Biophysical Reviews and Letters
Vol. 9, No. 1 (2014) 105–114

DOI: 10.1142/S1793048013500100

ELECTROSTATICS STUDY OF A SINGLE-STRANDED DNA: A PROSPECTIVE FOR SINGLE MOLECULE SEQUENCING

MOHAMMAD J. I. A. SAUDE and BASHIR I. MORSHED

Electrical and Computer Engineering Department
The University of Memphis
Memphis, TN 38152, USA

Received 11 November 2013
Revised 25 December 2013
Accepted 25 December 2013
Published 7 February 2014

Single molecule DNA sequencing requires new approaches to identify nucleotide bases. Using molecular dynamics simulations, we investigate the intrinsic electrostatics of single-stranded DNA by solving the nonlinear Poisson–Boltzmann equation. The results show variations of the molecular electrostatic potential (MEP) within 3 nm from the center of the sugar backbone, with suitably differentiable variations at 1.4 nm distance. MEP variations among four nucleotide bases are the most significant near $\sim 33.7°$ and $\sim 326.3°$ angular orientation, while the influences of the neighboring bases on MEPs become insignificant after the 3rd-nearest neighbors. This analysis shows potential for direct electronic sequencing of individual DNA molecules.

Keywords: Single molecule DNA sequencing; molecular dynamics; molecular electrostatic potential.

Special Issue Comment: This paper about DNA sequencing based on molecular electrostatic potential maps of the DNA in the channel is related to the Special Issue articles about: measuring enzymes,[32] and solving single molecules' trajectories with the RDF approach[33] and with the QuB software.[34]

1. Introduction

Deoxyribonucleic acid (DNA) is a long chain of sequentially ordered arrangement of four types of nucleotide bases: Adenine (A), Thymine (T), Cytosine (C), and Guanine (G). The distance between each subsequent nucleotide base is ~ 3.4 Å, and the diameter of a DNA double helix is between 22 to 26 Å.[1] Detecting this sequence of nucleotide bases is called DNA sequencing.[2] Fred Sanger devised the first method of DNA sequencing based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during *in vitro* DNA replication in 1977.[3] This first generation DNA sequencing method is still in use in various improved format such as slab gel and capillary electrophoresis.[4,5] Faster

and inexpensive DNA sequencing approaches have been developed, leading to the second generation of sequencing technologies that require collection of DNA strands, such as DNA microarrays based on hybridization of complementary strands, and automated DNA sequencers with fluorescence tagging of terminals with different fluorophores.[2,6] The next generation of sequencing technology, often referred to as the third generation, is envisioned to be based on individual molecules of DNA with the aim of achieving an unprecedented sequencing rate in order to sequence a complete human genome in several hours with less than $1,000 per genome.[7] This requires new sequencing technology development utilizing an interdisciplinary approach. 'Technology development for the $1,000 genome' program under the National Human Genome Research Institute (NHGRI) in the United States promote many high-risk high-reward single-molecule based sequencing technologies including single-molecule based nanopore and Förster resonance energy transfer (FRET) based sequencing technologies.[2] Early breakthroughs have been reported with promise for DNA nanopore,[8–10] and single molecule real time (SMRT).[11,12] In nanopore technology, an nm-sized pore is fabricated on a membrane through which a DNA strand is passed with a gradient electric field. An electric current is established through the nanopore that changes with the sizes of the bases when a DNA strand is translocated through the pore from one side of the membrane to the other. The changes of the current flow through the nanopore are monitored, and the sequence is decoded from the patterns of the relative current flow during this translocation process. By observing several independent parameters, such as time duration and temperature, along with the blocked current, the discrimination accuracy among various nucleotide types increases.[13] The SMRT technique utilizes a zero-mode waveguide (ZMW) to capture fluorescence emitted from individual tags as DNA is synthesized using DNA polymerase enzyme.[11,12]

