A SIMULATION STUDY OF THE EFFECT OF NATURALLY OCCURRING POINT MUTATIONS ON THE SRY-DNA COMPLEX

A.G. Nasou[1], E. Pantatosaki[1,2], G.K. Papadopoulos[1,3,*]

[1]School of Chemical Engineering, National Technical University of Athens, 157 80 Athens, Greece

[2]Department of Biomedical Engineering, University of West Attica, 122 43 Athens, Greece

[3]Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge MA 02139, US

(*gkpap@chemeng.ntua.gr)

ABSTRACT

We present our recently published results [1] on 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, by means of μs-long Molecular dynamics (MD) simulations. TFs are DNA-binding proteins that act as major regulators of gene expression and cellular differentiation. The wild-type SRY along with two mutants, the I13T and G40R, which comprise point mutations on the SRY chain (clinically identified in patients with sex developmental disorders), were modeled as DNA complexes. 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 our study. In addition, 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. Ultimately, 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, which in a broader context may contribute to novel therapeutic routes that target the transcriptome.

KEYWORDS: Transcription Factors, Gene Expression, Mutations, Computer Simulation

INTRODUCTION

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. 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, 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 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. 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.

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. In a general context, aberrant formation of transcriptional complexes has been linked to multiple congenital disorders and diseases, including cancer.

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.

METHODOLOGY

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) [2]. In addition, a 14mer double-stranded DNA of the same sequence as the SRY-bound DNA was reconstructed in atomistic detail to simulate the nucleic acid free in ionic solution for comparison. 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, wherein we created two point mutations: in the G40R mutant, glycine at position 40 on the SRY chain was replaced by arginine; in the I13T mutant, isoleucine at position 13 on the SRY chain was replaced by threonine.

Molecular Dynamics. Molecular dynamics simulations were conducted in the isobaric, isothermal ensemble to monitor the spatiotemporal evolution of the normal and mutant SRY-DNA complexes in ionic aqueous solution. For the equilibration of the system, prior to the MD production runs, we followed the equilibration protocol described in our previous works [1, 3].

Potential of Mean force. An indicative way for assessing the binding strength of the DNA-SRY complex is to sample the potential of mean force (PMF), defined as

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

It gives rise to the mean force,

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

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

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

with $p_{NPT}^{(n)}(\mathbf{r}^n)$ 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.

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

$Z(N,V,T)$ 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, $d\mathbf{r}^n \equiv d\mathbf{r}_1 \ldots d\mathbf{r}_n$, and integrating, we get

$$\int p_{NPT}^{(n)}(\mathbf{r}^n)d\mathbf{r}^n = N!/(N-n)!\,,$$

that is, the number of ways that N particles can be placed at the n mentioned positions. 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 by a non-Boltzmann sampling (umbrella sampling).

RESULTS & DISCUSSION

1. Conformational analysis.

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 1. The three α-helical core configuration remains stable 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.

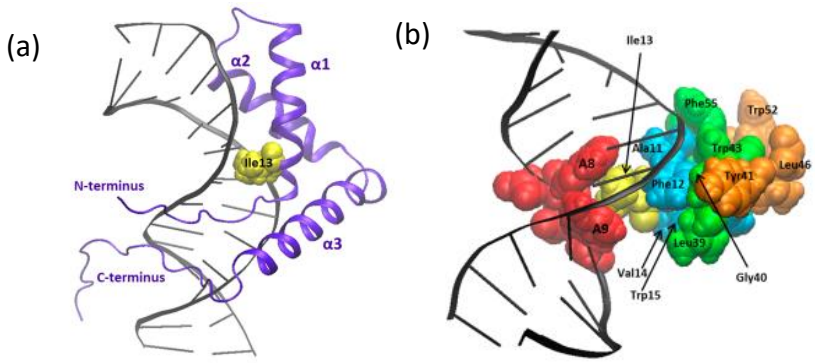

Figure 1. *Snapshots of the SRY-DNA extracted from the MD trajectory. (a) transcriptional complex of SRY (violet) with DNA (black). The α-helices (α1, α2, and α3) and N- and C-terminal regions of SRY are noted. Isoleucine at position 13 (Ile13) on α1, is shown (yellow spheres). (b): A8pA9 step (red spheres) wherein the Ile13 partially intercalates. The three contiguous layers of SRY hydrophobic amino acids are shown: layer I (cyan), layer II (green), and layer III (orange). Phe54 of layer II is behind Trp43 and hence is not visible.*

As shown in Figure 1b, the adenine (A) bases 8 and 9 unstack and roll open their planes to introduce a localized sharp bend known as DNA kink; this is attributed to the partial intercalation of isoleucine (Ile13) in the DNA double helix at the A8pA9 step. The computed average roll angle at the kinked

site was found to have a positive high value (50 ± 5°), reflecting the opening of the angle between the base planes toward the minor 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 ± 6°, close to the reported experimental value of 54 ± 2° [2]. 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 1. In addition to the hydrophobic layers, our computations showed that other hydrophilic SRY residues help to stabilize the resulting complex [1].

