A Simulation Study of the Effect of Naturally Occurring Point Mutations on the SRY-DNA Complex

Published as part of The Journal of Physical Chemistry virtual special issue "Doros N. Theodorou Festschrift".

Angeliki-Georgia Nasou,[†] Evangelia Pantatosaki,[†] and George K. Papadopoulos*

Read Online

ACCESS | 📊 Metrics & More | 📰 Article Recommendations | 🆘 Supporting Information

ABSTRACT: Molecular dynamics (MD) simulations were conducted in order to investigate the effect of the naturally occurring point mutations of the transcription factor (TF) sex-determining region Y (SRY) on the structure and dynamics of the SRY-DNA complex. The normal SRY, along with the two mutants I13T and G40R, comprising point mutations on the SRY chain, which have been clinically identified in patients with sex developmental disorders, were modeled as DNA complexes. Our modeling work aims at elucidating atomic-level structural determinants of the aberrant SRY-DNA complexation by means of μs-long MD. The results suggest that the observed disorders brought about by the G40R-DNA and I13T-DNA may arise predominantly from the destabilization of the complex being in accord with *in vitro* assays found elsewhere and from modifications of the DNA bending as revealed in this study. Comparative potential of mean force computations, over a sequence of short separation distances for the three complexes, verified a higher stability of the normal SRY-DNA. Examining the way the SRY mutations modulate the SRY-DNA complex dynamics at the microscopic level is important also toward elucidating molecular determinants of function for proteins capable of binding to DNA.
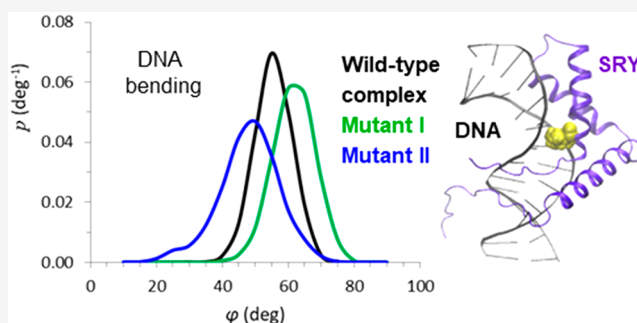
1. INTRODUCTION

Transcription factors (TFs) are DNA-binding proteins acting as major regulators of gene expression and cellular differentiation. The binding of TFs to particular DNA regions enables the recruitment of several accessory proteins forming supramolecular complexes, which guide RNA polymerase II to the target gene initiation site to enable transcription.[1] The exact molecular mechanism of transcriptional complex assembly is still under investigation. Several TFs have been reported to bind and bend the DNA helical axis,[2,3] thereby facilitating the establishment of precise geometry in the resulting higher-order structures; among them, is the sex-determining region Y (SRY) transcription factor, which acts as a master switch in mammalian sex determination[4] by binding to the minor groove of the DNA molecule, at the testes-specific enhancer element of the *SOX9* gene. This procedure forces the bending of DNA as a result of the partial intercalation of a hydrophobic amino acid (isoleucine) of SRY, in the interior of the DNA double helix.[5,6]

Therefore, establishment of a particular SRY-DNA spatial arrangement facilitates the transcription of the *SOX9* gene, thus upregulating the transcription factor SOX9. The latter is involved in a cascade of molecular transcriptional processes that can inhibit the creation of a female reproductive system during embryonic development, also promoting the development of male sexual organs.[7]

Point mutations (single amino acid substitutions) in the normal SRY polypeptide chain disrupt directly the ability of SRY to promote the assembly of a functional transcriptional complex, causing loss of the SRY function and hence several sex developmental disorders in humans.[8−11] In a general context, aberrant formation of transcriptional complexes has been linked to multiple congenital disorders and diseases, including cancer.[12,13] To elucidate the structure of transcriptional complexes, nuclear magnetic resonance (NMR) experiments have been performed on SRY-DNA complexes.[5,14,15] Also, computational molecular dynamics (MD) has been used to study the wild-type (normal) SRY-DNA.[16−18]

Understanding how TF mutations modulate the TF-DNA assembly is important for providing novel therapeutic routes that target the transcriptome.[19−21] In this work, we employ

fine-grained μs-long MD simulations to study the effect of SRY point mutations on the structure and dynamics of the SRY-DNA complex, aiming at elucidating atomic-level structural determinants of the aberrant SRY-DNA complexation. For this, we modeled the normal SRY-DNA complex and two mutant variants based on SRY mutations that have been clinically identified in sex-reversed patients, i.e., patients with complete 46,XY gonadal dysgenesis.

The paper is organized as follows: In the modeling section the *in silico* construction of the mutant SRY-DNA complexes is described, followed by the details on the employed computer simulation methods. In the Results and Discussion section the spatial arrangement of the biomolecules in the supramolecular assembly, the packing or sitting of residues at the SRY-DNA interface, and the deformation of the DNA are detailed, complemented by potential of mean force computations to assess the interaction of the two biomolecules in the complex at various separation distances.

2. MODELING DETAILS

2.1. Biomolecules Reconstruction. The initial structure of the wild-type SRY-DNA complex was taken from NMR spectroscopy data deposited in the Protein Data Bank (PDB code 1j46).[14] This NMR structure consists of a 14mer double-stranded DNA bound by the DNA-binding domain of SRY, i.e., a central fragment (85 amino acids) of the full-length (204 amino acids) SRY. This central fragment spans residues 57–140 of the full-length protein which correspond to residues 2–85 in the present numbering scheme, plus a methionine residue at position 1. The sequence of amino acids in the central fragment, hereafter called SRY, can be seen in Figure S1. In addition, a 14mer double-stranded DNA of the same sequence as the SRY-bound DNA was reconstructed in atomistic detail, using the Web 3DNA 2.0 server,[22] to simulate the nucleic acid free in ionic solution for comparison. The sequence of bases in the DNA strands can be seen in Figure S2.