The electrostatics around a molecule can be calculated using Poisson–Boltzmann equations,[14] and the charges, atomic arrangements, surrounding medium and shapes of the DNA and the dielectric boundary could affect the electrostatic potentials significantly.[15] It has been shown that the electrostatic interaction between two DNA molecules in a membrane depends on the charge density of the membrane and the distance between the two molecules.[16] Electrostatic interactions play a key role in various aspects of the structure and function of nucleic acids since they are highly charged molecules.[15] Using molecular electrostatic potential (MEP), it has been shown that there are different patterns for paired and mispaired nucleotides located in the central plane between bases.[17] Several solid-state devices have been proposed to detect single molecules of DNA by studying the change of capacitances induced by the DNA molecule within the nanopore.[18,19] These devices use a metal-oxide-semiconductor (MOS) capacitor that applies an electrical field to translocate the DNA molecule through the nanopore, while the response induced from the DNA backbone and its bases are recorded. Using a nanopore of 1 nm diameter in a MOS capacitor, the recorded voltage induced from different bases is between 2 to 9 mV. These potentially could be exploited to detect the number of nucleotides and could

lead to an electronic sequencing approach.[20] Furthermore, defective or mutated DNA could be detected using the nanopore capacitor approach by calculating the electrostatic potential, as the defective or mutated DNA has less charge than a normal base.[21] Single molecule measurements are emerging as a powerful tool for characterizing protein functions by providing newer insights and enabling potential for single molecule sensing.[22] Newer sensing techniques such as single molecule technique to identify aggregation, misfolding, deposition, and conformational change,[23] or the two state photon trajectory-based classification could lead to atomic level sensitivity and selectivity.[24–27] With new genetically encoded stress FRET-based sensor, mapping the stress field in a heterogeneous time-dependent environment has been demonstrated *in vivo* with a greater dynamic range and sensitivity.[28]

The motivation of this study is to explore the potential of harnessing the inherent electrostatic properties of a single stranded DNA to enable single molecule DNA sequencing. If such technology can be materialized, rapid DNA sequencing without time-consuming sample preparation steps and polymerize chain reaction might be conceived. Investigating the electrostatic potential induced by the DNA itself while translocating through the nanopore is largely uncharted territory. This paper presents this unique prospective and investigates the feasibility of single molecule DNA sequencing-based solely on the intrinsic electrostatics of DNA for a heterogeneous time-dependent sensing. Our analysis indicates that sufficient differentiation of electrostatics exists at a very close proximity to enable direct electronic sequencing of individual DNA molecules.

2. Modeling

We utilize molecular dynamics (MD) simulation as the tool for this analysis. We have simulated single-stranded DNA (ssDNA) molecule in this analysis, as the differences in electrostatic distribution among various nucleotide bases can be probed easily. For double-stranded DNA (dsDNA), nucleotide bases are inward of the helix structure and bound to complementary bases, whereas for ssDNA, nucleotide bases are exposed; enabling spatial probing. Hence, we studied electrostatic models of ssDNA, measure and analyze the molecular electrostatic potential (MEP) around the ssDNA, and explore detectable variations among four types of nucleotide bases (A, G, C, and T). The DNA molecules that are studied in this work are nucleic acids with conformation of B-DNA, which has been directly observed in functional organisms.[29] The Protein Data Bank (pdb) files of the DNA, which contains the atomic positions in three-dimensional format, have been created from the publicly available tool at 3D-DART.[30] This model is used for MD simulation by solvating the molecule with water and incorporating appropriate ionic atoms as mentioned below.

At first, the dsDNA is separated into two chains and then converted to ssDNA molecules using Visual Molecular Dynamics (VMD) software.[31] Afterwards, the Protein Structure File (psf) files of the DNA, which contains atomic information

including types, charge, mass, and radius, have been created using the "psfgen" plugin available in VMD by applying CHARMM (Chemistry at HARvard Macromolecular Mechanics) force field. Each ssDNA is then solvated with a box of TIP3P (a very popular model where the three-site model has three interaction sites, corresponding to the three atoms of the water molecule) water with dimensions of 100 Å, 100 Å, and 180 Å along the x-, y-, and z-axis, respectively. The system is ionized with Na^+ and Cl^- ions to achieve a concentration of 0.025 mol/L and to make the overall system neutral, using the "autoionize" plugin available in VMD. In this work, we have not looked into the effect of ionic concentration on MEP, however we anticipate the effect could be a shift of bias or offset of the observations reported here. Energy of each system, consisting over 96,000 total atoms, is minimized towards steady state using a scalable molecular dynamic simulation software, NAMD.[31] The minimization is simulated for 10,000 steps with 1 fs per step using CHARMM force field. The calculation presented in this paper is based on the obtained system minimized to the stable state.

