The Journal of Chemical Physics

RESEARCH ARTICLE | AUGUST 20 2013

Molecular dynamics simulation of a single-stranded DNA with heterogeneous distribution of nucleobases in aqueous medium

Kaushik Chakraborty; Sriteja Mantha; Sanjoy Bandyopadhyay

J. Chem. Phys. 139, 075103 (2013)
https://doi.org/10.1063/1.4818537

AIP Publishing

Molecular dynamics simulation of a single-stranded DNA with heterogeneous distribution of nucleobases in aqueous medium

Kaushik Chakraborty, Sriteja Mantha, and Sanjoy Bandyopadhyay[a)]
Molecular Modeling Laboratory, Department of Chemistry, Indian Institute of Technology, Kharagpur 721302, India

(Received 21 June 2013; accepted 1 August 2013; published online 20 August 2013)

The DNA metabolic processes often involve single-stranded DNA (ss-DNA) molecules as important intermediates. In the absence of base complementarity, ss-DNAs are more flexible and interact strongly with water in aqueous media. Ss-DNA–water interactions are expected to control the conformational flexibility of the DNA strand, which in turn should influence the properties of the surrounding water molecules. We have performed room temperature molecular dynamics simulation of an aqueous solution containing the ss-DNA dodecamer, 5′-CGCGAATTCGCG-3′. The conformational flexibility of the DNA strand and the microscopic structure and ordering of water molecules around it have been explored. The simulation reveals transformation of the initial base-stacked form of the ss-DNA to a fluctuating collapsed coil-like conformation with the formation of a few non-sequentially stacked base pairs. A preliminary analysis shows further collapse of the DNA conformation in presence of additional salt (NaCl) due to screening of negative charges along the backbone by excess cations. Additionally, higher packing of water molecules within a short distance from the DNA strand is found to be associated with realignment of water molecules by breaking their regular tetrahedral ordering. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4818537]

I. INTRODUCTION

Single-stranded DNA (ss-DNA) molecules are essential intermediates in various biological processes, such as DNA replication, transcription, recombination, and repair.[1] Double-stranded DNAs (ds-DNAs) form stable structures due to formation of strong hydrogen bonds between the complementary bases. In absence of such complementarity, ss-DNA oligomers are in general more flexible in aqueous solution.[2–4] In general, the conformational properties of an ss-DNA oligomer are found to depend on the composition of its nucleobase sequence.[5] This arises due to differential stacking ability between the purine and pyrimidine bases. To understand the physical properties of ss-DNA, it is necessary to probe such conformational flexibility. Ss-DNA being charged is expected to interact strongly with nearby water molecules. The ability of water to form hydrogen bonds with ss-DNA is particularly important. Therefore, it is natural that the water molecules present in the vicinity of the ss-DNA would play a critical role in controlling its flexibility in an aqueous environment, and hence determine the reaction pathways and rates associated with biological processes involving such molecules. Strong DNA–water interaction in turn should also influence the properties of the interfacial water molecules. A complete understanding of the hydration behavior of ss-DNA therefore requires microscopic knowledge of the conformational fluctuations of the DNA oligomer and the local structural and dynamical properties of the surface water molecules.

Due to inherent flexibility, the structural aspects of ss-DNA oligomers in solutions are not well understood experimentally. Different techniques, such as atomic force microscopy,[2] transient electric birefringence,[3] fluorescence resonance energy transfer (FRET),[4] etc., have been used to characterize the conformational features of ss-DNAs. These studies in general indicate that in aqueous media ss-DNAs neither adopt completely extended conformations nor exhibit significant ordering. Using single molecule force spectroscopy, Cui *et al.*[6] recently showed that the presence of bound water molecules results in shortening of ss-DNA chains in aqueous environment. Hydration properties of ds-DNA and ss-DNA molecules have been compared by Teixeira and co-workers[7] from quasielastic neutron scattering studies. Unlike ds-DNA, the degree of confinement of water molecules around the ss-DNA has been found to be minimum. Recently, the hydration behaviors of ss-DNAs comprising of thymine (dT_{10}) and adenine (dA_{10}) have been studied using infrared absorption spectroscopy.[8] These studies revealed that water molecules are bound to these nucleobases in a site-specific manner. In an important study, Liang *et al.*[9] demonstrated that the ss-DNA secondary structure can be characterized from electrophoretic analysis of DNA catenanes formed by them. From time-resolved fluorescence measurements using a fluorescein dye, Miyasaka and co-workers[10] recently studied the conformational dynamics of ss-DNA molecules in nanosecond to submillisecond time scale. They identified three main conformations of the DNA strand in aqueous solution with two interconversion time constants of several hundreds of nanoseconds and tens of microseconds.

Ss-DNA–water interactions in aqueous media are expected to influence the electrostatic, base pairing, and stacking interactions, which control the conformations of these oligomers.[11,12] Different statistical descriptions and simulations based on analytical models, such as freely jointed model,

a)Author to whom correspondence should be addressed. Electronic mail: sanjoy@chem.iitkgp.ernet.in

worm-like chain model, etc., have been used to characterize the conformational flexibilities of ss-DNAs.[11,13] From Monte Carlo (MC) studies, Zhang *et al.*[11] showed that at high salt concentration, ss-DNA forms hairpin-like folded structures due to pairing between complementary bases. Recently, it has been demonstrated further that the stability of such hairpin structures depends on the type of the salt and its concentration.[14,15] Interestingly, in contrast to experimental findings, Buhot and Halperin[12] from their numerical analysis showed that ss-DNA oligomers can form stacked domains in between melted random coil conformations. The elastic properties of ss-DNA molecules on stretching have also been explored from MC studies.[16] The effects of sodium ions and chain lengths on the microscopic structure and dynamics of ss-DNAs have been studied from molecular dynamics (MD) simulations.[17] It is shown that beside sodium ions interacting with the phosphate groups, enhanced base stacking with increase in chain length results in reduced flexibility of ss-DNAs. Sen and Nilsson[18] carried out MD simulations to analyze the conformational properties of different nucleic acids and their interactions with solvent water. They demonstrated that the structure and dynamics of the nucleic acid strands strongly depend on the covalent nature of the backbones. Significant hydrogen bonds have been found to form between the strands and water around them. Formation of water bridges around the phosphate groups has also been noticed. Chatterjee *et al.*[19] recently estimated thermodynamic properties of a ss-DNA molecule in aqueous solution from room temperature MD simulations. They observed that the conformational stability of the DNA strand depends on thermodynamic conditions.