Atomistic reconstruction of the initial configurations for the two mutant SRY-DNA complexes was performed using as a template the wild-type SRY-DNA complex,[14] wherein we created two point mutations depicted in Figure 1 that have been clinically identified in sex-reversed XY females, i.e., patients with complete 46,XY gonadal dysgenesis.[8,10,23,24]

In the G40R mutant, the glycine at position 40 (G40 or Gly40) on the normal SRY chain was replaced by arginine (R40 or Arg40). In the I13T mutant, the isoleucine at position 13 (I13 or Ile13) on the normal SRY chain was replaced by threonine (T13 or Thr13). For the notation of amino acids the one- or three-letter standard nomenclature is used.[25]

2.2. Molecular Dynamics. Molecular dynamics simulations were conducted in the isobaric, isothermal ensemble (P = 1 bar, T = 310 K) to monitor the spatiotemporal evolution of the normal and mutant SRY-DNA complexes in ionic aqueous solution. For the numerical integration of the differential equations of motion the GROMACS 2016.1 package[26–28] was used. The CHARMM27 force field was employed for the nucleic acid[29,30] and the proteins;[31,32] this force field has been recently reported to provide stable atomistic μs-long MD trajectories for the DNA and good agreement with experimentally determined DNA flexibility parameters.[33] The rhombic dodecahedron simulation box dimensions were set to ensure at least 2 nm thickness for the solvent surrounding the SRY-DNA complex in all directions. The TIP3P water model[34]
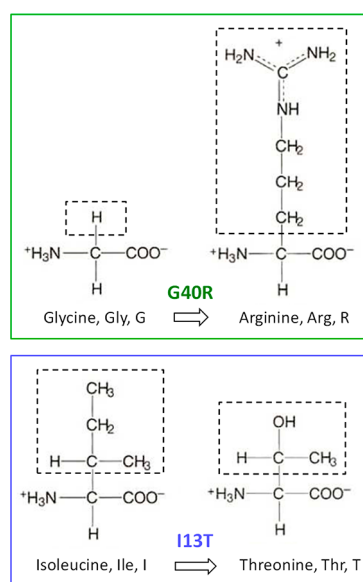
Figure 1. Modeled point mutations in the normal SRY chain for the G40R (green frame) and I13T (blue frame) mutants. Glycine at position 40 and isoleucine at position 13 on the SRY chain were replaced by arginine and threonine, respectively. The amino acid side chain is noted with a dashed rectangle.

was employed; randomly selected water molecules were replaced by the appropriate number of Na cations to neutralize the system. For the equilibration of the system, prior to the MD production runs, we followed the equilibration protocol described in our previous work.[35]

The Berendsen algorithm was utilized to allow for fluctuating the pressure tensor (diagonal components) isotropically under the constraint of a constant trace (the pressure),[36] with a time constant of 2.0 ps; the isothermal compressibility of water was set equal to 4.5×10^{-5} bar^{-1}. For the thermostating a velocity-rescaling scheme[37] was employed on two separate groups, namely, the protein with the DNA, and the ionic solvent, with a time constant of 0.1 ps. For the dispersion interactions a shifted cutoff of 1 nm was applied. Bond lengths were constrained as usual in most biomolecular systems.[38,39] The smooth particle-mesh Ewald method[40] was employed for the computation of the electrostatic interactions. MD trajectories were produced up to 1 μs, using a time step of 2 fs, and recorded every 10 ps. Visualization of the MD trajectory and image rendering was performed ultimately by means of VMD.[41]

In the postprocessing stage, the geometrical characteristics of the DNA molecules were extracted from the MD trajectories using Curves+.[42] The two terminal DNA base pairs were excluded from the calculations due to spurious values caused by base-pair fraying events, a known artifact in nucleic acids simulations.[43] The computation of the number of close contacts (ncc) between each protein amino acid and the DNA was performed using the cutoff distance shown in Table S1. A close contact is defined here as an amino acid atom lying at a distance less than the specified cutoff value from either one or multiple atoms of the DNA. Also, the time average of the minimum distance between all atoms in each pair of residues (inter-residue minimum distance) was computed for all 85 residues in the protein chain, over the last 300 ns of the MD trajectory; calculations were performed by means of Conan.[44]

2.3. Potential of Mean force. *2.3.1. Basics.* A rigorous way to assess the binding strength of the DNA-SRY complex is to sample the potential of mean force (PMF),[45] defined as

$$-\beta w^{(n)}(\mathbf{r}^n) \equiv \ln g^{(n)}(\mathbf{r}^n) \tag{1}$$

It gives rise to the mean force,

$$\langle \mathbf{f}_j^{(n)} \rangle_{N-n} = -\nabla_{\mathbf{r}_j} w^{(n)}(\mathbf{r}^n), \quad j = 1, \, ..., \, n \tag{2}$$

exerted on particle j among a subset of n particles, which are located at fixed positions specified by the $3n$ coordinates: $\mathbf{r}^n \equiv \{\mathbf{r}_1, ..., \mathbf{r}_n\}$, out of the total of N particles of the system in the configuration space: $\mathbf{r}^N \equiv \{\mathbf{r}_1, ..., \mathbf{r}_N\}$. The said force represents an average over all configurations of the $N - n$ particles: $n + 1$, ..., N not included in the fixed set $\mathbf{r}^n$. As seen by eq 1, the PMF involves the n-particle distribution function,

$$g^{(n)}(\mathbf{r}^n) = \frac{\rho_{NPT}^{(n)}(\mathbf{r}^n)}{\prod_{i=1}^{n} \rho_{NPT}^{(1)}(\mathbf{r}_i)} \tag{3}$$

with $\rho_{NPT}^{(n)}(\mathbf{r}^n)$ being the n-particle density in the isobaric−isothermal ensemble given by eq 4. Interestingly, eq 3 constitutes a metric of the deviation of a fluid structure from randomness.

$$\rho_{NPT}^{(n)}(\mathbf{r}^n) = \frac{N!}{(N-n)!}$$
$$\frac{\int_0^\infty \int \exp[-\beta \mathcal{V}(\mathbf{r}^N) - \beta PV] \, dV \, d\mathbf{r}_{n+1}...d\mathbf{r}_N}{\int_0^\infty \exp(-\beta PV) \, Z(N, V, T) \, dV} \tag{4}$$

Obviously, in the homogeneous system of this study, $\rho_{NPT}^{(1)}(\mathbf{r}_i) = \rho = N/V$, $i = 1, ..., n$; hence the denominator in eq 3 becomes ρ^n.

$Z(N,V,T)$ is the configuration integral in the current volume V. In words, eq 4 gives the probability density of the energy states containing n particles located at n fixed positions specified by Cartesian coordinates in the configuration space and averaged over all coordinates but $\mathbf{r}^n$, and volumes ranging from zero to infinite. If one is interested in a set of generalized coordinates instead, a proper Jacobian determinant enters eq 4. Multiplying both parts of eq 4 by the integrands, $d\mathbf{r}^n \equiv d\mathbf{r}_1...d\mathbf{r}_n$, and integrating, we get $\int \rho_{NPT}^{(n)}(\mathbf{r}^n) \, d\mathbf{r}^n = N!/(N-n)!$, that is, the number of ways that N particles can be placed at the n mentioned positions.

2.3.2. Implementation. The n-particle probability density cannot be obtained from an exact ensemble average since states with low probability (higher energy states) are hardly visited during the time captured by computer simulations. Therefore, the entailed configuration integral in eq 4 is misevaluated. For this, the system must be biased to sample these regions. The most usual algorithmic technique for such a non-Boltzmann sampling is known as umbrella sampling[46] (details in section 3.2).