To calculate the molecular electrostatic potential, all atoms must have charge information. To achieve this, the pdb files of the ssDNA are converted to pqr files that have the corresponding atomic charge information, applying the pdb2pqr software with the CHARMM force field. MEP is calculated using Adaptive Poisson–Boltzmann Solver (APBS) software (freely available macromolecular electrostatics calculation program governed by GPL). Each system was mapped onto a cube lattice of $161 \times 161 \times 161$ points and the grid lengths are 132 Å in any of x-, y-, and z-directions. MEP was calculated at each point inside the cube that can have a maximum of $\Delta/2$ error along x-, y-, or z-axis, where $\Delta(= 0.825\,\text{Å})$ is the distance between two measurement points (i.e. the grid spacing). The salt concentration is 0.025 mol/L, and the internal and external dielectric coefficients with respect to atomic volume are assumed to be 1 (e.g. vacuum within the atom) and 78.54 (e.g. surrounding water), respectively. If these conditions were to be relaxed, e.g. dielectric constant of 3 within any atom, the observed data will possibly shift correspondingly with similar trends as reported here. The simulation results are rendered with VMD, and the electrostatic potentials have been visualized with VMD, analyzed and plotted with Matlab (MathWorks Inc., MA, USA) and MicroSoft Excel (MicroSoft Corp., CA, USA).

3. Results

For numeric analysis a sufficient length of DNA strand needs to be considered that eliminates terminal effect but not very computationally expensive. The length of ssDNA sequence studied in this work is 25 base-pair (bp) long, i.e. the monomers are poly$(dA)_{25}$, poly$(dT)_{25}$, poly$(dG)_{25}$ and poly$(dC)_{25}$, and have monitored electrostatics around the residue-13 nucleotide (13th base at the center of the ssDNA sequence). Later, we show that the length of 25 bp is sufficiently long to ignore the terminal effects on the bases under observation.

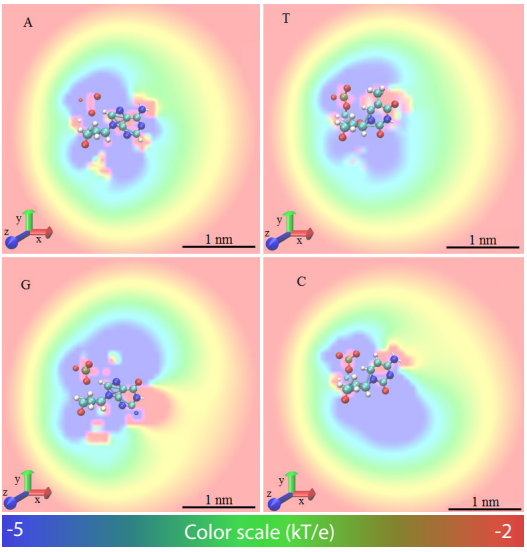

Fig. 1. (Color online) MEP distribution map around the residue-13 nucleotide (central base) in monomers of poly(dA)$_{25}$, poly(dT)$_{25}$, poly(dG)$_{25}$ and poly(dC)$_{25}$. The color scale shown at the bottom is from -5 kT/e (blue) to -2 kT/e (red).