It is apparent from the above discussion that despite limited attempts a proper understanding of the conformational properties of ss-DNA oligomers and their interactions with water molecules in aqueous media is still elusive. In this work, we attempt to probe the DNA–water interaction, and its influence on the ss-DNA conformations and the microscopic properties of water molecules hydrating the DNA strand. This is done by performing atomistic MD simulation of an aqueous solution of a ss-DNA dodecamer with heterogeneous base sequences, 5′-CGCGAATTCGCG-3′, at room temperature. Attempts have been made to study the conformational flexibilities of the ss-DNA molecule and explore its influence on the local structure and ordering of nearby water molecules. The rest of the article has been organized as follows. In Sec. II, we describe in brief the setup of the simulation system and the methods employed. The results are presented and discussed in Sec. III. The important findings and the conclusions reached from the study are summarized in Sec. IV.

II. SYSTEM SETUP AND SIMULATION DETAILS

We have used the NAMD code[20] to carry out the MD simulation. The all-atom CHARMM force field and potential parameters for nucleic acids[21,22] were used for the ss-DNA, while the TIP3P model[23] that is consistent with the chosen nucleic acid force field, was employed for water.

The initial coordinates of the ss-DNA oligomer with the base sequence 5′-CGCGAATT CGCG-3′ were obtained by unzipping chain A (i.e., by removing the complementary chain B) of the Dickerson-Drew dodecamer (DDD) from the protein data bank (PDB ID: 436D).[24] As mentioned before, this allowed us to study an ss-DNA oligomer with heterogeneous distribution of nucleobases. Hydrogen atoms were first added to the DNA strand and the two end residues were capped by replacing the phosphate groups with hydroxyl groups. The ss-DNA molecule was then immersed in a large cubic cell containing equilibrated water molecules. To avoid unfavorable contacts, the insertion process was carried out by carefully removing those water molecules that were present within 2 Å from the DNA atoms. The overall charge of the system was neutralized by adding 11 Na^+ ions. The final system contained the ss-DNA molecule (379 atoms) in a cubic cell containing 7632 water molecules and 11 Na^+ ions. The initial edge length of the simulation cell was 60 Å.

The system was first minimized to eliminate any initial stress using the conjugate gradient energy minimization method as implemented in the NAMD code.[20] The temperature of the system was then gradually increased to the room temperature of 300 K within a short MD run of about 100 ps. This was carried out at a constant pressure (P = 1 atm) under the isothermal-isobaric ensemble (NPT) conditions. It was then followed by an NPT equilibration run of about 5 ns duration at 300 K. While the Langevin dynamics method with a friction constant of 1 ps^{-1} was used to control the temperature of the system, the pressure was controlled by the Nosé-Hoover Langevin piston method.[25] During this period, the volume of the cell was allowed to fluctuate isotropically. At the end of this 5 ns NPT run, the volume of the cell attained a steady value corresponding to an edge length of around 61 Å. At this point the cell volume was fixed and the simulation conditions were changed from that of NPT to NVT ensemble (constant volume and temperature). The NVT equilibration run was continued further at 300 K for another 5 ns duration. After that, the simulation conditions were changed to that of constant energy in microcanonical ensemble (NVE). After a short equilibration (∼500 ps), a long NVE production run was carried out for about 100 ns duration. The average temperature of the system as obtained from the equilibrated NVE trajectory was found to be 301.1 (±1.1) K. The entire simulation was carried out with a time step of 1 fs, and the trajectory was stored every 400 fs for analyses. The minimum image convention[26] was employed to calculate the short-range Lennard-Jones interactions using a spherical cutoff distance of 12 Å with a switch distance of 10 Å. The SHAKE algorithm was used to constrain the bonds for the TIP3P water molecules to their equilibrium values.[26] The long-range electrostatic interactions were calculated by using the particle-mesh Ewald (PME) method.[27]

III. RESULTS AND DISCUSSION

A. Structural features of the ss-DNA

Several configurations of the ss-DNA molecule as obtained from the simulated trajectory at different time intervals are shown in Fig. 1. For comparison, the initial configuration of the DNA in the extended helical base-stacked

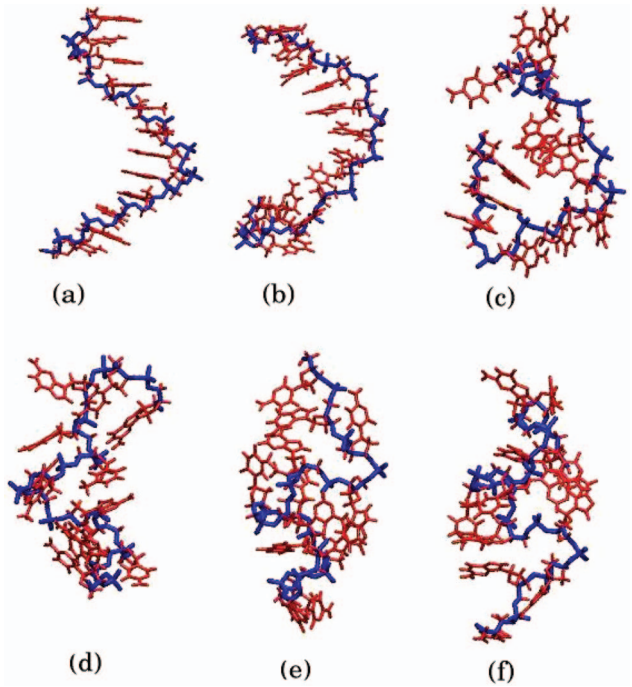

FIG. 1. (a) The initial and (b)–(f) snapshots of a few representative configurations of the ss-DNA molecule as obtained at 10, 25, 50, 75, and 100 ns of the simulation, respectively. The atoms of the sugar-phosphate backbone are drawn in blue and that of the nucleobases in red.

