free energy of a protein-DNA complex was previously correlated with binding stability (55). Therefore, we used the MM/PBSA protocol (49,52) to calculate the binding energy of the systems with protonated and neutral His-12 (supporting material section S-VI). The total binding energy of the system with protonated His-12 was lower than that with neutral His-12 (Fig. S6, *C* and *F*). Based on a decomposition of the binding free energy for each residue, we found that protonated His-12 had a greater influence on binding and was largely responsible for the lower free energy of that system (Fig. S6 *B*).

His-12 protonation-enhanced MGW narrowing had long-range conformational effects on the Exd-Scr complex, inducing side-chain rotation of two Exd residues, Arg5 and Lys61 (Figs. 2 *A* and S7; supporting material section S-VII). This conformational change occurred because of the rotation of α helices in the Exd protein, which was

driven by the change in His-12 protonation state (Fig. 2 *B*). Moreover, this long-range conformational change had major effects on the interface electrostatics of the Exd-Hox–DNA complex (Fig. 2, *C* and *D*). In the neutral state, Arg5, residing in the α helices of the Exd protein, was buried in the minor groove of the DNA. With His-12 protonation, Arg5 moved outside the DNA minor groove toward the protein surface, thereby increasing the surface charge (Fig. 2 *C*). Similarly, the rotation of Lys61 created an additional positive surface patch (Fig. 2 *D*). Enhanced surface charges can play a role in recruiting TFs (56). For Hox proteins, enhanced surface charges can change protein-protein interactions with chromatin-modifying enzymes, such as CREB-binding protein (CBP) (57), thereby altering transcription. These results highlight the significance of the His-12 protonation state on the conformation of the Exd-Scr–DNA complex.

Mutations of Arg3 or His-12 reveal the role of His-12 in modulating minor groove shape

When either the His-12 or Arg3 residue is mutated to alanine, the binding affinity of the mutated construct is reduced, but the mutation does not affect the complex with the Ubx-preferred sequence (19). Moreover, His-12
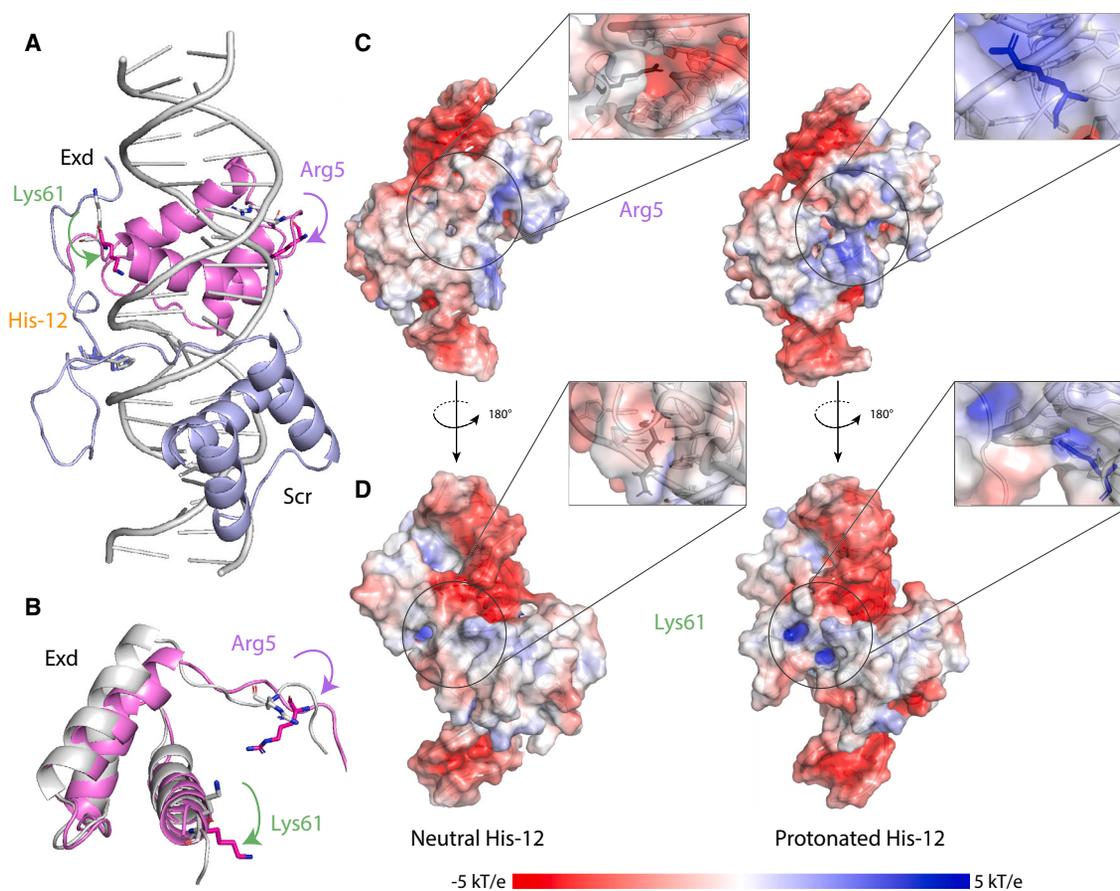
FIGURE 2 Conformational differences of lysine and arginine residues as a result of His-12 protonation. (*A*) Exd-Scr heterodimer bound to DNA with Exd shown in magenta and Scr shown in blue. Lys61, His-12, and Arg5 residues are shown in stick representation, with gray showing residues in the system with neutral His-12 and blue and magenta showing residues in the system with protonated His-12. Arrows indicate positional differences between residues in complex with neutral and protonated His-12. (*B*) Conformational changes in Exd α helices. Magenta and gray: Exd protein in the systems with protonated and neutral His-12, respectively. (*C*) Comparison of surface electrostatics between neutral and protonated His-12 systems for replica 1. Inset: region showing charge disparity as a result of Arg5 positioning. (*D*) Complexes are rotated 180° to the back, where Lys61 resides. Circle highlights difference in surface charges and shows position of residues beneath the protein surface.

and Arg3 contribute differently to the target DNA binding (19). Here, using MD simulations, we disentangle the mechanisms of each mutant, and we unravel the individual contributions of these mutants to binding affinity. We ran 300-ns MD simulations on 1) the WT system with DNA sequences from PDB: 2R5Z (19), 2) a system in which Arg3 was mutated to alanine (Scr^{R3A}), with both neutral and protonated His-12, and 3) a system in which His-12 was mutated to alanine (Scr^{H-12A}).