3. RESULTS AND DISCUSSION

3.1. Conformational analysis. *3.1.1. Wild-Type SRY-DNA.* The MD results for the wild-type SRY-DNA complex showed that the SRY remains minor-groove-bound to DNA throughout the μs-long created sequence of configurations. A representative configuration of the modeled SRY-DNA complex, also showing the three α-helices (α1, α2, and α3)

and the N- and C-terminal parts of the SRY chain, appears in Figure 2. The three α-helical core configuration remains stable
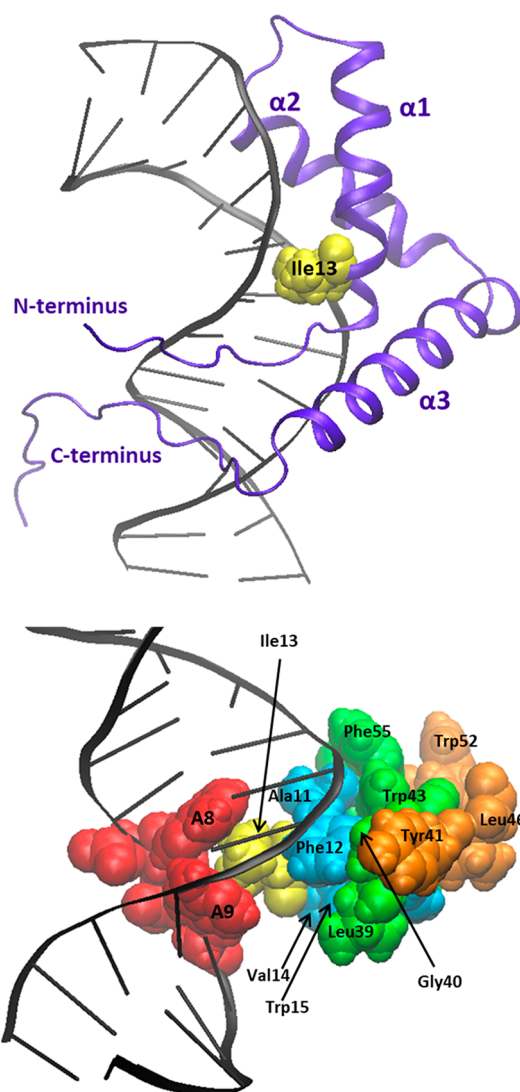
Figure 2. Snapshots of the SRY-DNA extracted from the MD trajectory. Top: transcriptional complex of SRY (violet) with DNA (black). The α-helices (α1, α2, and α3) and N- and C-terminal regions of SRY are noted. Also, isoleucine at position 13 (Ile13) on α1, which partially intercalates the DNA, is shown (yellow spheres). Bottom: A8pA9 step (red spheres) wherein the Ile13 partially intercalates (see text). The three contiguous layers of SRY hydrophobic amino acids (see text) are shown: layer I (cyan), layer II (green), and layer III (orange). Phe54 of layer II is behind Trp43 and hence is not visible.

throughout the whole MD trajectory, exhibiting a root-mean-square deviation (RMSD) of 0.18 ± 0.01 nm from its initial configuration, whereas the N- and C-termini show increased mobility.

As shown in Figure 2, the adenine (A) bases 8 and 9 unstack and roll open their planes to introduce a localized sharp bend known as DNA kink;[3] this is attributed to the partial intercalation of isoleucine (Ile13) in the DNA double helix at the A8pA9 step, i.e., between A8 and A9. The computed average roll angle at the kinked site shown in Figure 3 was found to have a positive high value (50 ± 5°), reflecting the opening of the angle between the base planes toward the minor
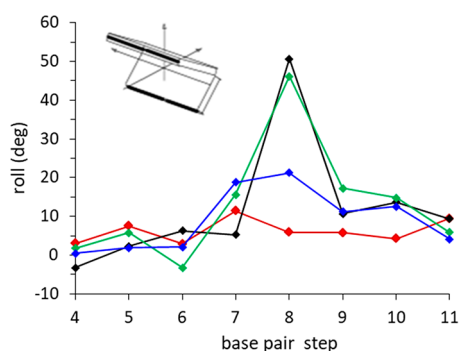
Figure 3. Computed average roll angle between the base planes of a base pair step (inset sketch) for the DNA bound to the normal SRY (black), and mutants G40R (green) and I13T (blue). The red data correspond to the free DNA in solution, for comparison. Lines are to guide the eye.

groove, while the minor groove exhibits a local expansion calculated to 0.41 nm with respect to its size in the free DNA. The overall bending angle of the DNA helical axis, induced by the SRY, was computed to $53 \pm 6°$, close to the reported experimental value of $54 \pm 2°$.[14] For comparison, the unbound DNA in solution showing a helical axis bending angle of $11 \pm 6°$ and a roll angle of $6 \pm 5°$ at the A8pA9 step is also depicted in Figure 3.

The partial intercalation of Ile13 is "supported" by the synergistic effect of multiple nonpolar and aromatic SRY residues forming three contiguous hydrophobic layers, as shown in Figure 2. Layer I (cyan) consists of four residues (Ala11, Val14, Phe12, and Trp15) belonging to α1 of SRY. Layer 2 (green) is formed cooperatively by three residues (Leu39, Gly40, and Trp43) belonging to α2 and two residues (Phe54, Phe55) located on α3. Layer III (orange) consists of Tyr41 of α2, and Leu46 and Trp52 of α3. The identified layers are consistent with the residue clusters described as "hydrophobic wedge" and "aromatic buttress" in earlier experimental studies.[5,47] The average inter-residue minimum distances for all residues in the SRY chain can be seen in the computed correlation contact map (Figure S3).

In addition to the hydrophobic layers, other hydrophilic SRY residues help to stabilize the resulting complex. As shown in a representative MD conformation (Figure S4, top), the lateral regions of Ile13 at the kinked site are occupied on one side by Met9 and Arg17 and on the other side by Ser33 and Ser36, all located in proximity to the DNA backbone. Furthermore, stabilization of the bound SRY chain is achieved by several cationic arginines (Arg7, Arg17, Arg21, Arg66) and lysines (Lys37, Lys44, Lys51), which exhibit high probability of lying close to the negatively charged phosphate backbone of the DNA (Figure S4).

Apart from the above SRY residues interacting closely with the central part of the DNA at the intercalation site, i.e., DNA base pairs (bp) 8 and 9, the protein also exhibits close contacts (distance of interacting pairs of <0.4 nm) with the two ends of the DNA 14mer, coded here as DNA-end-I and DNA-end-II, which correspond to DNA areas spanning bp 1−7 and bp 10−14, respectively, as shown by the computed relative frequency of the number of close contacts between the protein and each DNA end in Figure S5. Specifically, the DNA-end-I was found to interact closely with the protein N-terminus (residues 1−8), C-terminus (residues 69−85), and part of helix a3 (residues 56−68), whereas the DNA-end-II exhibited close contacts

mainly with the SRY residues located at the second half of **a1** (residues 18−26) and the beginning of **a2** (residues 31−35).