form is also displayed. Conformational flexibility of the DNA molecule in the single-stranded form in solution is clearly evident from the figure. During the simulation, the DNA molecule has been found to undergo transformation from the ordered base-stacked form to fluctuating collapsed coil-like conformations. Existence of such disordered conformations of the simulated ss-DNA molecule agrees well with experimental observations.[28]

To further explore the structural deviation of the simulated ss-DNA molecule with respect to its initial structure, we have computed the root mean square deviations (RMSD). RMSDs of a biomolecule can provide valuable information on its conformational flexibility originating from the local motions exhibited by the molecule in the available conformational space. The calculations involving the non-hydrogen heavy atoms are carried out separately for the sugar-phosphate backbone and the bases of the DNA over the entire simulated trajectory. The time evolutions of the RMSDs and their distributions are shown in Figs. 2(a) and 2(b). It can be seen that for both cases, the RMSD values change significantly within first 10 ns or so before attaining steady values (within 7–8 Å) with occasional oscillations for the rest of the simulation duration. This indicates considerable deviations of the simulated structures from the initial helical form. The results suggest that in general the DNA bases with relatively large RMSD values are slightly more flexible as compared to the covalently linked sugar-phosphate backbone. This is further evident from Fig. 2(c), where the average RMSD values of the backbone and base moieties of each of the residues are plotted. Except for the terminal residues, the RMSDs of most of the residue bases are noticeably higher than that of the backbone parts. The calculated average RMSD values for

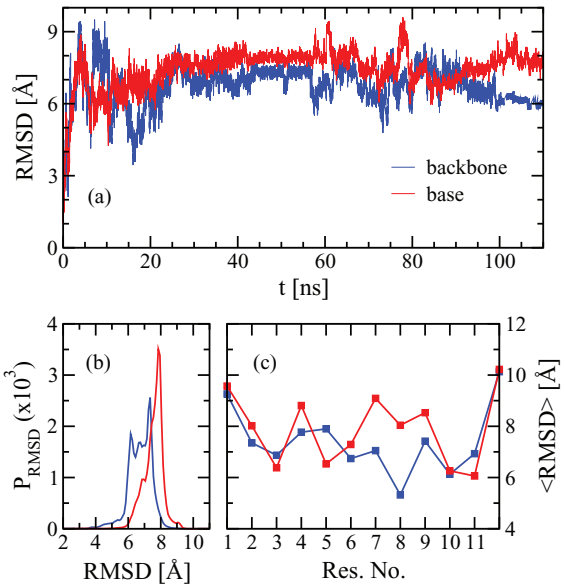

FIG. 2. (a) Time evolutions and (b) the distributions of the RMSDs of the backbone (blue curves) and the bases (red curves) of the ss-DNA molecule with respect to its initial configuration. (c) The time-averaged RMSD values of the backbone and base moieties of each of the ss-DNA residues.

the backbone and the bases are found to be 6.76 (±0.34) Å and 7.82 (±0.25) Å, respectively.

Radius of gyration (R_G) is another important parameter that can characterize the expansion and collapse (size and shape fluctuations) of a long-chain polymer, such as the ss-DNA. We have calculated the time evolution of R_G along with the corresponding distribution, as displayed in Fig. 3(a). A rapid decrease of R_G within about 10 ns of the simulation indicating transformation of the expanded initial conformation of the DNA to collapsed globular coil-like conformations is evident from the figure. The DNA molecule is found to maintain such flexible globular form for the rest (about 100 ns) of

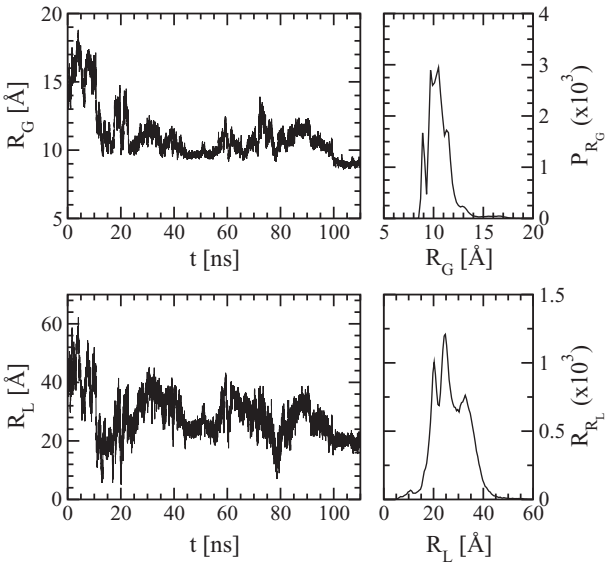

FIG. 3. Time evolutions and distributions of the radius of gyration, R_G (top panel), and the end-to-end distance, R_L (bottom panel), of the ss-DNA molecule.

the simulation with an average R_G value of 10.5 (±0.5) Å. In Fig. 3(b) we show the variation of the end-to-end distance (R_L) of the DNA molecule and its distribution. R_L is defined as the distance between the oxygen atoms of the 3′ and 5′ positions of the two terminal sugar residues. A correlation between the time evolutions of R_L and R_G can be easily seen from the figure. Like R_G, the rapid collapse of the DNA chain from extended stacked conformation to the globular coil form is accompanied by significant decrease in R_L. Once again, the flexibility of the chain in its collapsed form is evident from oscillations in R_L values. The calculated average R_L value over the last 100 ns of the trajectory has been found to be 26.7 (± 3.2) Å.