Comparing the average DNA shape profiles for these protein mutants to the WT system, we observed that protonation of His-12 resulted in a ≈ 1 Å and ≈ 1.2 Å reduction in MGW at the second minimum in the WT and Scr^{R3A} systems, respectively (Fig. 3, *A* and *B*), highlighting the role of His-12 protonation in modulating the MGW. The fluctuation of MGW is shown in Fig. S8. For the complete DNA sequence MGW profile see Fig. S3, *B* and *C*.

Near the 5′ end, the DNA shape profile at the first minimum diverged dramatically between the Scr^{R3A} and WT

systems with neutral His-12 (Fig. 3 *A*), revealing the role of Arg3 in controlling the MGW at a further distance with its charged guanidinium group. Compared to the WT system, the MGW at the 5′ end was ≈ 1.3 Å wider in the Scr^{R3A} system, whereas the His-12 mutation exhibited a smaller groove widening (≈ 0.4 Å). Arg3 likely influences the MGW at the 5′ end through electrostatic interactions, whereas neutral His-12 has a smaller effect at long distances. The strong long-range groove-narrowing ability of Arg3 can explain why mutating Arg3 results in a large reduction in binding affinity compared to the His-12 mutation in DNA-binding experiments (19). In contrast, the groove-narrowing role of Arg3 at the 5′ end diminishes when His-12 assumes a protonated state (Fig. 3 *B*), possibly because the additional charge from His-12 protonation compensates for the loss of charge in the Scr^{R3A} system, thereby narrowing MGW at the 5′ end.

To assess the structural impact of these mutations, we plotted the RMSF for the mutated systems. Based on the
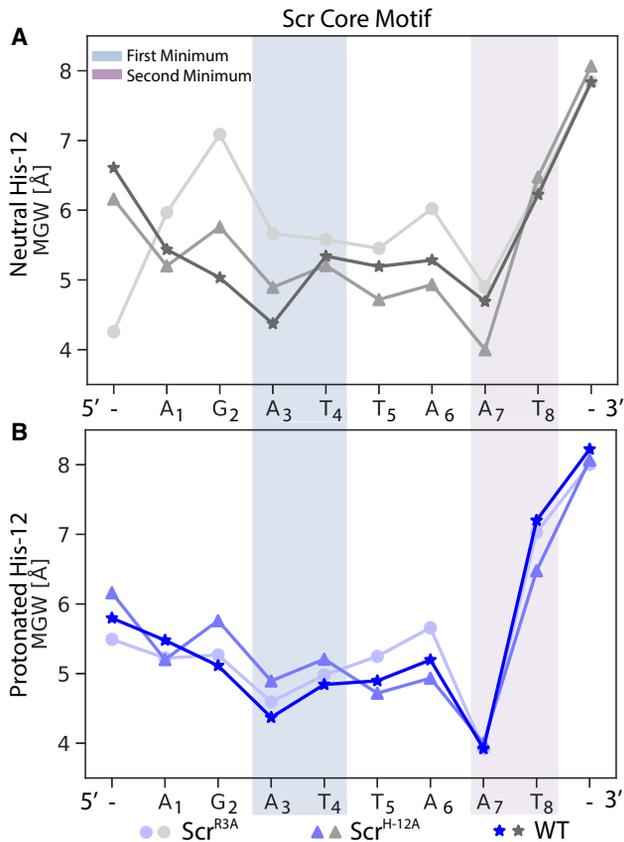
FIGURE 3   DNA shape profiles for wild-type (WT) and mutated systems with (*A*) neutral His-12 and (*B*) protonated His-12 averaged over three replicas. Shaded region shows the double minima for the Scr-preferred binding site. Blue and purple shaded regions indicate the first and second minima, respectively.

results, the system with the His-12 mutation showed the greatest fluctuation in the linker region (Fig. S9, *C* and *E*) and the widest MGW distribution at the second minimum (Fig. S9 *A*). This result serves as further evidence that His-12 is crucial in stabilizing the linker region. Moreover, the WT system had the fewest fluctuations (Fig. S9 *C*) and the narrowest distribution of MGW at A_7 (Fig. S9 *A*), confirming that both His-12 and Arg3 residues are important for binding of the Scr protein to the Scr target site (19,35). We further assessed the influence of Arg3 or His-12 mutations on binding stability by calculating the binding energy among protein mutants and WT. The WT had the lowest binding energy among all systems, supporting our previous findings that both residues together contribute to a more stable complex (supporting material section S-VI).

Ubx exhibits a different preference for the protonation state of His-12

The Scr-preferred site has a double-minimum DNA shape at the core motif. In contrast, the preferred site of its posterior homolog Ubx has a single-minimum DNA shape profile,

despite interacting with the same protein heterodimer. We next investigated how protein-DNA interactions orchestrate this preference, by performing 300-ns MD simulations on the Exd-Scr heterodimer in complex with a DNA sequence containing the Ubx-preferred site. Simulated systems included Scr^{R3A}, Scr^{H-12A}, and WT (PDB: 2R5Y) (19) with both histidine protonation states.