3.1.2. Mutant G40R-DNA. The results for this mutated residue showed that the bulky arginine, which replaces the short glycine on layer II, lies at closer distances exhibiting more close contacts with the DNA atoms (Figure S6). We found that arginine, due to its large size, creates a steric clash to the adjacent residues, verifying a hypothesis about the said mutant first made by Clore and co-workers,[5] and induces the following destabilizing effects as opposed to the wild-type complex: (i) The neighboring residues belonging to hydrophobic layers II (Leu39, Trp43) and III (Tyr41) are displaced further away from the DNA, as shown by the probability density of their minimum distances and the relative frequency of the number of close contacts with the DNA in Figures 4 and S7. (ii) Met9 on the lateral side of the kinked site moves to longer distances and loses a few close contacts with the DNA (Figure S8). (iii) Whereas the hydrophilic lysines 37 and 44 in the normal complex are bound to the DNA backbone near the kinked site, they move away from DNA in the mutant (Figure S9).

In addition, the residues of layer I, i.e., Val14 and Trp15 on **a1**, are found at larger minimum distances from Arg17 on **a1**, compared to the normal complex, as shown by the inter-residue minimum distances in the computed correlation contact maps in Figure 5. Also, the inter-residue distances of the N-terminus (Met9, Pro8, and Arg7 with Val5, Arg4 and Asp3, respectively) were found to decrease in the G40R mutant. Moreover, as shown in Figure S5, the ncc between the N-terminus and the DNA-end-I is diminished, as compared to the normal complex; this is also the case for part of helix **a3** (residues 56−68). Contrarily, an increase in the ncc between the C-terminus and the DNA-end-I was detected, whereas the overall ncc between the DNA-end-I and the mutant remains practically the same, as compared to the normal complex. Furthermore, an increase in the ncc between the DNA-end-II and the mutant is observed, as shown in the same figure; this is the result of close contacts that are formed between the DNA-end-II and the second half of **a1** (residues 18−26), the β-turn (residues 27−30), the beginning of **a2** (residues 31−35), and more importantly, the C-terminus of the mutant (Figure S5). The overall conformation of the G40R versus the SRY chain appears more compact, as reflected by the computed radius of gyration shown in Figure 6.

The computed roll at A8pA9 step (Figure 3) has a high positive value ($46 \pm 5°$), close to the roll value of the normal complex, denoting that the DNA remains kinked and the partial intercalation of Ile13 is preserved for the time length of 1 μs. The Ile13 preserves practically the same number of contacts with the DNA and binds to it at slightly shorter distances, as shown by the computed probability density distributions of the minimum distances and the relative frequency of the ncc between the two species (Figure S10). The bending angle of the DNA helical axis induced by the G40R mutant was found slightly higher, with an average value of $59 \pm 6°$, thus indicating that the above-described conformational changes of the G40R increase the bending of the oligonucleotide.
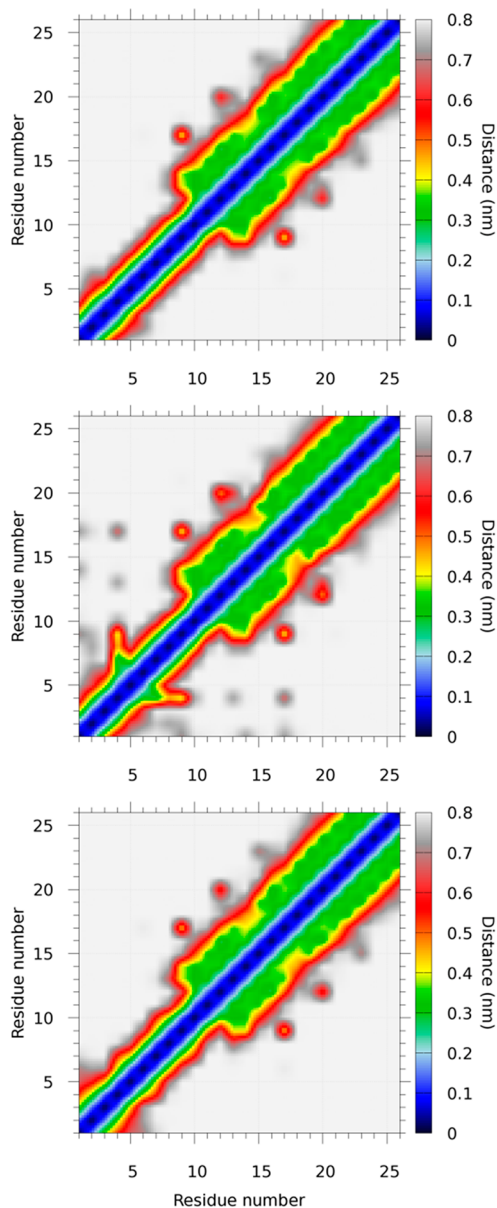
3.1.3. Mutant I13T-DNA. The replacement of the nonpolar isoleucine with the polar threonine at the intercalation site was found to destabilize the complex and alter the DNA deformation. We found that threonine exits from the interior of the DNA double helix moving away from adenine base 9, as shown by the probability density distribution of minimum
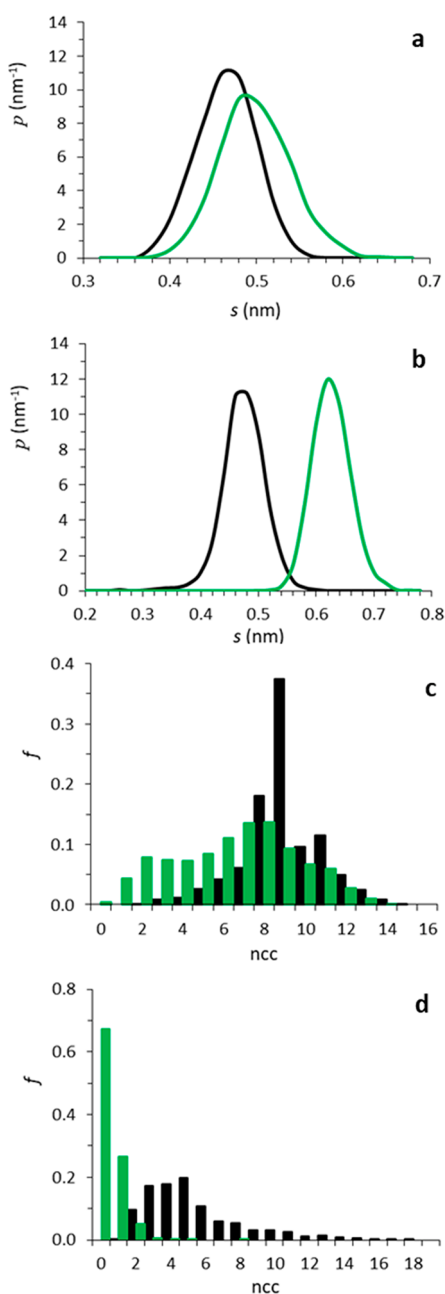
Figure 4. Probability density distribution of the minimum distance (a, b) and relative frequency of the number of close contacts (c, d) between DNA and Leu39 (a, c) and DNA and Tyr41 (b, d), presented for the normal SRY (black) and the G40R mutant (green).