It is apparent from the above discussion that the initial base-stacked structure of the ss-DNA chain collapses to the globular coil form due to breaking of sequential base-stacking. Further, conformational rearrangements of the residues may lead to the formation of new structural motifs with the possibility of non-sequential stacking of the bases. Such non-sequential base-stacking, if any, is expected to stabilize the ss-DNA in its collapsed form. Here, we explore the evidence for the existence of such structural motifs during the time scale of the simulation. Following the work by Trylska and co-workers,[29] we defined the bases of two residues (sequential or non-sequential) to form a stacked pair if the distance (R_S) between their centers of masses (considering only the ring atoms) is within 5 Å. Measurements of R_S between all possible base pairs reveal that the G10 · · C11 sequential base pair remains stacked for about 60% of the simulation time. However, interestingly, formation of the collapsed coil form is found to be associated with two non-sequential base pair stacking of G2 · · A5 and A5 · · T8, for about 50% and 65% of the simulation time, respectively. A few representative configurations of these base pairs forming stacked arrangements along the DNA strand as obtained using the distance criterion are displayed in Fig. 4. The figure also shows superimposed configurations of the nucleobases along the trajectory during which they are found to be arranged in stacked manners. Note the nearly parallel arrangements of the base pairs in the figure. This shows that an upper limit of 5 Å for R_S as employed here can identify the base-stacking arrangements along a DNA oligomer reasonably well. Time evolutions of R_S for these base pairs along with their respective interaction energies (E_S) over the entire simulated trajectory are shown in Fig. 5. It may be noted that in our calculations E_S includes both the electrostatic and the van der Waals interaction components between the tagged pairs of nucleobases. The percentage populations (P_P) of these stacked base pairs, average center-of-mass distances ($\langle R_S \rangle$), and interaction energies ($\langle E_S \rangle$) between them in the stacked forms are listed in Table I. The interaction energies of the stacked base pairs in the ss-DNA molecule as obtained from our simulation have been found to agree qualitatively with quantum mechanical *ab initio* calculations of stacked dimers of nucleic acid bases.[30] The results clearly indicate that noticeable populations of a few stacked base rings contribute to the stabilization of the collapsed coil form of the ss-DNA molecule. This is an important observation as under physiological conditions ss-DNAs are believed to contain stacked domains interspaced by random coils.[12]

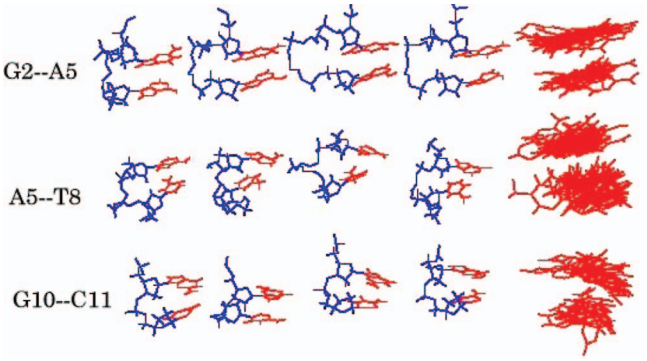

FIG. 4. A few representative configurations of selected base pairs of the ss-DNA molecule forming stacked arrangements with sufficiently long life times as obtained from the simulation. Superimposed configurations of those base pairs during the time periods when they are found to be in stacked forms are also shown. The atom coloring scheme is same as that in Fig. 1.

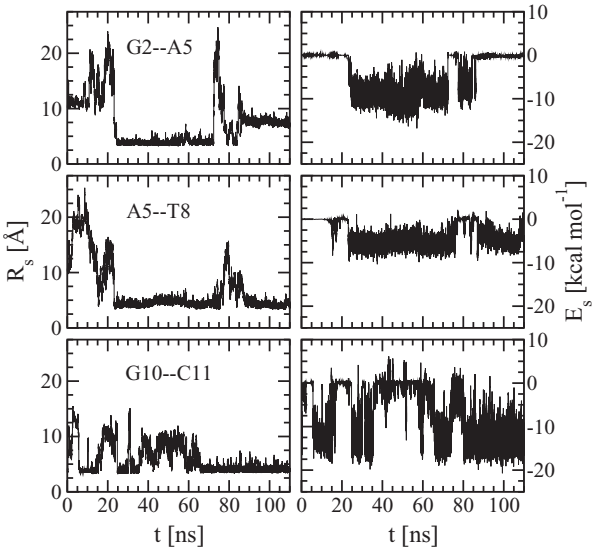

FIG. 5. Time evolutions of the center-of-mass distances, R_S (left panel), and the interaction energies, E_S (right panel), between those base pairs of the ss-DNA molecule that remain stacked for a significant duration of the simulation.

B. Configurational entropy of the ss-DNA

In this section we measure the configurational entropy of the ss-DNA molecule following the methodology as suggested by Schlitter.[31] This method based on one-dimensional quantum mechanical harmonic oscillator approximation is

TABLE I. The percentage populations (P_P), average center-of-mass distances ($\langle R_S \rangle$), and interaction energies ($\langle E_S \rangle$) between different stacked base pairs. The values in the parentheses are the standard deviations.

Base pair	P_P	$\langle R_S \rangle$ (Å)	$\langle E_S \rangle$ (kcal mol^{-1})
G2 · · A5	50	3.8 (0.06)	− 8.85 (1.28)
A5 · · T8	65	4.2 (0.07)	− 5.34 (0.76)
G10 · · C11	60	3.9 (0.05)	− 11.43 (0.99)

capable of calculating absolute entropies of macromolecules directly from their Cartesian coordinates.[32] According to Schlitter's method, the absolute entropy (S_{abs}) can be approximated as

$$S_{abs} < S = \frac{1}{2} k_B \ln \det \left[\mathbf{1} + \frac{k_B T e^2}{\hbar^2} \mathbf{M}^{1/2} \boldsymbol{\sigma} \mathbf{M}^{1/2} \right], \quad (1)$$

where k_B is Boltzmann's constant, T is the temperature, e is Euler's number, $\mathbf{M}$ is the $3N$ dimensional diagonal matrix containing N atomic masses of the solute, $\boldsymbol{\sigma}$ is the covariance matrix of atom positional fluctuations, and $\hbar$ is Planck's constant divided by 2π. Note that the inequality in the equation arises as Schlitter's entropy provides upper bound to the true absolute entropy (S_{abs}) of the system.[31] The covariance matrix elements (σ_{ij}) are given by