The MD simulation of the Ubx core motif revealed a DNA MGW shape with a single minimum at the 5′ end, in contrast to the double-minimum DNA shape profile of the Scr core motif (Fig. 4 *A*). These different DNA shape observations for Ubx- versus Scr-preferred sites are in agreement with data from co-crystal structures (19) and Monte Carlo-based predictions of unbound target sites (8). Despite small differences in DNA shape in the mutated systems with the Ubx site, the His-12 or Arg3 mutation did not appear to alter the fundamental single-minimum "V" shape in the core motif region. For the fluctuation of MGW, see Fig. S10. Additionally, the role of these residues in anchoring the linker region was reduced because either mutation resulted in a smaller RMSF change in the linker region compared to the Scr-preferred site in all simulations (Fig. S9, *D* and *F*). These differences suggest that His-12 and Arg3
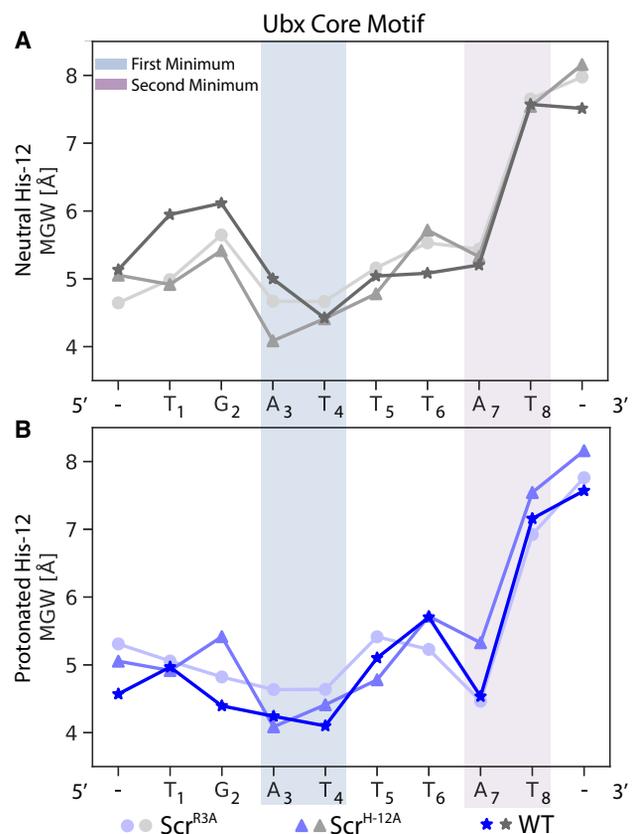
FIGURE 4   DNA shape profiles for the wild-type (WT) and mutated systems with (*A*) neutral His-12 and (*B*) protonated His-12 averaged over three replicas. Shaded region shows the locations of the double minima for the Ubx-preferred motif. Blue and purple shaded regions indicate first and second minima, respectively.

are not as crucial for binding to the Ubx core motif as they are for the Scr core motif. This difference is likely due to the replacement of A_6 in the Ubx core motif with the T_6 base, which results in an elongated A-tract sequence and a more rigid DNA region (Fig. S11 and supporting material section S-IX). This rigidity cannot be overcome by the attraction of protein residues to the DNA backbone, thus the minor groove remains wide. However, similar to the effect on the Scr core target site, protonation of His-12 can effectively overcome this rigidity and narrow the DNA minor groove at the second minimum of the Ubx core target site upon Exd-Hox binding (Fig. 4 B).

Unlike Scr, the structure of Ubx is made more flexible by the protonation of His-12 (Fig. S4). We observe that the overall RMSF was much higher across the Ubx protein chain (arrow in Fig. S9 D). In addition, the distribution of the MGW at the A_7 position was wider, exhibiting greater fluctuation despite the groove-narrowing effect of the protonated His-12 (Fig. S9 B). The increased overall flexibility of the Scr protein with the Ubx core motif reflects the ability of the Scr protein to undergo conformational fluctuations to accommodate the Ubx-preferred sequence motif.

Using PROPKA 3, H++, and DelPhiPKa for His-12 in complex with the Ubx-preferred site, we predicted pK_a to be 6.16, 5.75, and 6.77, respectively (see materials and methods), indicating a higher likelihood for His-12 to be neutral. When His-12 was neutral, the binding energy was notably higher in WT systems with the Ubx core motif than with the Scr core motif (Fig. S6, C and F). While forcing protonation reduces binding energy for the Ubx core, the calculated pK_a values suggest that His-12 is more likely to be neutral, resulting in a less stable system compared to the one with the Scr-preferred site. Taken together, these results suggest that His-12 plays an important role in differentiating similar cognate motifs across various homologs.

Flanking regions influence Exd-Hox–DNA binding affinity by modulating DNA shape

Flanking sequences have been shown to influence the binding affinity for Exd-Hox–DNA complexes (14). To investigate the influence of His-12 on flanking sequences with different binding affinities, we constructed DNA sequences with a high-affinity Scr core motif, $T_1G_2A_3T_4T_5A_6A_7T_8$, which is different from the core sequence present in the co-crystal structure (Fig. 5 A). To focus on the role of His-12 in the presence of different flanks, we only varied flanks on the 3′ end of the DNA, which is where the Scr protein linker interacts with the minor groove. Flanking sequences with high, medium-high, medium-low, or low affinity, and a sequence with long A-tract (very-low affinity), were chosen from the SELEX-seq experiments (35) (see materials and methods). We performed 300-ns MD simulations for each of the five different flanks with both protein heterodimers and His-12 protonation states and their respective unbound DNA targets.