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. We found that arginine, due to its large size, creates a steric clash to the adjacent residues, 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. (ii) Met9 on the lateral side of the kinked site moves to longer distances and loses a few close contacts with the DNA. (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. Computations of the probability density of the residue minimum distances and the relative frequency of the number of close contacts with the DNA can be found in Ref. [1]. The computed roll at A8pA9 step has a high positive value (46 ± 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 bending angle of the DNA helical axis induced by the G40R mutant was found slightly higher, with an average value of 59 ± 6°, thus indicating that the above-described conformational changes of the G40R increase the bending of the oligonucleotide.

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 [1]. Consequently, the bending angle of the DNA helical axis is reduced and computed to 45 ± 9°. 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.

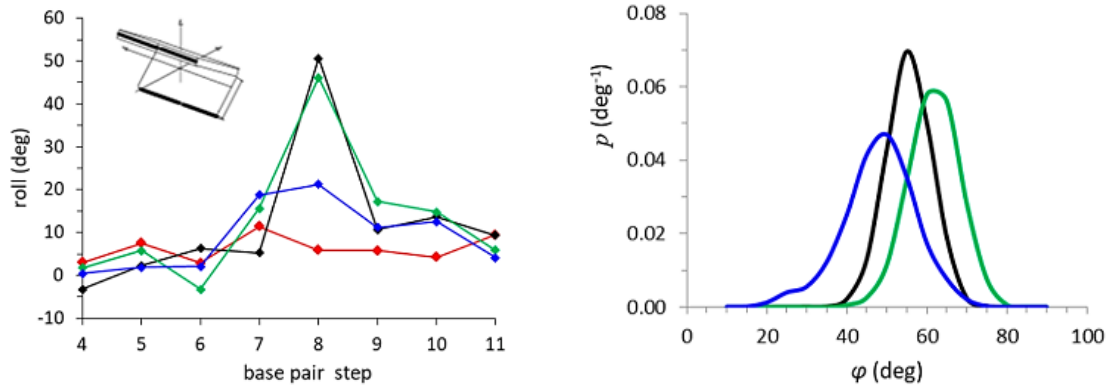

Figure 2. *(left) Computed average roll angle between the base planes of a base pair step (inset sketch) for the DNA bound to the normal SRY (black), and mutants G40R (green) and I13T (blue). The red data correspond to the free DNA in solution, for comparison. Lines are to guide the eye. (right) Probability 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).*

The bending angle distributions for the DNA in the three complexes, computed for the last 300 ns of the MD trajectory, were also computed [1]. The broader distribution for the I13T-bound DNA indicates a more labile and hence destabilized complex. 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.

2. Potential of Mean Force Calculations

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 × 9 × 16 nm^3), with the DNA helical axis oriented perpendicular to the z-axis.

Because the system under study is homogeneous and also isotropic, a reasonable choice of ξ is the distance 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 post-processing 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). Thus, the normalized frequency 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. To sample the aforementioned regions, the reaction coordinate was restrained and pulled to target separation distances of the two macromolecules along the z-axis by applying a harmonic potential [1]. 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. Comparative potential of mean force computations showed a higher stability of the normal SRY-DNA (Figure 3).

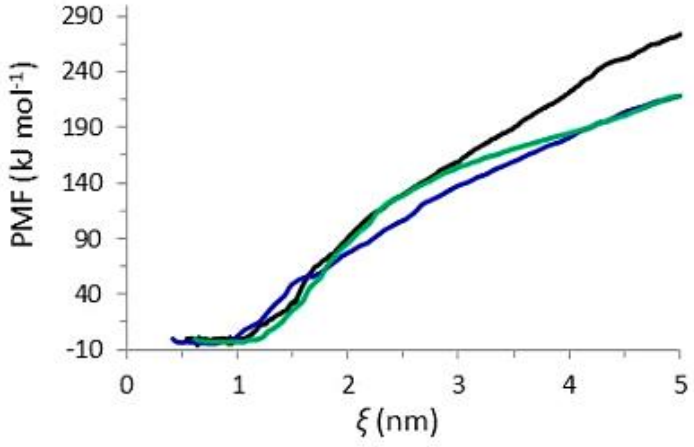

***Figure 3**. Potential of mean force computed along the reaction coordinate (see text) for the normal SRY-DNA (black), G40R-DNA (green), and I13T-DNA (blue) complexes.*

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, i.e., lysine, proline, glutamine, aspartic acid, and arginine 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 [1]. As the 3-helical core of the SRY moves away from the DNA, both the DNA roll angle at A8pA9 step and DNA bending angle decrease for all complexes. The A8 and A9 bases restack and the DNA straightens up toward recovering the unperturbed geometry of the unbound DNA in solution [1].

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 computations, 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.

ACKNOWLEDGMENTS

G.K.P. acknowledges support by the European Union through a fellowship under Grant H2020-MSC-IF (Project 796794-ENGEMED) and by the Greek Research and Technology Network for CPU-time provided in the National HPC facility “ARIS”. E.P. acknowledges a grant of Scientific Excellence by the Hellenic Foundation for Research and Innovation (Project B.1003-ENGETACT).

REFERENCES

[1] Nasou A. G., Pantatosaki E., Papadopoulos G. K. (2022). J. Phys. Chem. B, 126(44), 8921–8930.
[2] 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.
[3] 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.