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle, \quad (2)$$

where x_i denote the Cartesian coordinates of the macromolecular atoms after removal of the center-of-mass translation and rotation around the center-of-mass with respect to the reference structure of the macromolecule. This ensures that the calculated entropy S is the configurational entropy of the molecule. In this calculation, we have considered the initial configuration of the analyzed trajectory as the reference structure for the ss-DNA. The covariance matrix elements are calculated by time averaging over the simulated configurations. Further details of this method can be found in earlier works.[31,33,34] Since the matrix $\mathbf{1} + \frac{k_B T e^2}{\hbar^2} \mathbf{M}^{1/2} \boldsymbol{\sigma} \mathbf{M}^{1/2}$ is a symmetric positive-definite one, its determinant has been evaluated using a triangularization procedure, known as the Cholesky decomposition method.[35] It may be noted that one particular advantage of this method is that one can calculate the entropy contribution for a particular component of a macromolecule like ss-DNA. It may be noted at this stage that Andricioaei and Karplus[35] developed a similar quasiharmonic method that can also measure entropy directly from Cartesian coordinate covariance matrices.

We have estimated the cumulative configurational entropy of the ss-DNA oligomer by including its non-hydrogen heavy atoms, as shown in Fig. 6. The calculations are performed with the assumption that the conformational sampling of the ss-DNA is complete within the time scale of the present simulation. In addition, the entropy contributions originating from the backbone and the bases of the DNA chain are also calculated separately and shown in the figure. The calculations are performed over the equilibrated trajectory of the system. Note that, we have used $T = 301$ K, the average temperature of the system as obtained from the equilibrated trajectory to estimate the entropies using Eq. (1). Rapid entropy buildups and convergence to steady values for all cases are clearly evident from the figure. This is a signature of near complete sampling of the conformational space by the DNA molecule. The cumulative entropy of the whole ss-DNA molecule has been found to be 2 kcal K^{-1} mol^{-1}, while that for its backbone and bases are found to be 1.32 kcal K^{-1} mol^{-1} and 1.09 kcal K^{-1} mol^{-1}, respectively. Such noticeable entropy gained by the ss-DNA molecule is consistent with its flexible globular coil-like conformations as discussed earlier. Besides, the difference between the entropy of the whole DNA molecule and the sum of the entropies of the backbone and the bases as evident from the data signify correlated flexibilities of the backbone and the nucleobases of the DNA molecule in its single stranded form.

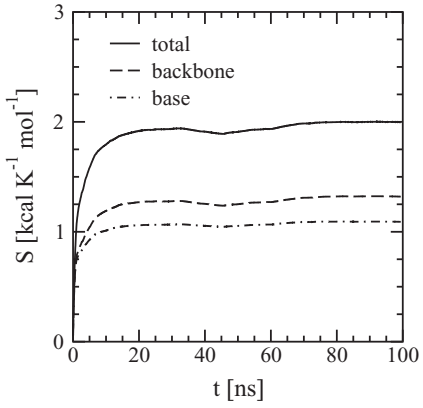

FIG. 6. Cumulative configurational entropy of the whole ss-DNA molecule, as well as that separately for its sugar-phosphate backbone and the bases. The calculations are carried out by considering the non-hydrogen atoms of the respective components.

C. Ss-DNA–water interactions

In an aqueous environment, the polar and charged sites present in the nucleobases and the backbone phosphate and sugar groups of the ss-DNA chain are expected to interact strongly with the surrounding solvent molecules. Such strong DNA–water interactions are expected to be crucial in controlling the conformational fluctuations of the DNA oligomer as discussed in Sec. III A. The local properties of the water molecules hydrating the ss-DNA chain should also be influenced by the DNA residues. This may lead to the surrounding water molecules behaving differently than pure bulk water. In this section, we probe the local structure, ordering, and interaction energies of the water molecules around the ss-DNA and compare them with that for water in pure bulk state.

1. Structural arrangements of water

Presence of a charged macromolecule like ss-DNA in polar aqueous medium should lead to restructuring of water molecules around it. To probe that, we have calculated the pairwise correlation function (g(r)) of the water oxygen atoms with respect to the DNA phosphorus (P) and oxygen (O) atoms. The results are shown in Fig. 7. Presence of sharp first peaks around 2.5 Å from the backbone oxygen atoms and around 4 Å from the phosphorus atoms clearly indicates structuring of water molecules in the first hydration layer around the DNA due to its strong interaction with water. It is evident that such interaction results in higher packing of water molecules in the volume available at the interface. It may be noted that similar higher packing of water molecules near the surface of biomolecules have been observed in x-

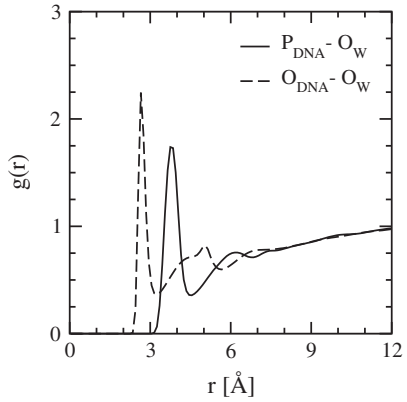

FIG. 7. Pairwise correlation function, g(r), of the water oxygen atoms as a function of distance from the phosphorus (solid line) and the oxygen atoms (dashed line) of the ss-DNA molecule.