Analyzing DNA shape for high- and low-affinity sequences, we identified the consensus shape in their MGW profiles. High-affinity sequences displayed a more prominent double-minimum "W" shape overall, with lower MGW at the second minimum (Fig. 5, B and C, purple-shaded regions). In contrast, low-affinity sequences had a flatter consensus shape across the core motif when His-12 was neutral (Fig. 5 B). Upon histidine protonation, the DNA shape of low-affinity sequences was more significantly affected at both the 3′ and 5′ ends compared to high-affinity sequences. To illustrate this difference, we examined the MGW distribution at the second minimum (A_7) and identified large discrepancies in the groove-narrowing effect of the His-12 protonation between high- and low-affinity sequences (Fig. 5 D). Specifically, His-12 protonation induced a more prominent groove narrowing for low-affinity sequences, with a ≈ 1.5 Å reduction in the mean MGW and a noticeable shift in its distribution. In contrast, His-12 protonation had a minor impact on high-affinity sequences, showing almost no change in the distribution of the MGW at the second minimum. Because the high-affinity sequences already closely matched the desired shape profile for Scr binding, protonation had less of an effect on MGW. By contrast, the lower affinity sequences exhibited more substantial MGW profile changes upon His-12 protonation to attain a more stable binding conformation. We further validated the groove-narrowing effect of His-12 on low-affinity sequences by investigating the MGW fluctuation over the trajectory at the second minimum (Fig. S12, A and B). The time-series data confirmed that histidine protonation significantly affected the DNA MGW of low-affinity sequences while having minimal impact on high-affinity sequences (supporting material section S-X).

It is possible that the Exd-Hox protein heterodimer not only selects the depth of the MGW at the second minimum but also the "shape" of the second minimum. To quantify this shape, we calculated the MGW differences between A_7 and the neighboring nucleotides (supporting material section S-X). High-affinity sequences exhibited a much narrower MGW at A_7 compared to the adjacent A_6 and T_8 nucleotides, resulting in a more pronounced "V" shape, unlike the low and medium-low sequences (Fig. S12, C and D). We defined this aspect as a higher ΔA_7 MGW, indicating a distinct narrowing of the MGW at A_7. This deepening of the "V" shape at the second MGW minimum was more pronounced in high-affinity sequences than in lower-affinity sequences, indicating a sharp MGW narrowing at A_7. This result indicates that the MGW of the DNA basepairs around the second minimum is also important in differentiating high- and low-affinity sequences.
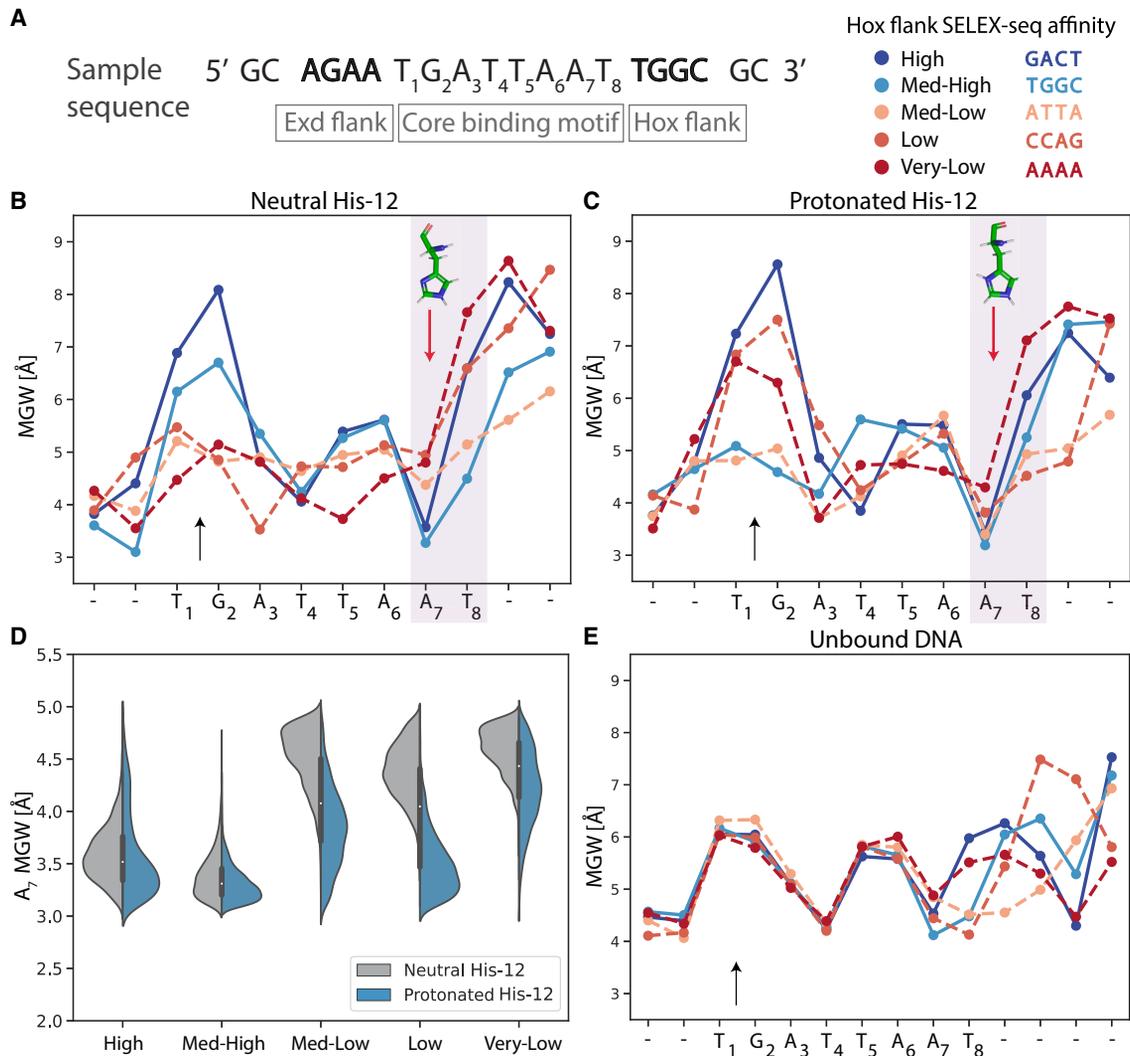
FIGURE 5 Effects of flanking sequences on DNA shape. (*A*) Sample sequence, highlighting flanks on the 5′ and 3′ ends and the core motif used for Scr. (*B* and *C*) DNA shape in core motif region for system with neutral (*B*) and protonated (*C*) His-12. Sequences with lower binding affinity (*red dashed lines*) and higher affinity (*blue solid lines*) are marked by a red arrow for the relative position of His-12, purple-shaded region for the position of the second minimum, and black arrow for the divergence in DNA shape at the 5′ end. (*D*) Split violin plots showing MGW distribution at A_7 basepair for five flanking sequences, with systems containing neutral His-12 (*gray*) or protonated His-12 (*blue*). (*E*) DNA MGW shape profile for unbound DNA target sites.