Figure 5. Correlation contact maps for the wild-type SRY (top) and the mutants G40R (middle) and I13T (bottom), presenting the average inter-residue minimum distance for residues 1−26 in the protein chain. Distances greater than 0.8 nm were truncated to 0.8 nm for clarity.
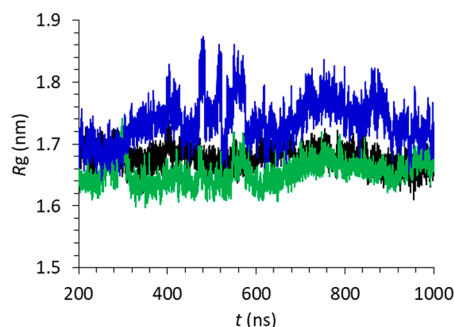
Figure 6. Time evolution of the radius of gyration for the DNA-bound protein modeled variants: wild-type SRY (black) and mutants G40R (green) and I13T (blue).

distance and the relative frequency of the number of close contacts of threonine with the base atoms (Figure S11); this finding verifies a hypothesis about the said mutant first made by Clore and co-workers.[5] In particular, while being outside of the DNA helix, threonine forms on average two strong hydrogen bonds (HB) with the water molecules, with an average HB length and angle of 0.28 nm and 9°, respectively; details on the computation scheme of HBs can be found in previous studies.[35,48]
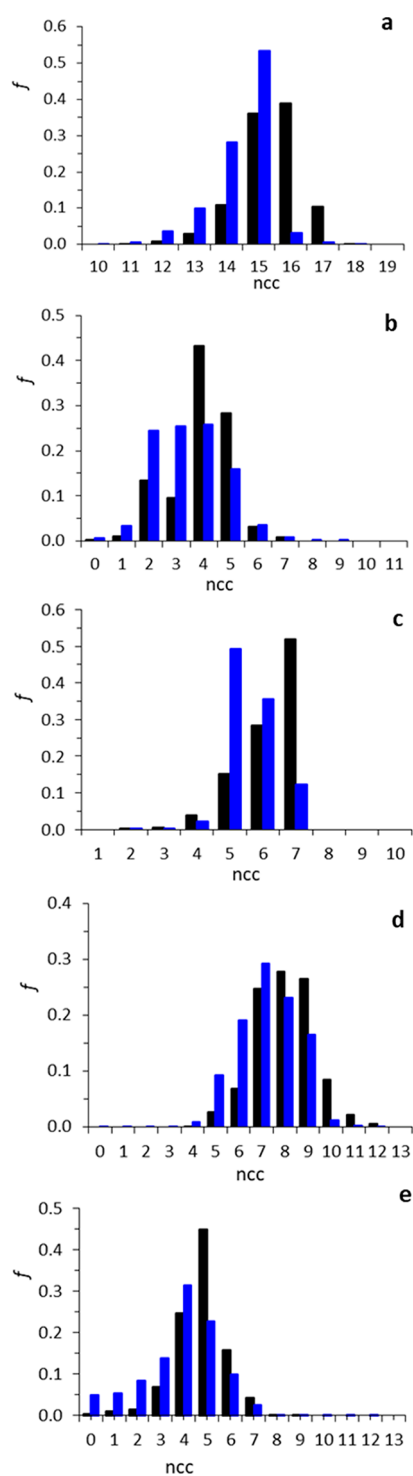
The displacement of threonine out of the helix affects the residues Met9 and Arg17 located proximal to the A8pA9 step and also the Phe12, Trp15, Gly40, Trp43, and Phe54 comprising the contiguous hydrophobic layers, which all lose some of their contacts with the DNA as seen in Figures 7 and

Figure 7. Relative frequency of the number of close contacts between the DNA and Phe12 (a), Trp15 (b), Gly40 (c), Trp43 (d), or Phe54 (e), presented for the normal SRY (black) and the I13T mutant (blue).

S12, with respect to the normal complex, thereby destabilizing the SRY-DNA interface. In addition, as shown in the correlation contact map in Figure 5, the inter-residue minimum distances between Val14, Trp15 and Arg17, Asp18 increase in the I13T mutant.

Our computations also show an increase in the ncc between the I13T and the DNA-end-I and wider ncc distributions, as compared to the normal SRY, indicating labile contacts in the

complex interface (Figure S5). This is especially evident for the interaction of the DNA-end-I with the N-terminus, the C-terminus, and part of helix **a3** (residues 56−68). As shown in the same figure, an increase in the ncc is also observed between the DNA-end-II and the beginning of **a2** (residues 31−35), whereas the ncc formed between the DNA-end-II and the second half of **a1** (residues 18−26) remains practically the same as in the normal complex. In addition, the I13T C-terminus was found to exhibit zero close contacts with the DNA-end-II, similar to the normal SRY and in contrast to the G40R mutant.

Furthermore, the interactions of the two key residues, Val5 and Phe54, which are implicated in other clinically identified mutations resulting in severe sex developmental disorders[10,23] are altered in the I13T-DNA complex as follows: (i) Phe54 appears to have less contacts with the DNA and displaces to longer distances as shown in Figures 7 and S13. (ii) The hydrophobic cluster created by the close packing of Val5 against the aromatic Tyr69, Tyr72, and Tyr74 at the SRY terminal tails in the normal complex (Figure S14) breaks apart in the mutant, as shown by the computed inter-residue distances for the protein (Figure S3), thus allowing Val5 and Tyr74 to move to respectively shorter and longer distances from DNA (Figure S15). The above destabilizing effects are reflected in an increase in the radius of gyration for the mutant protein compared to the wild-type as can be seen in Figure 6, indicating the acquisition of a less compact chain.

As shown in Figure 3, the computed roll value ($21 \pm 14°$) at the A8pA9 step is decreased, due to the removal of threonine from the kinked site, thus allowing for the bases to partially restack toward recovering the unperturbed geometry of the unbound DNA. Consequently, the bending angle of the DNA helical axis is reduced and computed to $45 \pm 9°$. The bending angle distributions for the DNA in the three complexes, computed for the last 300 ns of the MD trajectory, can be seen in Figure 8. The broader distribution for the I13T-bound DNA indicates a more labile and hence destabilized complex.
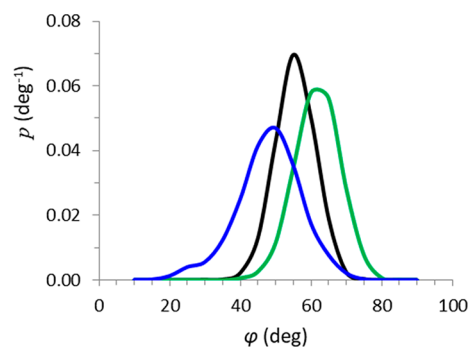
Figure 8. Probability density distributions of the bending angle of the DNA helical axis, induced by the normal SRY (black) and the mutants G40R (green) and I13T (blue).

