3D Structural Integrity and Interactions of Single-Stranded Protein-Binding DNA in a Functionalized Nanopore

Mohammed Arif I. Mahmood,†‡§ Waqas Ali,†‡§ Ashfaq Adnan,‖ and Samir M. Iqbal*†‡§‖⊥

†Nano-Bio Lab, ‡Department of Electrical Engineering, §Nanotechnology Research Center, Shimadzu Institute for Research Technologies, ‖Department of Mechanical and Aerospace Engineering, †Department of Bioengineering, ⊥Joint Graduate Studies Committee of Bioengineering Program, University of Texas at Arlington and University of Texas Southwestern Medical Center at Dallas, University of Texas at Arlington, Arlington, Texas 76019, United States

ABSTRACT: Biomarker-binding nucleotide sequences, like aptamers, have gained recent attention in cancer cell isolation and detection works. Self-assembly and 3D conformation of aptamers enable them to selectively capture and bind diseased cells and related biomarkers. One mode of utilizing such an extraordinary selective property of the aptamers is by grafting these in nanopores. Coating the inside walls of the nanopore with biomarker specific ligands, like DNA, changes the statistics of the dynamic translocation events. When the target protein passes through the nanopore, it interacts with ligand coated inside the nanopore, and the process alters the overall potential energy profile which is essentially specific to the protein detected. The fundamental goal in this process is to ensure that these detection motifs hold their structure and functionality under applied electric field and experimental conditions. We report here all-atom molecular dynamics simulations of the effects of external electric field on the 3D conformation of such DNA structures. The simulations demonstrate how the grafted moieties affect the translocation time, velocity, and detection frequency of the target molecule. We also investigated a novel case of protein translocation, where DNA is prebound to the protein. As model, a thrombin-specific G-quartet and thrombin pair was used for this study.

1. INTRODUCTION

Many diseases can be diagnosed using one or multiple biomarkers. These biomarkers may consist of alien entities inside a host body or disease-induced proteins overexpressed or downregulated by the host itself as part of its defense mechanisms. These proteins become available in the circulatory bloodstream at early stages of the disease and work as disease precursors. Detection and identification of these proteins is important for diagnosis and prognostic approaches henceforth.

Over the past several years, significant progress has been made on nanopore-based DNA detection technologies.1,2 Recently, it has been shown that such nanopore-based detection can also be used toward the detection of protein biomarkers.3 In a nanopore-based system, DNA traveling through functionalized nanopore has been shown to alter due to ligand specific affinity.1 We’ve shown here that proteins allowed to pass through a nanopore whose interior wall is functionalized with the protein-specific DNA would exhibit similar discriminatory effects. In other words, proteins would slow down or even chemically bind to the surface of the nanopore depending on the nature of stimuli applied inside the nanopore (e.g., electric field, mechanical forces, etc.).

DNA and protein are two intertwined moieties by virtue of their functions in cellular mechanisms. Proteins are synthesized by DNA transcription; on the other hand, certain proteins play significant roles in regulation of such transcriptions. This regulation is accomplished by selectivity between the DNA segments and proteins. This selectivity is a useful property that can be used in vitro for detection of certain proteins. Nanopore is a highly suitable yet simple platform for utilizing such extraordinary property. Compared to the gel electrophoresis, the widely employed protein detection scheme, a nanopore-based method, not only promises easy and quick detection of as few as a single copy of rare biomarkers without need of expert supervision, but it also eliminates the requirement of a strict lab environment.4 Therefore, it can be perceived why detection and identification of protein or DNA based on the translocation behavior through a nanopore has received recent growing attention.4 In this method, protein or DNA is allowed to pass through a nanopore of comparable size in an ionic solution and under applied electric bias (Figure 1).

The ionic current is measured, and parameters like translocation time, velocity, and current dip are calculated as electronic signatures. When proteins pass through the functionalized nanopores, these create statistically different ionic current dips than those through bare nanopores. Unraveling the mystery of how protein transport takes place in a cellular environment has opened up new windows toward such methodologies. Enzyme assisted protein translocation through nanopores has indicated promise of protein sequencing through nanopores.5 However, many biophysical phenomena are yet to
be understood for a complete understanding of working at such small dimensions.