Interestingly, although the 5′ flanking sequence was the same, the system with neutral His-12 deviated in MGW at the 5′ end between high- and low-affinity sequences (*black arrow* in Fig. 5 *B*). For low-affinity sequences, neutral His-12 only influenced DNA shape at the second minimum but failed to create the first minimum seen in the unbound DNA (*black arrow* in Fig. 5, *B* and *E*). Protonation of His-12 resulted in a first minimum for low-affinity sequences (*black arrow* in Fig. 5 *C*). This result can be explained by the charge difference between the two His-12 protonation states. Protonated His-12 has an additional positive charge that can impact the MGW at a farther distance through electrostatic interaction. In addition, the strength of electrostatic interaction is stronger, such that it can subdue the unfavorable DNA shape of low-affinity sequences and change it to that of a preferred "W" shape.

Hydrogen bonds between the Exd-Hox linker and DNA can determine binding affinity

We next investigated the molecular forces underlying differences between high- and low-affinity sequences. To understand which residues are crucial in determining binding affinity, we calculated the RMSF of the backbone heavy atoms for all five sequences with both His-12 protonation states (Fig. 6, *A* and *B*). We found a positive association between RMSF in the linker region and binding affinity among the five sequences, showing the involvement of the linker region in determining the binding affinity between Scr and its target DNA.

To gain deeper insights into the selection mechanism of the linker region, we identified the hydrogen bonds that the linker region forms with DNA during the simulation.
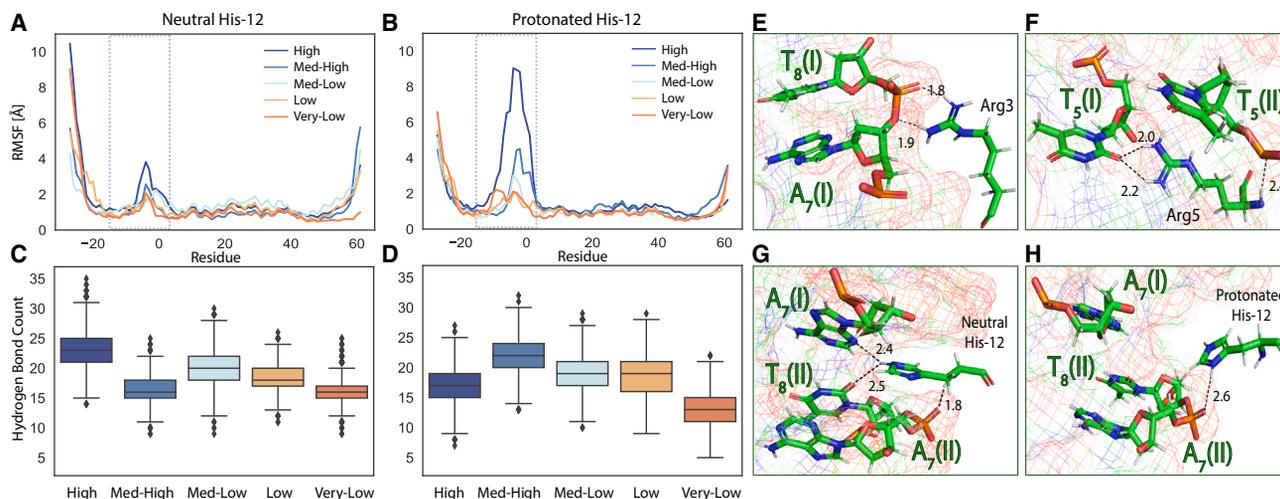
FIGURE 6 Protein-DNA linker region hydrogen-bonding interactions for sequences with different binding affinities. (*A* and *B*) RMSF values of Scr with neutral (*A*) and protonated (*B*) His-12 system. RMSF values are colored by affinity obtained for sequence by SELEX-seq (35), with blue and orange indicating higher and lower binding affinity, respectively. Blue rectangle indicates the Scr linker region. (*C* and *D*) Bar plots show the distribution of hydrogen-bond counts between linker region and DNA. The edges of boxes are defined as the 25th to 75th percentiles, the whiskers represent the minimum and maximum values excluding the outliers, and the diamonds indicate outliers. (*E–G*) Visualizations of specificity-conferring hydrogen bonds in system with neutral His-12. Hydrogen-bonding pairs are highlighted by dashed black lines with distances labeled in Å. (*H*) Loss of specific hydrogen bonds in system with protonated His-12 compared to system with neutral His-12 in (*G*). In (*E*)–(*H*), I indicates DNA strand 1, and II indicates DNA strand 2.

As binding affinity increased, the DNA formed a greater number of hydrogen-bonding contacts with linker residues (e.g., Arg3 and His-12) (Fig. 6, *C* and *D*). Additionally, a greater number of significant hydrogen-bonding pairs (with hydrogen bond occupancy >50%) were observed in the neutral His-12 system compared to the protonated His-12 system (Fig. S13).