MD results show local distortions for the I13T-bound DNA, namely, moderately perturbed Watson−Crick base pairing for base pair 8, as depicted by the computed opening angle between the opposing bases of a base pair, shown in Figure 9. The opening angle is defined here as the relative angle $\Delta\theta = \theta - \theta_0$, where θ_0 is the value corresponding to the relaxed opening angle of the bases in the unperturbed DNA. Positive and negative $\Delta\theta$ values correspond to base opening into the major and minor groove, respectively. The opening angle (−22
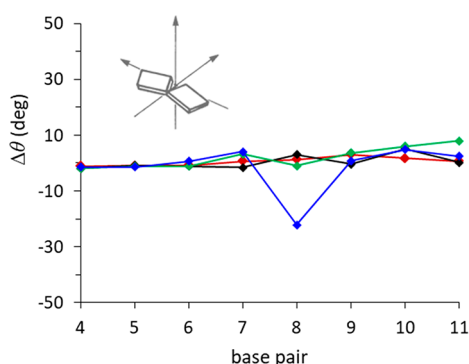
Figure 9. Computed average opening angle (see text) between the opposing bases in the base pairs (inset sketch) for the DNA bound to the normal SRY (black), and mutants G40R (green) and I13T (blue). The red data correspond to the free DNA in solution, for comparison. Lines are to guide the eye.

± 16°) of base pair 8 is caused by the strong interaction of threonine with adenine 8; although threonine exits from the interior of the helix, it moves to shorter distances and forms more close contacts with adenine 8 (Figure S16), as compared to isoleucine in the normal complex, thereby destabilizing the position of adenine 8 in the helical duplex. Nevertheless, the DNA base pairing is preserved, since the rupture of hydrogen bonds in a base pair (and the subsequent flipping of a base out of the helix) requires a relative opening angle of at least ±50° compared to the unperturbed DNA.[49]

Notably, threonine exits from the interior of the double helix but the DNA remains still bent during the simulated time of 1 μs, indicating that the partial intercalation of the protein chain to DNA is not the sole factor contributing to DNA bending; that is to say, the deformation of the nucleic acid is the result of the overall arrangement of the SRY chain along the binding surface. The difference in the DNA bending angle (computed to approximately 10°) as compared to the normal complex may result in large displacements of the accessory proteins that the SRY recruits on site, to the extent that the formation of a transcriptionally competent nucleoprotein complex may no longer be possible.

3.2. Potential of Mean Force. The system was retained at $P = 1$ bar and $T = 310$ K for all the MD runs. As starting configuration for the SRY-DNA complex, the equilibrated structure extracted from the μs-long simulations was used within a rectangular simulation box ($10 \times 9 \times 16$ nm³), with the DNA helical axis oriented perpendicular to the z-axis. In fact, a rigorous way to locate such a reaction coordinate would be the computation of committor probabilities as discussed in the paper of Roux and co-workers.[50] Water molecules with the appropriate number of Na cations were also added to neutralize the system, and an introductory simulation run of 12 ns was performed to equilibrate the system temperature, pressure, and solvent density.

Because the system under study is homogeneous and also isotropic, a reasonable choice of ξ is the distance $\mathbf{r}_{j,12} = \mathbf{r}_{c2} - \mathbf{r}_{c1} = r_{j,12}$ between the position vectors of two central atoms on SRY and DNA for the particular region j, where $j \in [1, 60]$. The next region is sampled by increasing the distance by 0.08 nm each time. At the postprocessing stage, the results are combined to extract ultimately the (unbiased) PMF profile; an optimum way for the latter is being performed via the weighted histogram analysis method (WHAM).[51]

Thus, instead of the mathematically strict $\rho^{(2)}(\mathbf{r}_{12})$, the normalized frequency $\mathcal{P}(\xi)$ of finding the system in the interval of configuration space specified by ξ and $\xi + d\xi$ is usually sampled, where ξ can be any fixed set of Cartesian or generalized coordinates. In particular, one is usually interested in calculating the PMF profile over some path (reaction coordinate), which is split into regions with each of them covering a small interval over the whole range of ξ. Then, these regions are sampled individually to obtain the distributions $\mathcal{P}(\xi)$.

To sample the aforementioned regions, the reaction coordinate was restrained and pulled to values $\xi_{0,j}$, that is to say, the target separation distances of the two macromolecules along the z-axis by applying the harmonic potential

$$u_j(\xi) = K(\xi - \xi_{0,j})^2/2 \tag{5}$$

with a strength constant K. It is stressed that this potential must be added to the potential energy term of eq 4 so that eventually after simple algebra the unbiased distributions of interest, and hence PMFs, are extracted. A K value of 3000 kJ mol⁻¹ nm⁻² was used for all complexes; in the normal complex, however, the simulations were repeated using a constant of 20 000 kJ mol⁻¹ nm⁻² (narrower distributions and increased number of j-regions to preserve overlapping) and a simulation time of 30 ns in each region to ensure no significant differences in the PMF profile (Figure S17).

The computed PMF sequences for the three complexes are presented in Figure 10; the distributions $\mathcal{P}(\xi)$ are shown in Figures S18−S20.
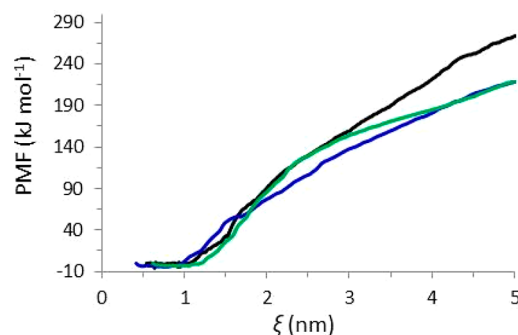

Figure 10. Potential of mean force computed along the reaction coordinate (see text) for the normal SRY-DNA (black), G40R-DNA (green), and I13T-DNA (blue) complexes.

By substituting eqs 3 and 4 to eq 1, we get the PMF, $w^{(2)}(\mathbf{r}_1,\mathbf{r}_2) = w^{(2)}(\mathbf{r}_{12}) = G(\xi)$, as a function of a local partition function (integrating out all degrees of freedom but ξ); thus, a Gibbs free energy, $G_j(\xi)$, for each region along the reaction coordinate is readily achieved. Figure 10 provides such an evolution of PMF for various values of distance between DNA and the protein. It also shows according to eq 2 the mean force acting on, say, the central atom c2 of SRY by the c1 of DNA, averaged over the configurations of all system atoms (degrees of freedom) except those held at fixed distance, c1 and c2.