Figure 1 shows two-dimensional electrostatic potential distribution map in a plane perpendicular to the DNA helix (oriented along the z-axis of the simulation box) around the residue-13 nucleotide (central base) of ssDNA monomers. The ssDNA is oriented such that the center of the sugar backbone and the center of the nucleotide base are on the x-axis. MEP evidently demonstrates slightly different patterns for each nucleotide. We note that the atomic electrostatic variations diminish within $\sim$3 nm distance to bulk potentials. Hence an effective electrostatic sequencing mechanism must probe within this distance from the ssDNA strand.

To study the variations around the ssDNA molecules, we analyzed MEP around the bases at four different radii (1 nm, 1.4 nm, 2 nm, 3 nm) on the xy-plane. The center of the sugar backbone of each nucleotide configuration is used as the reference, and the angle is measured counterclockwise from the center of the nucleotide along the x-axis of plots in Fig. 1. The values are plotted in Fig. 2 at every 2° for all four nucleotide bases.

Figure 2 depicts that the variations among the MEPs are the most diverse between radii 1 to 3 nm. We note that at distance of 1 nm, substantial variations are observed within 45° and after 315° due to the measurement points being inside the nucleotide atoms. The figure also indicates that there are two angular ranges where the variation among four bases peaks, one of them reside around 33.7°, while the other is around 326.3°. The corresponding data is presented in Table 1. We have selected the radius of 1.4 nm from the center of the nucleotide backbone, a feasible probing distance, with an angular orientation of 33.7° from the x-axis for

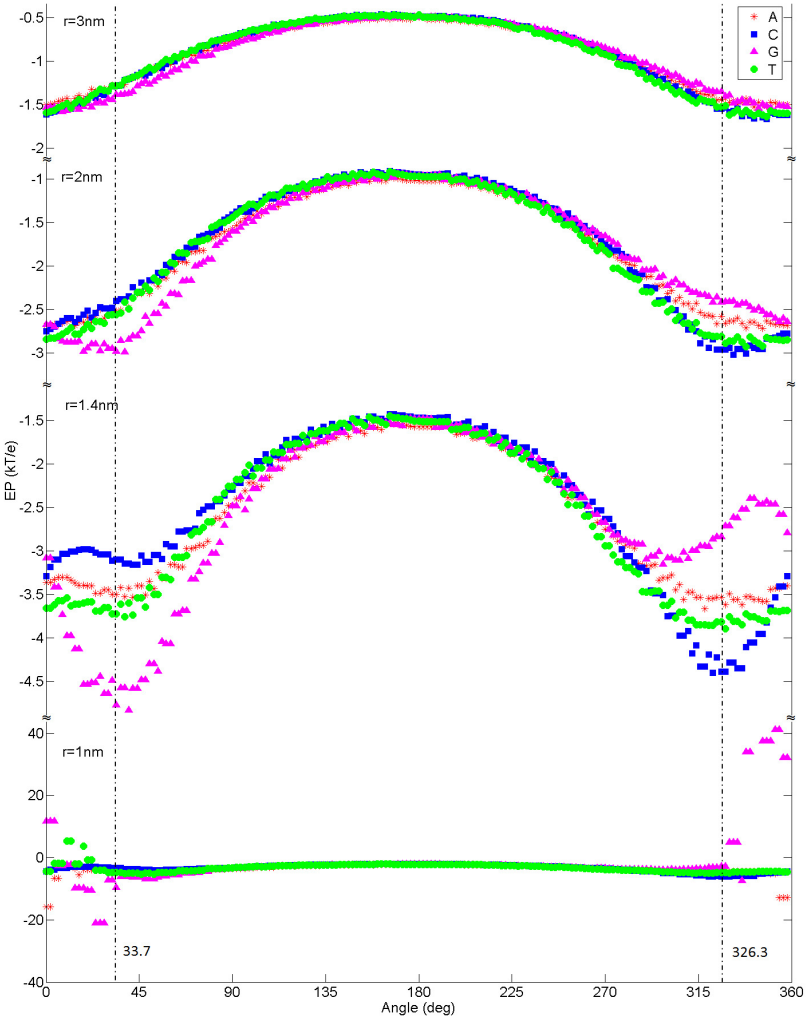

Fig. 2. (Color online) MEP on circles with various radii with the center at the suger backbone at various distances (r) for four types of nucleotides (A, C, G, and T). Data are plotted for every 2°, counterclockwise from the x-axis.