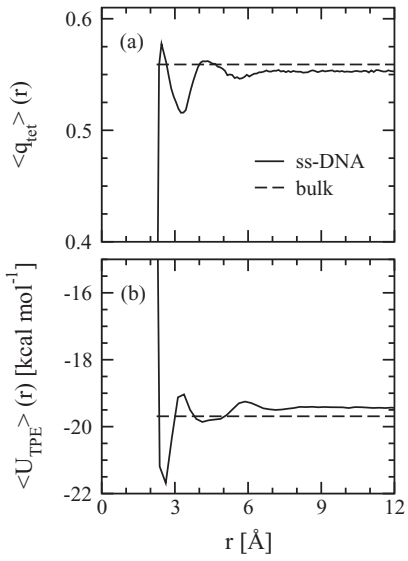

FIG. 8. (a) Average tetrahedral order parameter, $\langle q_{tet}\rangle$(r), and (b) the tagged potential energy (TPE), $\langle U_{TPE}\rangle$(r), of water molecules as a function of distance from the ss-DNA molecule. The corresponding average values for pure bulk TIP3P water at room temperature are marked for comparison.

ray and neutron scattering experiments,[36] as well as in simulation studies.[37,38] The DNA oxygen atoms also exhibit a noticeable second peak around 5 Å as evident from the figure. It is important to note that in general such pairwise distributions for the ss-DNA oligomer as obtained from our calculations are similar to that observed for ds-DNAs.[39] We have calculated the average number of water molecules present in the first coordination shells around the DNA phosphorus and oxygen atoms. This is obtained by integrating the g(r) curves for the oxygen and phosphorus atoms within 3.3 Å and 4.5 Å, respectively. On average 3 water molecules are found within the first coordination shells of each of the DNA oxygens, while 7-8 water molecules are found around each of the phosphorus atoms.

2. Water ordering and energetics

We now explore the relative tetrahedral ordering of water molecules at the surface of the ss-DNA chain with respect to that of water in pure bulk state. This is done by calculating the order parameter, q_{tet}, for the water molecules. q_{tet} is defined as[40–42]

$$q_{tet} = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(\cos\psi_{jk} + \frac{1}{3}\right)^2, \qquad (3)$$

where ψ_{jk} is the angle between the bond vectors $\mathbf{r}_{ij}$ and $\mathbf{r}_{ik}$, where j and k are the four nearest neighbor atoms of the ith water molecule. The calculations are carried out based on the assumption that the non-hydrogen atoms of the DNA strand can also be the nearest neighbors for the water molecules close to the surface. In Fig. 8(a) we show the variation of the average tetrahedral order parameter ($\langle q_{tet}\rangle$(r)) of water molecules as a function of distance from the DNA. The average order parameter of 0.56 for pure TIP3P water as obtained from our calculation is also marked in the figure for comparison. It can be seen that water molecules within a short distance of about 2 Å from the DNA exhibit q_{tet} values slightly higher than that for pure bulk water. These are essentially the first layer of water molecules that are bound to the DNA chain. Flexibility of the DNA backbone and the presence of oxygen atoms on it allow these bound water molecules to orient appropriately and satisfy their tetrahedral ordering. Interestingly, with increase in distance, q_{tet} values first decrease within 3–4 Å from the surface before attaining the limiting value close to that of bulk water beyond 5–6 Å. Importantly, it is found that reduced tetrahedral ordering around 3–4 Å from the surface of the ss-DNA is associated with increased water structuring as discussed in Sec. III C 1 (Fig. 7). This signifies that realignment of water molecules necessary for their higher packing in the region around 3–4 Å away from the ss-DNA leads to breaking of water tetrahedral ordering.

The ordering patterns of water molecules around the ss-DNA chain should be correlated with the corresponding tagged potential energies (TPE), U_{TPE}. This is essentially the binding energy of a tagged water molecule, and is defined as the interaction energy of the water with rest of the system.[43,44] We have calculated the variation of the average TPE ($\langle U_{TPE}\rangle$(r)) of water molecules as a function of distance from the DNA strand, as shown in Fig. 8(b). Once again, the average TPE of pure TIP3P water is marked in the figure for comparison. A deep minimum around 2–3 Å from the DNA molecule followed by a gradual approach to the limiting value close to that of pure bulk water (-19.7 kcal mol^{-1}) beyond 5–6 Å can be seen. The $\langle U_{TPE}\rangle$(r) value of about -22 kcal mol^{-1} at the minimum corresponds to the strongly bound first layer of water molecules at the surface, as discussed earlier. Importantly, a correlation between $\langle q_{tet}\rangle$(r) and $\langle U_{TPE}\rangle$(r) for water molecules present in the first hydration layer (5–6 Å) around the ss-DNA is evident from the figure. The maximum in $\langle q_{tet}\rangle$(r) curve is found to match with the minimum in $\langle U_{TPE}\rangle$(r) and vice versa. Such anticorrelation between these two quantities exists for small peptides and proteins.[38,44] The present results demonstrate that the existence of such anticorrelation is a general phenomenon near the surface of any biomolecule irrespective of its chemical and structural features.

D. Effect of salt on DNA conformation

One crucial issue in DNA hydration is the effect of additional salt on its conformations.[45,46] This has relevance in biology as cations present in cellular environment shield the negative charges on the backbone phosphate groups, thereby reducing the repulsive force between them. In this section we performed preliminary calculations to explore the effect of added salt on ss-DNA conformations. For that an additional simulation of the ss-DNA molecule with identical base sequence and initial configuration as described in Sec. II has been carried out in presence of 500 mM sodium chloride (NaCl) salt. To achieve that, we have followed the same protocol as described in Sec. II and inserted 65 Na^+ and 65 Cl^- ions (in addition to 11 neutralizing Na^+ ions) into the solution of ss-DNA in a cubic cell of 60 Å edge length containing 7479 water molecules. Following the same procedure, the system was equilibrated at 300 K first under NPT conditions for 5 ns and then under NVT conditions over same duration. The simulation cell volume attained a steady value during the first 5 ns of the NPT run with an edge length of around 61 Å. This was followed by over 60 ns of production run under NVE ensemble conditions with a calculated average temperature of 301.2 (±1.2) K.