It is widely acknowledged that hydrogen bonds can convey specificity between protein residues and DNA basepairs (58,59). Such specificity depends not only on the total number of hydrogen-mediated bonds but also on the spatial arrangement between donors and acceptors (7,59). Bidentate hydrogen bonds, where donors and acceptors differ between two hydrogen bonds, form a unique geometry that allows for the highest degree of readout specificity (7). A reduced level of specificity is conferred by bifurcated hydrogen bonds, wherein two hydrogen bonds share the same donor atom. Single hydrogen bonds convey a comparably lower level of specificity. A recent study (60) showed that low-affinity sites are critical in determining Hox-binding specificity, whereas high-affinity sites do not confer specificity in the same way. Indeed, we identified many bidentate and bifurcated hydrogen bonds unique to low-affinity sequences (Fig. S13). Specific examples include: Arg3 forming a bidentate hydrogen bond between its guanidinium group and the oxygen atoms of the $A_7(I)$ and $T_8(I)$ bases (Fig. 6 *E*); Arg5 forming bifurcated hydrogen bonds with the oxygen atoms of the $T_5(I)$ base (Fig. 6 *F*); and neutral His-12 forming intricate hydrogen bonds contacting the $T_8(II)$, $A_6(II)$, and $A_7(II)$ bases (Fig. 6 *G*) (I and II indicate the two DNA strands). These residues form hydrogen bonds that specif-

ically recognize bases within the core motif, conveying specificity in the presence of neutral His-12. However, with His-12 protonation, some specificity-conferring hydrogen bonds are lost. For instance, protonated His-12 no longer contacts the $T_8(II)$ and $A_7(II)$ bases due to its slight shift toward the phosphate group of the minor groove backbone (Fig. 6 *H*). Although His-12 protonation increases binding affinity (Fig. S6), it does not elevate specificity. Similar losses of specific hydrogen bonds were observed in the bifurcated hydrogen bonds made between Thr-9 and Ser-8 residues when His-12 was protonated (Fig. S13). We conclude that hydrogen bonding likely serves as a key mechanism by which Hox proteins determine binding specificity.

DISCUSSION

Due to its unique pK_a, histidine is a notable residue in protein-DNA interactions. It plays an essential role in protein-DNA recognition. Our study underscores the importance of considering the histidine protonation state at the protein-DNA binding interface. We showed that different histidine protonation states alter electrostatic interactions and hydrogen-bonding patterns in the Exd-Scr–DNA complex. Through these molecular interactions, histidine aids in differentiating the Scr versus Ubx Hox homologs, shedding light on Hox gene specificity. Our approach can likely extend to other TF-DNA complexes, such as the early growth response 1 (EGR1) (61) or zinc finger protein families (62), to elucidate the interaction dynamics between DNA and histidine at the binding interface that gives rise to specific binding across homologs.

Although SELEX-seq experiments provide high-throughput insights into the flanking sequence preference for TF-DNA complexes, it remains challenging to decipher the mechanisms underlying the observed binding affinity. Structural studies on these complexes are essential for understanding these intricate systems. Our approach uncovers molecular mechanisms to achieve differential binding affinity. Specifically, although the flanking sequences of the Exd-Hox heterodimer seem to have little in common among low- or high-affinity sequences, we identified a consistent MGW profile among flanking sequences of high-affinity sequences, which differed from the consensus shape profile of the low-affinity sequences. This DNA shape profile distinction can be key to explaining the difference in binding affinity.

Furthermore, our approach delves deep into the relationship between binding specificity and affinity within the Exd-Hox–DNA complex. Through MD simulations of SELEX-seq derived sequences, we observed that protonation of histidine enhanced binding affinity (Fig. S6) but reduced the number of specificity-conferring hydrogen bonds (Figs. 6 and S13), which could be correlated with decreased binding specificity in low-affinity sequences. Previous research has also emphasized the significance and functional relevance of low-affinity binding sites in their interactions with Hox proteins (60,63). Our analysis pinpointed specific hydrogen-bond pairs between low-affinity DNA sequences and Scr protein, shedding light on the potential mechanism behind their specificity. Given the crucial role of histidine in TF-DNA binding uncovered in this study, our approach can be extended to investigate the binding mechanisms in other Hox co-factor systems.

Our simulation protocol can also be adapted for various TF-DNA complexes using binding affinity data from experimental methods such as SELEX-seq, high-throughput SELEX (64), or genomic-context protein-binding microarrays (6). We can uncover detailed structural mechanisms to understand the binding specificity and affinity of many TF-DNA complexes that, despite being in the same family, carry out different genomic functions. Coupled with DNA shape analysis, our approach helps uncover how proteins achieve sequence and shape recognition. Many proteins utilize different readout mechanisms for DNA targets, raising questions about the preference for base or shape readout (7,65–67). Our simulation protocol offers an avenue to explore shape recognition. For instance, in our analysis of DNA readout by Hox proteins, we identified two distinct consensus shape profiles associated with high- or low-affinity flanking sequences. Despite seemingly arbitrary sequence compositions, these shape profiles showcase the Scr protein's ability to fine-tune its target site through DNA shape. Our MD simulation protocol can also extend to larger Hox systems such as the Homothorax-Exd-Hox complex (68).

CONCLUSION

We revealed the significance of the His-12 residue and its protonation state in the Scr Hox protein. Our findings indicate that His-12 plays a crucial role in the interaction between the linker region of the Hox protein and DNA targets by regulating the MGW shape profile, which is essential for the specificity of Hox binding. The uniqueness of His-12 lies in its ability to adopt two protonation states at biological pH. Our MD simulations demonstrated that the differences in behavior between the two protonation states of His-12 can not only account for its preference toward different core motifs but also provide insights into its preference for different DNA targets of high and low affinity.