Indeed, the work of this force for moving reversibly the SRY from the separation distance at the region j up to infinite, under constant P and T, equals the corresponding Gibbs energy of the complex dissociation. From the same figure it is seen that this force becomes weak faster with varying separation distance for both mutants, compared to the wild-

type complex, denoting thus a less stable system; this is in agreement with the preceding conformational analysis, being also consistent with *in vitro* assays[8,10,24] reporting significantly reduced binding affinity of the G40R and I13T mutants.

As the separation distance increases, the 3-helical core of the SRY is displaced away from the DNA, whereas the disordered N- and C-termini remain bound to the nucleic acid by acquiring elongated conformations; this is especially true for the longer C-terminal tail. Both N- and C-termini are rich in disorder-promoting amino acids,[52,53] i.e., lysine, proline, glutamine, aspartic acid, and arginine (Figure S1), favoring interactions with the water environment, thus allowing for extended chain conformations. The elongation of the protein chain is reflected in the gradual increase of the protein radius of gyration along the reaction coordinate, as shown in Table S2; the corresponding representative SRY-DNA configurations are shown in Figure 11. As the 3-helical core of the SRY moves
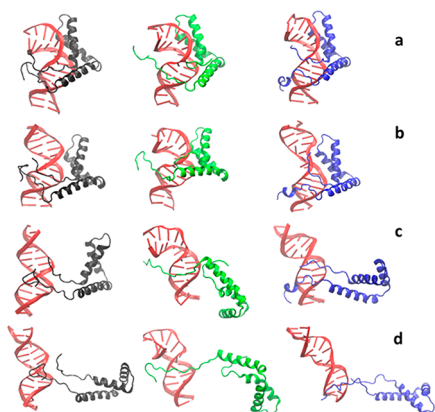
Figure 11. Representative SRY-DNA configurations from umbrella sampling simulations in the initial window (a) and along the reaction coordinate for ξ values of 1.5 (a), 2.5 (b), and 3.5 (c) nm, presented for the normal SRY (black) and mutants G40R (green) and I13T (blue).

away from the DNA, both the DNA roll angle at A8pA9 step and DNA bending angle decrease for all complexes (Table S2). The A8 and A9 bases restack and the DNA straightens up toward recovering the unperturbed geometry of the unbound DNA in solution.

4. CONCLUSIONS

Molecular dynamics simulations of the normal SRY-DNA complex, along with two mutant complexes comprising the naturally occurring mutations G40R and I13T on the SRY chain, showed that a point mutation in the SRY induces conformational changes, which can severely impact the DNA-binding performance.

We found that destabilization of the G40R-DNA interface occurs near the kinked site, while the overall conformation of the G40R appears more compact (radius of gyration decreases), albeit the DNA bending increases. On the contrary, in the case of I13T-DNA, the DNA bending diminishes as the mutated residue threonine at the kinked site exits the double helix interior, inducing local distortions in the geometry of DNA and destabilization of the complex interface. In addition, the conformation of the I13T terminal tails is affected, whereas the radius of gyration of the mutant protein is increased. Furthermore, comparative potential of mean force computa-

tions, over a sequence of short separation distances for the three complexes, showed a higher stability of the normal SRY-DNA.

Our results elucidate the G40R and I13T induced molecular-level structural changes in the normal SRY-DNA complex, verifying earlier speculations about the said mutants made by Clore and co-workers. Also, our findings suggest that the sex-developmental disorders brought about by the G40R and I13T mutants may arise predominantly from the destabilization of their complex with the DNA, being consistent with *in vitro* assays showing impaired DNA binding performance for both mutants, as compared to the normal SRY.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.2c04852.

Additional structural and conformational data and illustrations for the SRY and DNA; modeling details; correlation contact maps; probability density distributions of the minimum distance and relative frequency of the number of close contacts between DNA and various residues; additional potential of mean force computations and frequency distributions along the reaction coordinate (PDF)

■ AUTHOR INFORMATION

Corresponding Author

George K. Papadopoulos − *School of Chemical Engineering, National Technical University of Athens, 157 80 Athens, Greece;* ⓞ orcid.org/0000-0001-5935-5461; Email: gkpap@chemeng.ntua.gr

Authors

Angeliki-Georgia Nasou − *School of Chemical Engineering, National Technical University of Athens, 157 80 Athens, Greece*

Evangelia Pantatosaki − *School of Chemical Engineering, National Technical University of Athens, 157 80 Athens, Greece;* ⓞ orcid.org/0000-0002-2063-097X

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpcb.2c04852

Author Contributions

[†]A.-G.N. and E.P. contributed equally.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 5th ed.; W H Freeman. 2002.

(2) Bewley, C. A.; Gronenborn, A. M.; Clore, G. M. Minor groove-binding architectural proteins: Structure, Function, and DNA Recognition. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 105−131.

(3) Werner, M. H.; Gronenborn, A. M.; Clore, G. M. Intercalation, DNA kinking, and the control of transcription. *Science* **1996**, *271*, 778−784.

(4) Kashimada, K.; Koopman, P. Sry: the master switch in mammalian sex determination. *Development* **2010**, *137*, 3921−3930.

(5) Werner, M. H.; Huth, J. R.; Gronenborn, A. M.; Clore, G. M. Molecular Basis of Human 46X,Y Sex Reversal Revealed from the Three-Dimensional Solution Structure of the Human SRY-DNA Complex. *Cell* **1995**, *81*, 705−714.

(6) King, C.-Y.; Weiss, M. A. The SRY high-mobility group box recognizes DNA by partial intercalation in the minor groove - a topological mechanism of sequence specificity. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11990−11994.

(7) De Santa Barbara, P.; Bonneaud, N.; Boizet, B.; Desclozeaux, M.; Moniot, B.; Sudbeck, P.; Scherer, G.; Poulat, F.; Berta, P. Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Müllerian hormone gene. *Mol. Cell. Biol.* **1998**, *18*, 6653−6665.

(8) Pontiggia, A.; Rimini, R.; Harley, V. R.; Goodfellow, P. N.; Lovell-Badge, R.; Bianchi, M. E. Sex-reversing mutations affect the architecture of SRY-DNA complexes. *EMBO J.* **1994**, *13*, 6115−6124.

(9) Hawkins, J. R.; Taylor, A.; Berta, P.; et al. Mutational analysis of SRY: nonsense and missense mutations in XY sex reversal. *Hum. Genet.* **1992**, *88*, 471−474.

(10) Rimini, R.; Pontiggia, A.; Spada, F.; Ferrari, S.; Harley, V. R.; Goodfellow, P. N.; Bianchi, M. E. Interaction of Normal and Mutant SRY Proteins with DNA. *Philos. Trans. R. Soc. London B* **1995**, *350*, 215−220.

(11) Jäger, R. J.; Harley, V. R.; Pfeiffer, R. A.; Goodfellow, P. N.; Scherer, G. A familial mutation in the testis-determining gene SRY shared by both sexes. *Hum. Genet.* **1992**, *90*, 350−355.

(12) Engelkamp, D.; van Heyningen, V. Transcription factors in disease. *Curr. Opin Genet Dev.* **1996**, *6*, 334−342.