Although nanopore-based measurements are simpler and robust, this method comes with certain challenges. One of these is the uncertainty that comes from the relative orientation and conformation of DNA structure. We know that, in their natural state, nucleic acid sequences can fold into various secondary and tertiary 3D structures (hairpin loop, G-quartet, etc.), creating binding sites for other biomolecules. For example, a G-quartet is a naturally occurring DNA structure involved in regulating genetic expressions in cellular machinery. These act as translational repressors of specific target mRNAs. On the other hand, these 3D DNA structures have also been employed, in vitro, to screen and isolate target biomolecules from random populations. Biomarkers such as proteins also fold into 3D structures in their natural states (determined mostly by their amino acid sequences). This property gives them specific biological functionalities. Both of these 3D entities, when in proximity, can create strong binding, even though complementarity is not clearly evident.

Target specific DNA sequences or aptamers have been reported to successfully isolate cancerous biomarkers and even tumor cells from the samples with high selectivity. In these cases, devices were typically functionalized with the specific aptamer sequences. When a sample was passed over/through the device, target molecules that came in the vicinity of the aptamers got loosely attached via nonspecific electrostatic attractions. However, as determined by the disassociation constant, these got detached from the binding site within a short period and through a process called “facilitated diffusion” these finally reached the target segment of the nucleotide sequence where multiple binding sites from both molecules “complemented” each other, hence creating a much stronger bond (Figure 1B). If the bond was strong enough against thermal and mechanical perturbations, the target molecule got immobilized.

The binding interactions are, however, dependent on the external force competitively applied over the protein–DNA complex. A simple mechanical force can dislodge the protein bound to the DNA. Shear stress applied by fluid flow has been shown to release cells that were bound through such protein–DNA interactions. In addition, perturbation of the DNA structure may also be a contributory factor, as has been shown with joule heating by microheaters fabricated beneath the surface. A major challenge for the robust development of a selective method is that it requires control of multiple variables to acquire reasonable affinity. Calibrations/optimizations of these variables via experiments only are very tedious and time-consuming. Numerical methods such as molecular dynamics simulations can define the key variables of the method.

Here, single-molecule-level DNA stability is characterized inside nanopores using classical molecular dynamics simulations. After developing the molecular models in section 2, the stability of the DNA structure under different applied electric fields is quantified and the effect of the DNA structure on the translocation of protein through nanopores is simulated in section 3. We also investigated how such stability might affect the translocation events. A short simulation is also presented to show the effect of prebound DNA to the protein on its translocation through the nanopore.

2. MOLECULAR DYNAMICS SIMULATION DETAILS

The stability of DNA structure under an applied electric field, the mechanism of protein translocation through DNA functionalized nanopore, and the travel of protein–DNA complex through a bare nanopore were studied using all-atom MD simulations. The freely available massively parallel MD simulation package known as nanoscale molecular dynamics (NAMD) was used to perform the simulation. Visual molecular dynamics (VMD) was used to analyze and visualize the output. All simulations were performed on the Lonestar supercomputer cluster at Texas Advanced Computing Center (TACC). Two cases were studied: first, the stability of the DNA structure in a typical experimental condition was investigated, i.e., in ionic solution and under applied bias; second, the translocation events of the protein through the nanopore for three different cases were simulated, showing how the DNA structure affected the translocation time.

2.1. Molecular Models. To examine the stability of the 3D DNA structure, a protein databank (PDB) file of an X-ray resolved nucleotide sequence was obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) PDB. It is
known that the quadruplex topology of DNA depends on both the number and nature of the nucleotides participating in the construction of the motif. We found that the 11-mer single-stranded DNA (ssDNA) (PDB id 2AVJ) was a suitable match for our study. The X-ray resolved crystal structure of the DNA was in its native G-quartet form. The residue sequence of this DNA was 5′-G-G-G-T-T-T-G-G-G-G-3′. It should be noted that this DNA has a TTT linker which is responsible for forming the folds. It also provides the DNA its signature preferred loop conformation.