DATA AND CODE AVAILABILITY

Simulation protocols, binding free energy input files, and additional supplemental information can be found at and retrieved from https://doi.org/10.6084/m9.figshare.22695778.v7. The available README file describes the usage of each input file and the order in which each input file was used.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2023.12.013.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Slattery, M., T. Zhou, …, R. Rohs. 2014. Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* 39:381–399.

2. Inukai, S., K. H. Kock, and M. L. Bulyk. 2017. Transcription factor–DNA binding: beyond binding site motifs. *Curr. Opin. Genet. Dev.* 43:110–119.

3. von Hippel, P. H., W. A. Rees, ..., K. S. Wilson. 1996. Specificity mechanisms in the control of transcription. *Biophys. Chem.* 59:231–246.

4. Kribelbauer, J. F., C. Rastogi, ..., R. S. Mann. 2019. Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annu. Rev. Cell Dev. Biol.* 35:357–379.

5. Lambert, S. A., A. Jolma, ..., M. T. Weirauch. 2018. The human transcription factors. *Cell.* 172:650–665.

6. Gordân, R., N. Shen, ..., M. L. Bulyk. 2013. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.* 3:1093–1104.

7. Rohs, R., X. Jin, ..., R. S. Mann. 2010. Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* 79:233–269.

8. Slattery, M., T. Riley, ..., R. S. Mann. 2011. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell.* 147:1270–1282.

9. Levy, Y., J. N. Onuchic, and P. G. Wolynes. 2007. Fly-casting in protein− DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* 129:738–739.

10. Chiu, T. P., S. Rao, ..., R. Rohs. 2017. Genome-wide prediction of minor-groove electrostatic potential enables biophysical modeling of protein–DNA binding. *Nucleic Acids Res.* 45:12565–12576.

11. Afek, A., and D. B. Lukatsky. 2013. Positive and negative design for nonconsensus protein-DNA binding affinity in the vicinity of functional binding sites. *Biophys. J.* 105:1653–1660.

12. Yella, V. R., D. Bhimsaria, ..., M. Bansal. 2018. Flexibility and structure of flanking DNA impact transcription factor affinity for its core motif. *Nucleic Acids Res.* 46:11883–11897.

13. Hancock, S. P., S. Stella, ..., R. C. Johnson. 2016. DNA sequence determinants controlling affinity, stability and shape of DNA complexes bound by the nucleoid protein Fis. *PLoS One.* 11, e0150189.

14. Ghoshdastidar, D., and M. Bansal. 2022. Flexibility of flanking DNA is a key determinant of transcription factor affinity for the core motif. *Biophys. J.* 121:3987–4000.

15. Horton, C. A., A. M. Alexandari, ..., P. M. Fordyce. 2023. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science (Washington, D.C.).* 381, eadd1250.

16. McGinnis, W., M. S. Levine, ..., W. J. Gehring. 1984. A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature.* 308:428–433.

17. Pearson, J. C., D. Lemons, and W. McGinnis. 2005. Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* 6:893–904.

18. Mann, R. S. 1995. The specificity of homeotic gene function. *Bioessays.* 17:855–863.

19. Joshi, R., J. M. Passner, ..., R. S. Mann. 2007. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell.* 131:530–543.

20. Li, S., and M. Hong. 2011. Protonation, tautomerization, and rotameric structure of histidine: a comprehensive study by magic-angle-spinning solid-state NMR. *J. Am. Chem. Soc.* 133:1534–1544.

21. Chiu, T. P., J. Li, ..., R. Rohs. 2022. It is in the flanks: Conformational flexibility of transcription factor binding sites. *Biophys. J.* 121:3765–3767.

22. Bradshaw, E. M., D. G. Sanford, ..., W. W. Bachovchin. 2004. T antigen origin-binding domain of simian virus 40: determinants of specific DNA binding. *Biochemistry.* 43:6928–6936.

23. Eliseo, T., I. E. Sánchez, ..., D. O. Cicero. 2009. Indirect DNA readout on the protein side: coupling between histidine protonation, global structural cooperativity, dynamics, and DNA binding of the human papillomavirus type 16 E2C domain. *J. Mol. Biol.* 388:327–344.

24. Onufriev, A. V., and E. Alexov. 2013. Protonation and pK changes in protein–ligand binding. *Q. Rev. Biophys.* 46:181–209.

25. Chattopadhyay, A., A. Esadze, ..., J. Iwahara. 2016. NMR scalar couplings across intermolecular hydrogen bonds between zinc-finger histi-

26. Lundbäck, T., S. van den Berg, and T. Härd. 2000. Sequence-specific DNA binding by the glucocorticoid receptor DNA-binding domain is linked to a salt-dependent histidine protonation. *Biochemistry.* 39:8909–8916.

27. Chang, Y. P., M. Xu, ..., X. S. Chen. 2013. Mechanism of origin DNA recognition and assembly of an initiator-helicase complex by SV40 large tumor antigen. *Cell Rep.* 3:1117–1127.

28. Langella, E., R. Improta, and V. Barone. 2004. Checking the pH-induced conformational transition of prion protein by molecular dynamics simulations: effect of protonation of histidine residues. *Biophys. J.* 87:3623–3632.

29. Sarkar, C. A., K. Lowenhaupt, ..., D. A. Lauffenburger. 2002. Rational cytokine design for increased lifetime and enhanced potency using pH-activated "histidine switching". *Nat. Biotechnol.* 20:908–913.

30. Kampmann, T., D. S. Mueller, ..., B. Kobe. 2006. The role of histidine residues in low-pH-mediated viral membrane fusion. *Structure.* 14:1481–1487.

31. Park, J. S., T. H. Han, ..., Y. W. Cho. 2006. N-acetyl histidine-conjugated glycol chitosan self-assembled nanoparticles for intracytoplasmic delivery of drugs: endocytosis, exocytosis and drug release. *J. Contr. Release.* 115:37–45.