To obtain a preliminary knowledge of the structural deviations of the ss-DNA molecule in presence of added salt, we have computed the RMSDs of the sugar-phosphate backbone and the bases of the DNA with respect to its initial structure. The time evolutions of the RMSDs and their distributions are shown in Figs. 9(a) and 9(b). The average RMSD values are found to be 8.88 (±0.56) Å and 9.58 (±0.61) Å for the DNA backbone and the bases, respectively. Compared to the results described in Sec. III A (Fig. 2), it is clear that the presence of salt leads to increased deviations of the ss-DNA molecule from its extended helical initial state configuration. Recently, Chen *et al.*[46] measured salt-dependent end-to-end distance (R_L) of ss-DNAs using FRET spectroscopy. By performing experiments at different NaCl concentrations they showed that reduced repulsion between the backbone phosphate groups results in decreased R_L values of the DNA strands. To verify that from our simulation study, we have computed the time evolution of R_L and its distribution in presence of added NaCl, as shown in Figs. 9(c) and 9(d). The calculated average R_L value has been found to be 16.1 (±4.0) Å, which is about 40% lower than that estimated in absence of excess salt (Sec. III A). Such significant decrease in end-to-end distance of the DNA strand in presence of the salt agrees well with experimental data[46] and shows enhanced compaction of the molecule with formation of more collapsed structure due to screening of negative charges along the backbone. The results are also in agreement with salt-induced structural collapse of ss-DNAs of different sequence lengths as observed in recent small-angle x-ray scattering[47] and Monte Carlo simulation[48] studies.

We demonstrated earlier (Sec. III A) that the collapse of the ss-DNA strand from its initial extended helical form is associated with formation of stable non-sequentially stacked base pairs. The increased compaction of the strand with modified conformation in presence of excess cations from the added salt is expected to rearrange such stacking pattern. Using the criterion as described earlier, we have monitored the distance R_S (defined earlier) between all possible base pairs. It is noticed that the collapsed conformation of the DNA strand in presence of salt is associated with two prominent non-sequentially stacked base pairs, A6· · ·C11 and T8· · ·G10, with overall stacking durations of about 88% and 78% of the total simulation time, respectively. In addition, the sequential base pair G4· · ·A5 is found to remain stacked for about 70% of the duration. In Fig. 10, we display the time evolutions of R_S for these three base pairs along with their interaction

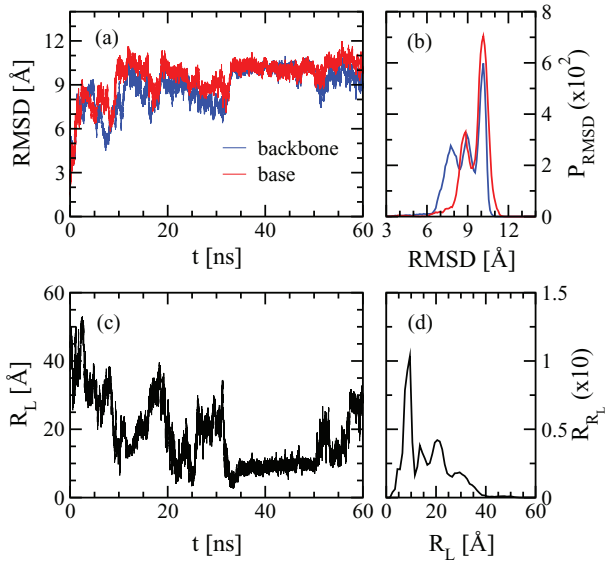

FIG. 9. (a) Time evolutions and (b) the distributions of the RMSDs of the backbone (blue curves) and the bases (red curves) of the ss-DNA molecule with respect to its initial configuration in presence of 500 mM NaCl concentration. (c) Time evolutions and (d) distributions of the end-to-end distance, R_L, of the ss-DNA molecule in presence of the salt.

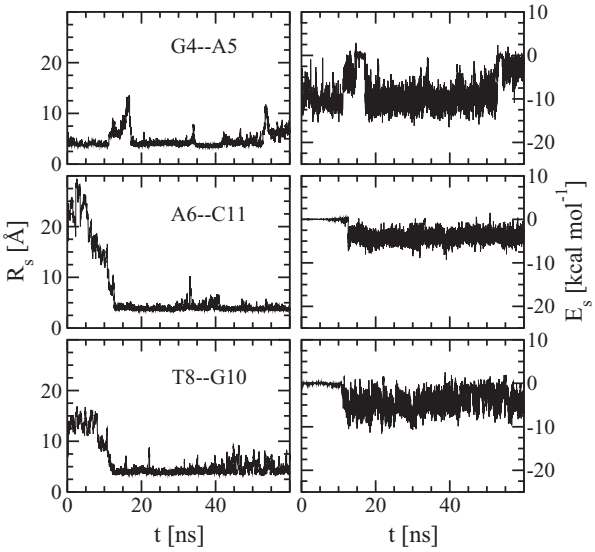

FIG. 10. Time evolutions of the center-of-mass distances, R_S (left panel), and the interaction energies, E_S (right panel), between those base pairs of the ss-DNA molecule that remain stacked for a significant duration of the simulation in presence of 500 mM NaCl concentration.

energies (E_S). Comparison of the results with that described earlier (Fig. 4) indicates that the presence of additional salt in the solution results in formation of non-sequentially stacked base pairs with longer lifetimes. This in turn is likely to be responsible for increasingly collapsed conformation of the DNA strand with reduced end-to-end distance in presence of salt as discussed above. Our results agree well with earlier studies of ss-DNA structuring at high salt concentrations.[14]

We would like to mention at this stage that in this preliminary calculations we have studied the effect of salt on ss-DNA conformation at one concentration. However, systematic studies by varying salt concentrations are necessary to obtain a more quantitative understanding of such biologically important issue involving nucleic acids. Efforts are on in our laboratory to investigate this in further detail.

IV. CONCLUSIONS

In this work, we have performed atomistic MD simulation to study the microscopic properties of an aqueous solution of a ss-DNA molecule. During the simulation, the DNA strand has been found to undergo transformation from an initial ordered base-stacked form to a fluctuating collapsed globular coil-like conformation. It is noticed that the DNA bases with relatively large RMSD values are slightly more flexible as compared to the covalently linked sugar-phosphate backbone. The calculations revealed that formation of non-sequential G2 · · A5 and A5 · · T8 base pair stackings contributes to the stabilization of the collapsed coil form of the ss-DNA. Form a preliminary analysis in presence of 500 mM NaCl salt, it is demonstrated that screening of negative charges along the backbone by excess cations leads to increasingly collapsed conformation of the ss-DNA strand with enhanced non-sequential base stackings and reduced end-to-end distance. This is found to be in agreement with recent experimental studies.[46,47] The difference between the configurational entropy of the whole DNA molecule and the sum of the contributions of the sugar-phosphate backbone and the nucleobases demonstrated existence of noticeable correlation between the DNA backbone and bases in single-stranded form.