Table 1. MEPs at $r = 1.4\,\text{nm}$ distance from the sugar backbone for residue-13 nucleotide (central base) in monomers of poly(dA)$_{25}$, poly(dT)$_{25}$, poly(dG)$_{25}$ and poly(dC)$_{25}$.

	MEPs for various nucleotide base types (kT/e)			
Angular orientation	poly(dG)	poly(dT)	poly(dA)	poly(dC)
33.7°	−4.767	−3.722	−3.498	−3.098
326.3°	−2.52	−3.816	−3.531	−4.386

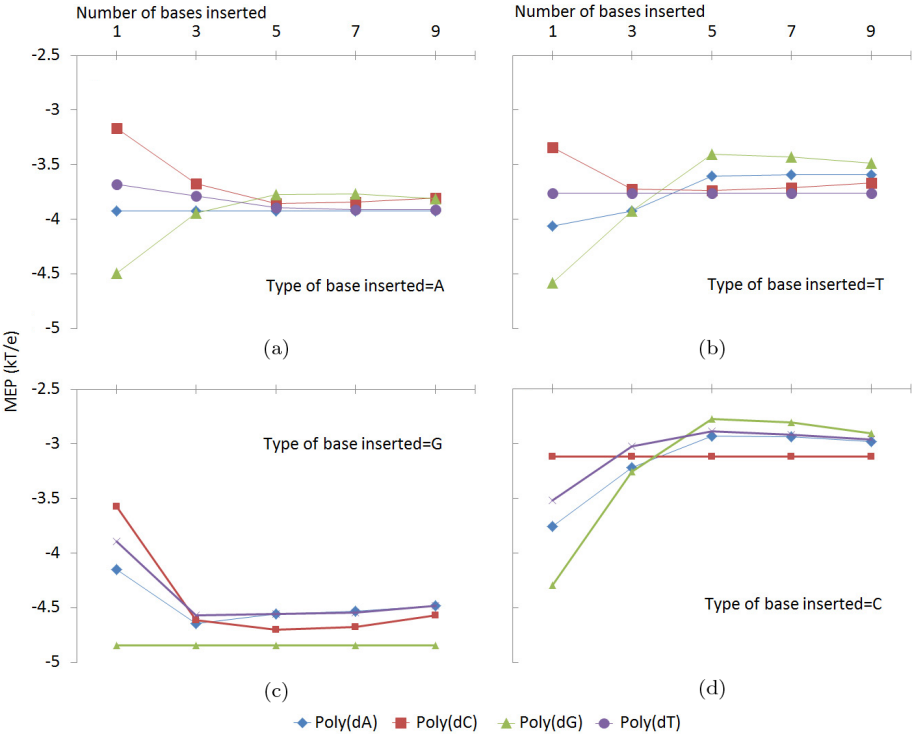

Fig. 3. (Color online) The influnce of the neighboring nucleotides on MEP at the probing points (1.4 nm distance from the center of the sugar backbone of the ssDNA at 33.7° counterclockwise orienation from the x-axis). (a) Insertion of base-A. (b) Insertion of base-T. (c) Insertion of base-G. (d) Insertion of base-C. In all plots, the effect of inserting higher number of bases diminishes rapidly towards the monomer MEP of the inserted nucleotide type.

further analysis. Analysis along other radii and angular orientations are similar. The calculations are performed with various grid/lattice sizes with values ranging between ± 0.01 to ± 0.1 without any visually noticeable changes, indicating minimum sensitivity of the data on the grid/lattice size.