For the nanopore translocation model, the human alphathrombin was modeled after the structure defined in 1HAP PDB, reported first in 1996. In the crystallographically resolved structure, this is a complex of thrombin protein and a 15-mer DNA. The nucleotide sequence was shown to have two stranded DNA (ssDNA) (PDB id 2AVJ) was a suitable match for our study.14 The nanopore segment, the simulation super cell consisted of a rigid nanopore and a solvated DNA protein complex. Modeling of the nanopore was done by freely available crystal nanopore building software. The nanopore was constructed using silicon nitride (Si3N4) with a constricted nanopore diameter at the center. The smallest diameter of the nanopore was 6 nm. The total atom count was 103 204 after addition of water and ions. In experiments, the 3′ or 5′ end of the DNA is usually modified with certain chemical groups and immobilized to the surface through other linkers. In our model, we fixed the 3′ end guanine base on the nanopore wall. The surface tethered probe was positioned inside the nanopore so that it came in contact with the thrombin protein when it passed through.

2.2. Force Field. The interatomic and intermolecular interactions between different species were modeled using the CMAP corrected CHARMM force field. Repulsive and attractive dispersion for short-range interactions were described by a Lennard-Jones potential with a cutoff distance of 1.1 nm and a switching distance between 1 and 1.2 nm. The total potential energy in the system consisted of bonded energies and nonbonded pair interaction energies, described by

\[
U = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{improper}} + E_{\text{elec}} + E_{\text{vdw}} + E_{\text{others}}
\]

\[
= \sum k_b (b - b_0)^2 + \sum k_{\theta\phi} (\theta - \theta_0)^2
\]

\[
+ \sum k_{\phi}(1 + \cos(n(\phi - \delta))) + \sum k_{\omega}(\omega - \omega_0)^2
\]

\[
+ \sum \left( \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \sum_{i<j} \left( \frac{r_{ij}^6}{r_{ij}^6} - 2 \left( \frac{r_{ij}}{r} \right)^6 \right) \right)
\]

\[
+ \sum U_{\text{CMAP}}(\phi, \psi)
\]  

(1)

Here, \(E_{\text{bond}}\) accounts for the bond stretches where \(k_b\) is the bond force constant and \(b_0\) is the equilibrium bond length. The \(E_{\text{angle}}\) term stands for the bond angles where \(k_{\theta\phi}\) is the angle force constant and \(\theta_0\) is the equilibrium bond angle. The third term refers to the dihedral energy where \(k_{\phi}\) is the dihedral force constant, \(n\) is the multiplicity factor, \(\phi\) is the dihedral angle and \(\delta\) is the phase shift. The \(E_{\text{improper}}\) term stands for the improper energy where \(k_{\omega}\) is the improper force constant and \(\omega\) is the out-of-plane angle. The next two terms represent electrostatic and van der Waals energies, respectively. Here, the van der Waals (VDW) energy is calculated with a standard Lennard-Jones potential and the electrostatic energy with a Coulombic potential. \(U_{\text{CMAP}}\) represents the dihedral energy correction term with backbone and torsional correction (CMAP, \(\phi, \psi\)).

2.3. Simulation Details. For both cases, the simulations were performed under constant-temperature, constant-volume...
(known as NVT) ensemble and conducted in two major steps. First, the systems were energy-optimized using the conjugate gradient method. After that, the temperature was raised and kept at 295 K using a Langevin thermostat. Long-range electrostatic interactions were calculated using the particle-mesh Ewald (PME) method at a time step of 1 fs. Nonbonded forces were evaluated at every two time steps. The systems were later equilibrated without applying any external field.

2.4. Applied Forces and Electric Field. For evaluating the DNA stability, a uniform electric field comparable to experimental value was applied in the z-direction. The field was applied only during the simulation (16 ns). In experiments, a typical membrane with a nanopore is ~100 nm thick where the applied voltage across it is in the 5 V range. Considering most of the potential drop is across the nanopore, a typical electric field across the nanopore is in the range of 0.05 V/nm. However, the profile inside the nanopore varies depending on the inner roughness. Usually, a nanopore is more constricted in the center due to the etching steps of the fabrication process. Thus, the field is higher in the central nanopore region. To verify that the DNA structure is not perturbed in the experimental voltage range, simulations were performed with gradually increasing applied electric field (Figure 3).