32. Hansen, A. L., and L. E. Kay. 2014. Measurement of histidine pKa values and tautomer populations in invisible protein states. *Proc. Natl. Acad. Sci. USA.* 111:E1705–E1712.

33. Riley, T. R., M. Slattery, ..., H. J. Bussemaker. 2014. SELEX-seq: a method for cha[[[racterizing the complete repertoire of binding site preferences for transcription factor complexes. *In* Hox Genes Springer, pp. 255–278.

34. Schrodinger LLC. 2015. The PyMOL Molecular Graphics System, Version 1.8.

35. Abe, N., I. Dror, ..., R. S. Mann. 2015. Deconvolving the recognition of DNA shape from sequence. *Cell.* 161:307–318.

36. Cramer, P. 2021. AlphaFold2 and the future of structural biology. *Nat. Struct. Mol. Biol.* 28:704–705.

37. Lu, X. J., and W. K. Olson. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 31:5108–5121.

38. Adams, P. D., P. V. Afonine, ..., P. H. Zwart. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66:213–221.

39. Maier, J. A., C. Martinez, ..., C. Simmerling. 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theor. Comput.* 11:3696–3713.

40. Ivani, I., P. D. Dans, ..., M. Orozco. 2016. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods.* 13:55–58.

41. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An N• log (N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

42. Hess, B., H. Bekker, ..., J. G. E. M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

43. Smith, L. J., X. Daura, and W. F. van Gunsteren. 2002. Assessing equilibration and convergence in biomolecular simulations. *Proteins.* 48:487–496.

44. Daura, X., W. F. van Gunsteren, and A. E. Mark. 1999. Folding–unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. *Proteins.* 34:269–280.

45. Søndergaard, C. R., M. H. M. Olsson, ..., J. H. Jensen. 2011. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values. *J. Chem. Theor. Comput.* 7:2284–2295.

46. Olsson, M. H. M., C. R. Søndergaard, ..., J. H. Jensen. 2011. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theor. Comput.* 7:525–537.

47. Anandakrishnan, R., B. Aguilar, and A. V. Onufriev. 2012. H++ 3.0: automating p K prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 40:W537–W541.

48. Wang, L., M. Zhang, and E. Alexov. 2016. DelPhiPKa web server: predicting p K a of proteins, RNAs and DNAs. *Bioinformatics.* 32:614–615.

49. Baker, N. A., D. Sept, …, J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.* 98:10037–10041.

50. Dolinsky, T. J., P. Czodrowski, …, N. A. Baker. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 35:W522–W525.

51. Lavery, R., and H. Sklenar. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91.

52. Kumari, R., R. Kumar, …, A. Lynn. 2014. g_mmpbsa A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* 54:1951–1962.

53. Cuervo, A., P. D. Dans, …, L. Fumagalli. 2014. Direct measurement of the dielectric polarization properties of DNA. *Proc. Natl. Acad. Sci. USA.* 111:E3624–E3630.

54. Gilson, M. K., and B. H. Honig. 1986. The dielectric constant of a folded protein. *Biopolymers.* 25:2097–2119.

55. Kollman, P. A., I. Massova, …, T. E. Cheatham, 3rd. 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889–897.

56. Vascon, F., M. Gasparotto, …, I. Righetto. 2020. Protein electrostatics: From computational and structural analysis to discovery of functional fingerprints and biotechnological design. *Comput. Struct. Biotechnol. J.* 18:1774–1789.

57. Ladam, F., and C. G. Sagerström. 2014. Hox regulation of transcription: more complex (es). *Dev. Dynam.* 243:4–15.

58. Lin, M., and J. T. Guo. 2019. New insights into protein–DNA binding specificity from hydrogen bond based comparative study. *Nucleic Acids Res.* 47:11103–11113.

59. Garvie, C. W., and C. Wolberger. 2001. Recognition of specific DNA sequences. *Mol. Cell.* 8:937–946.

60. Crocker, J., N. Abe, …, D. L. Stern. 2015. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell.* 160:191–203.

61. Mikles, D. C., V. Bhat, …, A. Farooq. 2013. pH modulates the binding of early growth response protein 1 transcription factor to DNA. *FEBS J.* 280:3669–3684.

62. Besold, A. N., A. A. Oluyadi, and S. L. J. Michel. 2013. Switching metal ion coordination and DNA Recognition in a Tandem CCHHC-type zinc finger peptide. *Inorg. Chem.* 52:4721–4728.

63. Delker, R. K., V. Ranade, …, R. S. Mann. 2019. Low affinity binding sites in an activating CRM mediate negative autoregulation of the Drosophila Hox gene Ultrabithorax. *PLoS Genet.* 15, e1008444.

64. Jolma, A., T. Kivioja, …, J. Taipale. 2010. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res.* 20:861–873.

65. Esmaeeli, R., A. Bauzá, and A. Perez. 2023. Structural predictions of protein–DNA binding: MELD-DNA. *Nucleic Acids Res.* 51:1625–1636.

66. Yang, L., Y. Orenstein, …, R. Rohs. 2017. Transcription factor family-specific DNA shape readout revealed by quantitative specificity models. *Mol. Syst. Biol.* 13:910.

67. Dantas Machado, A. C., B. H. Cooper, …, R. Rohs. 2020. Landscape of DNA binding signatures of myocyte enhancer factor-2B reveals a unique interplay of base and shape readout. *Nucleic Acids Res.* 48:8529–8544.

68. Mann, R. S., and M. Affolter. 1998. Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* 8:423–429.

69. Cooper, B. H., T. P. Chiu, and R. Rohs. 2022. Top-Down Crawl: a method for the ultra-rapid and motif-free alignment of sequences with associated binding metrics. *Bioinformatics.* 38:5121–5123.