Further, we investigate the influence on the electrostatic field distributions of the central nucleotide base from the neighboring nucleotides of the ssDNA strand. To reduce the number of combinations to be simulated, we limit the combinations of inserts to only identical bases symmetrically (with respect to residue-13) inserted at the middle of a monomer of ssDNA, replacing the corresponding bases of the monomer. The 64 combinations are simulated and categorized into four groups based on the nucleotides being inserted (Fig. 3). In the first group, MEPs are calculated for 1, 3, 5, 7, and 9 base insertions of nucleotide base-A in the middle of monomers of poly$(dA)_{25}$, poly$(dT)_{25}$, poly$(dG)_{25}$, and poly$(dC)_{25}$. Similarly, other groups of simulations inserted nucleotide bases of T, G, and C into all four types of monomers.

An insertion of 1 base represents MEP influence by all adjacent bases of the same or different kinds (denoted as 1st-nearest neighbor influence). A symmetrical insertion of 3 bases yields the first neighbors on both sides to be of the same type, while from the 2nd-nearest neighbors to the terminals, original monomer bases are preserved (denoted as 2nd-nearest neighbor influence), and so on. Evidently the data shows the influence of neighboring bases diminishes for 5 or more inserted bases (3rd-nearest or higher neighbor influence). As the number of inserted base increases, MEP results trend to the MEP consisting of polymers of the inserted base. For instance, the MEP of nucleotide 13 in the DNA polymer of $\text{poly(dG)}_{10}\text{poly(dT)}_5\text{poly(dG)}_{10}$ will be closer to poly(dT)_{25} than poly(dG)_{25}, as depicted in Fig. 3(b). From this prospective, we show that simulations with 25 bp ssDNA for probing electrostatics at the residue-13 will have insignificant terminal effect on the analysis.

4. Conclusions

In this work, we have analyzed and studied the electrostatic distribution and MEP around ssDNA to demonstrate the feasiblity of direct indentification of nucleotide bases for single molecule DNA sequencing. We show that a proximal probing location around 1.4 nm at 33.7° might distinctly identify four types of nucleotide base polymers for a time-dependent heterogenous sensing of DNA polymers. However, the influence of at least 3rd-nearest neighbor must be taken into account, possibly using an adaptive or decision-directed sequencing algorithm. Furthermore, a simultaneous probing of both 33.7° and 126.3° at proximities of ~ 1.4 nm would additionally improve the ability to uniquely identify each nucleotide base. Due to the structural similarities, the nucleotide-bases of A and T types have very small noise margin for distinction (Figs. 2 and 3). This discrimination difficulty of "naked" bases of A and T might be improved through simultaneous probing at multiple locations or differentially ionizing one of the bases. Influence of noise such as thermal noise and shot noise are critical on such single molecule probing as the noise floor essentially dictates the sensitivity and specificity. Signal processing techniques such as oversampling, notch filtering, and iterative numerical analysis methods can be exploited to reduce the noise in the processed signals to obtain a cleaner signal with a higher signal to noise ratio. Future research direction includes studying a random sequence of DNA strand, effect of ionic concentration of salts and dynamics, effect of probing such as nanopore wall, studying the MEPs around ssDNA during heating and equilibration to explore the relative oscillations and rotations of the nucleotide bases within the DNA strand and the resultant sensitivity on the observed MEP at the probing point. The potential probing technique could be two-state photon trajectories,[25–27] quantum dot, or functionalized conductive carbon nanotube, among others that could allow probing relatively away from the nanopore wall or sensor structure to minimize their effects on MEP measurement.

Acknowledgments

The authors would like to gracefully acknowledge the resources provided for the large-scale simulations performed on the High Performance Computing (HPC) center hosted at the University of Memphis. The authors are also grateful to Dr. Yongmei Wang, Professor, Department of Chemistry, the University of Memphis, and Dr. Russell Deaton, Professor, Department of Electrical and Computer Engineering, the University of Memphis, for their critical feedback to substantially improve the quality of this manuscript. The authors would like to thank the anonymous reviewers and editors for their valuable comments, suggestions and insights to improve the quality of this manuscript.