(13) Lee, T. I.; Young, R. A. Transcriptional Regulation and its Misregulation in Disease. *Cell* **2013**, *152*, 1237−1251.

(14) Murphy, E. C.; Zhurkin, V. B.; Louis, J. M.; Cornilescu, G.; Clore, G. M. Structural Basis for SRY-dependent 46-X,Y Sex Reversal: Modulation of DNA Bending by a Naturally Occurring Point Mutation. *J. Mol. Biol.* **2001**, *312*, 481−499.

(15) Li, B.; Phillips, N. B.; Jancso-Radek, A.; Ittah, V.; Singh, R.; Jones, D. N.; Haas, E.; Weiss, M. A. SRY-directed DNA Bending and Human Sex Reversal: Reassessment of a Clinical Mutation Uncovers a Global Coupling between the HMG Box and its Tail. *J. Mol. Biol.* **2006**, *360*, 310−328.

(16) Bouvier, B.; Lavery, R. A Free Energy Pathway for the Interaction of the SRY Protein with Its Binding Site on DNA from Atomistic Simulations. *J. Am. Chem. Soc.* **2009**, *131*, 9864−9865.

(17) Bouvier, B.; Zakrzewska, K.; Lavery, R. Protein-DNA Recognition Triggered by a DNA Conformational Switch. *Angew. Chem., Int. Ed.* **2011**, *50*, 6516−6518.

(18) Singh, R. M.; Sasikala, W. D.; Mukherjee, A. Molecular Origin of DNA Kinking by Transcription Factors. *J. Phys. Chem. B* **2015**, *119*, 11590−11596.

(19) Kazantsev, A. G.; Hersch, S. M. Drug Targeting of Dysregulated Transcription in Huntington's Disease. *Prog. Neurobiol.* **2007**, *83*, 249−259.

(20) Chen, A.; Koehler, A. N. Transcription Factor Inhibition: Lessons Learned and Emerging Targets. *Trends Mol. Med.* **2020**, *26*, 508−516.

(21) Bhagwat, A. S.; Vakoc, C. R. Targeting Transcription Factors in Cancer. *Trends Cancer* **2015**, *1*, 53−65.

(22) Li, S.; Olson, W. K.; Lu, X. J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* **2019**, *47*, W26−W34.

(23) Harley, V. R.; Jackson, D. I.; Hextall, P. J.; et al. DNA Binding Activity of Recombinant SRY from Normal Males and XY Females. *Science* **1992**, *255*, 453−456.

(24) Haqq, C. M.; King, C.-Y.; Ukiyama, E.; Falsafi, S.; Haqq, T. N.; Donahoe, P. K.; Weiss, M. A. Molecular Basis of Mammalian Sexual Determination: Activation of Müllerian Inhibiting Substance Gene Expression by SRY. *Science* **1994**, *266*, 1494−1500.

(25) Joint Commission on Biochemical Nomenclature (JCBN).. Nomenclature and Symbolism for Amino Acids and Peptides. *Eur. J. Biochem.* **1984**, *138*, 9−37.

(26) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19−25.

(27) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(28) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(29) Foloppe, N.; MacKerell, A. D., Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86−104.

(30) MacKerell, A. D., Jr.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21*, 105−120.

(31) MacKerell, J. A. D., Jr.; Feig, M.; Brooks, C. L., III Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(32) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(33) Minhas, V.; Sun, T.; Mirzoev, A.; Korolev, N.; Lyubartsev, A. P.; Nordenskiöld, L. Modeling DNA Flexibility: Comparison of Force Fields from Atomistic to Multiscale Levels. *J. Phys. Chem. B* **2020**, *124*, 38−49.

(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(35) Pantatosaki, E.; Papadopoulos, G. K. Binding Dynamics of siRNA with Selected Lipopeptides: A Computer-Aided Study of the Effect of Lipopeptides' Functional Groups and Stereoisomerism. *J. Chem. Theory Comput.* **2020**, *16*, 3842−3855.

(36) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(37) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(38) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(39) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.

(40) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(41) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.

(42) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917−5929.

(43) Zgarbová, M.; Otyepka, M.; Šponer, J.; Lankas, F.; Jurečka, P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014**, *10*, 3177−3189.

(44) Mercadante, D.; Gräter, F.; Daday, C. CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps. *Biophys. J.* **2018**, *114*, 1267−1273.

(45) Theodorou, D. N. Progress and Outlook in Monte Carlo Simulations. *Ind. Eng. Chem. Res.* **2010**, *49*, 3047−3058.

(46) Torrie, G. M.; Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **1974**, *28*, 578−581.

(47) Racca, J. D.; Chen, Y.-S.; Maloy, J. D.; Wickramasinghe, N.; Phillips, N. B.; Weiss, M. A. Structure-Function Relationships in Human Testis-determining Factor SRY. *J. Biol. Chem.* **2014**, *289*, 32410−32429.

(48) Kolokathis, P. D.; Pantatosaki, E.; Papadopoulos, G. K. Atomistic Modeling of Water Thermodynamics and Kinetics within MIL-100(Fe). *J. Phys. Chem. C* **2015**, *119*, 20074−20084.

(49) Giudice, E.; Várnai, P.; Lavery, R. Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. *Nucleic Acids Res.* **2003**, *31*, 1434−1443.

(50) Pan, A. C.; Sezer, D.; Roux, B. Finding Transition Pathways Using the String Method with Swarms of Trajectories. *J. Phys. Chem. B* **2008**, *112*, 3432−3440.

(51) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(52) Quaglia, F.; Mészáros, B.; Salladini, E.; Hatos, A.; Pancsa, R.; Chemes, L. B.; et al. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res.* **2022**, *50*, D480−D487.

(53) Campen, A.; Williams, R. M.; Brown, C. J.; Meng, J.; Uversky, V. N.; Dunker, A. K. TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.* **2008**, *15*, 956−963.

📖 Recommended by ACS

A Dynamic Switch in Inactive p38γ Leads to an Excited State on the Pathway to an Active Kinase

Phillip C. Aoto, Peter E. Wright, *et al.*

DECEMBER 03, 2019
BIOCHEMISTRY

READ 🗗

Binding Models of Aβ42 Peptide with Membranes Explored by Molecular Simulations

Ke Wang, Wensheng Cai, *et al.*

AUGUST 19, 2022
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 🗗

Electronic Polarization at the Interface between the p53 Transactivation Domain and Two Binding Partners

Alexsandra N. Corrigan and Justin A. Lemkul

JUNE 24, 2022
THE JOURNAL OF PHYSICAL CHEMISTRY B

READ 🗗

Ycs4 Subunit of *Saccharomyces cerevisiae* Condensin Binds DNA and Modulates the Enzyme Turnover

Rupa Sarkar, Valentin V. Rybenkov, *et al.*

NOVEMBER 01, 2021
BIOCHEMISTRY

READ 🗗

Get More Suggestions >