For the second case of protein translocation, the process of moving it through the nanopore using solely electric field (e-field) was slow. This was due to the fact that, at a physiological pH of 7.4, only a handful of protein residues are charged. Initially, after prepositioning the protein on the channel opening, we applied increasingly higher voltage to facilitate faster translocation. However, it still took long for the protein to translocate through. In physical experiments, protein translocation normally occurs at the millisecond scale. A small force in the z-direction was applied on the protein to guide it through the nanopore. This force would be akin to the forces stemming from the flow of cations, osmotic flow, salt concentration gradient, etc., in the experiments. A word of caution here is necessary; a large force could potentially alter the 3D conformation of protein. This could, in turn, affect the binding affinity with the DNA. The root-mean-square deviation (RMSD) of the protein was checked for different applied accelerations, and an acceleration of 0.5 Å/ps² was simulated. This was mild enough to not disturb the native 3D structure. To further accelerate the process, a slightly higher applied voltage (0.2 V/nm) was used, which is comparable to values reported in experiments. This approximated the electric potential applied on both ends of the nanopores in the experiments that establish an ionic current through the nanopores. However, as can be seen from Figure 3D, the
RMSD deviation from the initial structure stayed within 0.5 for this field; i.e., this electric field did not affect the DNA structure.

3. RESULTS AND DISCUSSION

3.1. Isolated DNA in Electric Field. Figure 3 shows the changes in the DNA structure with applied voltage. The RMSD of each molecule of the DNA structure from the initial equilibrated position was calculated at the end of each simulation. It was observed that, at lower electric fields, the RMSD of the nucleic acid sequence from the initial structure was minimal, indicating that the 3D structure was stable at lower applied fields. As the field increased to 0.8 V/nm, it can be seen from Figure 3D that the RMSD had an upward slope for as long as 20 ns, indicating structural instability. Any higher field quickly elongated the negatively charged DNA. Figure 3C shows the elongation of the DNA strand in the z-direction at higher applied voltage. As shown later, this elongation of DNA structure affected the protein translocation time due to loss of binding affinity.

3.2. Nanopore Protein Translocation. The X-ray resolved thrombin protein has a maximum length of ~5 nm. The nanopore size was maintained at comparable but slightly larger diameter (7 nm). The thickness of the nanopore was also kept at 7 nm. Before running the actual protein translocation simulation, an electric field was applied for 2 ns. The RMSD of both protein and DNA indicated no disruption in their structures due to this applied field. In all cases, the translocation time was calculated from the electrostatic and VDW energies between the molecules and the nanopore (Figure 4A). Three different scenarios were studied: (i) a single protein translocation through a bare nanopore, (ii) protein translocation through a nanopore coated with ssDNA, and (iii) a complex of protein and ssDNA translocating through the nanopore. For the first case, the DNA sequence was removed from the complex before the protein was allowed to translocate. The protein was pulled toward the nanopore using a small force on the molecule using grid molecular dynamics. For the second case, the DNA sequence was tethered inside the nanopore wall and the protein was separated from the complex. The protein was then pulled toward the nanopore and was allowed to come in contact with the DNA. The surface tethered probe DNA intercepted the protein through electrostatic and VDW interactions. For the third case, the DNA was not tethered to the surface and the protein–DNA complex was allowed to translocate instead.

The study of translocation of the DNA–protein complex has two important implications: First, the selective binding of the target protein in a solution with the aptamer can essentially result in changes in the physical dimensions and the isoelectric point (pI) of the molecule. The size and charge of the complex can change the translocation time much more profoundly. Second, the nanopore does not need to be functionalized. The complexity of the process can be reduced significantly and the same framework can be used for detection of multiple target proteins.