References

1. M. Mandelkern, J. Elias, D. Eden and D. Crothers, *J. Mol. Biol.* **152**(1), 153–161 (1981).
2. E. Mardis, *Nature* **470**, 198–203 (2011).
3. F. Sanger, S. Nicklen and R. Coulson, *Proc. Natl. Acad. Sci.* **74**(12), 5463–5467 (1977).
4. C. Kan, C. P. Fredlake, E. A. S. Doherty and A. E. Barron, *Electrophoresis* **25**(21–22), 3564–3588 (2004).
5. O. Morozova and M. A. Marra, *Genomics* **92**, 255–264 (2008).
6. F. Collins, *Nature* **464**, 674–675 (2010).
7. E. R. Mardis, *Genome Biol.* **7**, 112 (2006).
8. D. W. Deamer and D. Branton, *Accounts of Chem. Res.* **35**(10), 817–825 (2002).
9. M. J. Holst, The Poisson–Boltzmann equation, PhD thesis, University of Illiuois, Urbana-Champaign, IL, USA (1994).
10. J. J. Kasianowicz, E. Brandin, D. Branton and D. W. Deamer, *Proc. Natl. Acad. Sci.* **93**, 13770–13773 (1996).
11. J. Eid, A. Fehr, J. Gray *et al.*, *Science* **323**, 133–138 (2009).
12. J. Korlach, K. P. Bjornson, B. P. Chauduri *et al.*, *Methods Enzymol* **472**, 431–455 (2010).
13. A. Meller, L. Nivon, E. Brandin, J. Golovchenko and D. Branton, *Proc. Natl Acad. Sci.* **97**, 1079–1084 (2000).
14. M. J. Holst, Multilevel methods for the Poisson–Boltzmann equation, *Technical Report UIUCDCS-R-93-1821* (University of Illinois at Urbana-Champaign, IL, USA, 1994).
15. B. Jayaram, K. A. Sharp and B. Honig, *Biopolymers* **28**, 975–993 (1989).
16. G. Caracciolo and R. Caminiti, *Phys. Lett.* **400**, 314–319 (2004).
17. I. Otero-Navas and J. M. Seminario, *J. Mol. Modeling* **18**(1), 91–101 (2012).
18. M. E. Gracheva *et al.*, *Nanotechnology* **17**, 622–633 (2006).
19. P. Mali and R. K. Lal, *IEEE Trans. Electron. Devices* **51**(12), 2004–2012 (2004).
20. M. E. Gracheva, A. Aksimentiev and J. P. Leburton, *Nanotechnology* **17**, 3160–3165 (2006).
21. G. Sigalov, J. Comer, G. Timp and A. Aksimentiev, *Nano Lett.* **8**, 56–63 (2008).
22. N. S. Hatzakis, *Biophysical Chemistry* (in press, 2013) doi:10.1016/j.bpc.2013.11.003.
23. N. Cremades, S. I. A. Cohen, E. Deas *et al.*, *Cell* **149**(5), 1048–1059 (2012).
24. O. Flomenbom, Quantitave Biology, submitted (2013).
25. O. Flomenbom and R. J. Silbey, *Proc. Natt. Acad. Sci.* **103** (29), 10907–10910 (2006).
26. O. Flomenbom and R. J. Silbey, *Phys. Rev. E* **78**(6), 066105(1–20) (2008).
27. O. Flomenbom, *Advances in Chemical Physics* **146**, 367 (2011).

28. F. Meng and F. Sachs, *J. Cell Science* **125**(3), 743–750 (2012).
29. A. Ghosh and M. Bansal, *Acta Crystallogr D* **59**(4), 620–626 (2003).
30. M. van Dijk and A. M. J. J. Bonvin, *Nucl. Acids Res.* **37**, W235–W239 (2009).
31. VMD and NAMD software, publicaly available through the University of Illinois at Urbana-Champaign, IL, USA.
32. S. K. Jørgensen and N. S. Hatzakis, *Biophys. Rev. Lett.* **8**, 137–160 (2013).
33. O. Flomenbom, *Biophys. Rev. Lett.* **8**, 109–136 (2013).
34. C. Nicolai and F. Sachs, *Biophys. Rev. Lett.* **8**, 191–211 (2013).