In the absence of any surface friction, nanochannels would allow a free translocation environment for protein molecules. However, molecular interactions with the confining walls of the nanopore can affect simple electrophoresis events. A bare nanopore allows faster translocation than the nanopore grafted with ssDNA. In other words, the translocation time for protein is significantly higher when nanopore is coated with tethered ssDNA (Figure 4B). Protein with a specific affinity to the sequence of the DNA strand can slow down more through a nanopore and hence give a distinct signature in the ionic current.

In the case of the protein–DNA complex, when a bias is applied to the electrolyte solution, the complex starts moving toward the oppositely charged electrode. Force on a single protein is lower than force on a protein–DNA complex because the latter has additional charges of DNA. It can be noted that DNA binding itself may change protein folding. The changes in protein folding can also change the surface charge distribution as well as the pI of the protein. This force causes the protein–DNA complex to move faster within the electrolyte toward the nanopore (Figure 5A). Also, the following competitive factors take place: (a) The protein–DNA complex has a higher volume and surface area compared to the protein alone, hence causing higher interaction with the nanopore wall and thus slowing it down. (b) Due to the additional charges, the complex moves faster within the electrolyte before entering the nanopore (Figure 5B). The moment when the protein and the complex just enter the...
nanopore, the latter has more velocity. Due to it having more mass, it has more inertia, and due to that, the complex should translocate faster through the nanopore in comparison to the case when protein alone is translocating. If the surface friction is not high enough, then the complex will translocate faster through the nanopore in comparison to the protein alone. As can be seen from Figure 4B, in our scenario, the dominating factor is the second one and that is why the complex translocates faster through the nanopore as compared to the protein alone.

The velocity of protein translocation was also calculated from the derivative of the translocation profile. Due to interaction with the tethered ssDNA, the protein slowed down significantly while translocating, causing longer translocation times. The velocity dropped to as low as one-fourth inside the nanopore with DNA coating (Figure 5A). The interactions of the protein with the walls of a bare nanopore also slowed the protein down but not as much. In a larger nanopore, with more space to move around inside the nanopore, such an effect would be less noticeable, thus underlining the necessity of similar sized nanopores to impart selectivity. The velocity and nanopore-size relationship can help design nanopores for optimized signal.

Nanopore size also plays an important role for such detection modalities. To have better selectivity and sensitivity, it is important to have pore dimensions close to the target molecule size. Pores with larger dimensions have more area inside for proteins to move around. As a result, it is possible for the traveling molecules to pass through the nanopore without actually interacting with the surface-bound ligands. In such a case, both the target and nonspecific proteins will show similar translocation behavior. The thrombin protein is 5 nm on its largest axis. We simulated the translocation event of thrombin protein on both 6 and 8 nm nanopores. As can be seen from Table 1, increasing the nanopore size just by 25% increased the protein travel speed by more than 6 times. It is also important to note that the larger nanopore lost selectivity between specific and nonspecific molecules. The translocation of thrombin through 8 nm functionalized and bare nanopore was not significantly different. One might argue that the selectivity stems from the nanoconfinement itself where the two molecules have to come in intimate contact. However, the selective binding of the two molecules is already established. The nano confinement may very well be enhancing the selective binding.

3.3. Deviation in 3D DNA Structure during Protein Translocation. As the protein translocated through the nanopore, the electrostatic and VDW forces between the two molecules slowed it down. The DNA strand also stretched along the path, losing its initial preferred loop structure (Figure 6A). This indicated that, after the protein was released, the
DNA would take some time to relapse back to its preferred structure. During this time, any other protein translocation would be rapid and nonselective. As can be seen from Figure 6B, the translocation times through a bare nanopore and the translocation time through a nanopore with elongated DNA are very close. Hence, the device would lose its selectivity. This latency places a theoretical boundary on the upper limit of the molecule detection frequency.

4. CONCLUSIONS

The close biological connections between DNA and protein dictate the energy interaction contributions of one molecule toward detection of the other. Nanopore is an emerging platform for such molecular sensing. Biophysical phenomena have shown that DNA strand can hold 3D conformation with protein traveling through. The 3D structure of the DNA between nanopores with DNA coating and a bare nanopore surely bolster the idea of such detection modalities.