The calculations further showed significant structuring of water molecules within the first hydration layer around the ss-DNA. Interestingly, such higher packing of water molecules within a short distance of 3–4 Å from the surface of the DNA strand is found to be associated with their realignment and breaking of tetrahedrally ordered arrangement. Further, it is noticed that the ordering of the surface water molecules is correlated with their tagged potential energies. It would be interesting to explore the dynamic properties of water molecules close to the surface of the DNA strand. Besides, effects of variation of base sequence in presence and absence of additional salt on the conformational behavior of the ss-DNA and the microscopic properties of surrounding water molecules are other important issues with crucial biological relevance that need to be addressed. We are currently exploring these in great detail.

ACKNOWLEDGMENTS

This study was supported in part by a grant from the Department of Science and Technology (DST) (SR/S1/PC-23/2007), Government of India. Part of the work was carried out using the computational facility created under DST-FIST programme (SR/FST/CSII-011/2005). K.C. thanks CSIR, Government of India, for providing a scholarship.

[1] L. Zou and S. J. Elledge, Science **300**, 1542 (2003).
[2] C. Rivetti, C. Walker, and C. Bustamante, J. Mol. Biol. **280**, 41 (1998).
[3] J. B. Mills, E. Vacano, and P. J. Hagerman, J. Mol. Biol. **285**, 245 (1999).
[4] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, Biophys. J. **86**, 2530 (2004).
[5] V. A. Bloomfield, D. M. Crothers, and I. Tinoco, Jr., *Nucleic Acids: Structures, Properties and Functions* (University Science Books, 2000).
[6] S. Cui, C. Albrecht, F. Kühner, and H. E. Gaub, J. Am. Chem. Soc. **128**, 6636 (2006).
[7] M. Bastos, V. Castro, G. Mrevlishvili, and J. Teixeira, Biophys. J. **86**, 3822 (2004).
[8] K. Miyamoto, K. Onodera, R. Yamaguchi, K. Ishibashi, Y. Kimura, and M. Niwano, Chem. Phys. Lett. **436**, 233 (2007).
[9] X. Liang, H. Kuhn, and M. D. Frank-Kamenetskii, Biophys. J. **90**, 2877 (2006).
[10] T. Kaji, S. Ito, S. Iwai, and H. Miyasaka, J. Phys. Chem. B **113**, 13917 (2009).
[11] Y. Zhang, H. Zhou, and Z. C. Ou-Yang, Biophys. J. **81**, 1133 (2001).
[12] A. Buhot and A. Halperin, Phys. Rev. E **70**, 020902(R) (2004).
[13] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).
[14] S. V. Kuznetsov, C. Ren, S. A. Woodson, and A. Ansari, Nucleic Acids Res. **36**, 1098 (2008).
[15] Z. Tan and S. Chen, Biophys. J. **95**, 738 (2008).
[16] M. N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette, Phys. Rev. Lett. **89**, 248102 (2002).
[17] J. M. Martìnez, S. K. C. Elmroth, and L. Kloo, J. Am. Chem. Soc. **123**, 12279 (2001).
[18] S. Sen and L. Nilsson, J. Am. Chem. Soc. **123**, 7414 (2001).
[19] S. Chatterjee, B. Gersten, S. Thakur, and A. Burin, Mol. Simul. **33**, 573 (2007).
[20] J. C. Philips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorsid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).
[21] N. Foloppe and A. D. MacKerell, Jr., J. Comput. Chem. **21**, 86 (2000).
[22] A. D. MacKerell, Jr. and N. Banavali, J. Comput. Chem. **21**, 105 (2000).
[23] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[24] V. Tereshko, G. Minasov, and M. Egli, J. Am. Chem. Soc. **121**, 470 (1999).
[25] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, J. Chem. Phys. **103**, 4613 (1995).
[26] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford, Clarendon, 1987).
[27] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[28] G. Vesnaver and K. J. Breslauer, Proc. Natl. Acad. Sci. U.S.A. **88**, 3569 (1991).
[29] J. Panecka, C. Mura, and J. Trylska, J. Phys. Chem. B **115**, 532 (2011).
[30] J. Sponer, J. Leszczynski, and P. Hobza, J. Phys. Chem. **100**, 5590 (1996).
[31] J. Schlitter, Chem. Phys. Lett. **215**, 617 (1993).
[32] E. Suárez, N. Díaz, and D. Suárez, J. Chem. Theory Comput. **7**, 2638 (2011).
[33] S. D. Hsu, C. Peter, W. F. van Gunsteren, and A. M. J. J. Bonvin, Biophys. J. **88**, 15 (2005).
[34] J. Dolenc, R. Baron, C. Oostenbrink, J. Koller, and W. F. van Gunsteren, Biophys. J. **91**, 1460 (2006).
[35] I. Andricioaei and M. Karplus, J. Chem. Phys. **115**, 6289 (2001).
[36] D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai, Proc. Natl. Acad. Sci. U.S.A. **95**, 2267 (1998).
[37] N. Smolin and R. Winter, J. Phys. Chem. B **108**, 15928 (2004).
[38] S. K. Sinha and S. Bandyopadhyay, J. Chem. Phys. **134**, 115101 (2011).
[39] S. Pal, P. K. Maiti, and B. Bagchi, J. Chem. Phys. **125**, 234903 (2006).

[40]P. L. Chau and A. J. Hardwick, Mol. Phys. **93**, 511 (1998).
[41]J. R. Errington and P. G. Debenedetti, Nature (London) **409**, 318 (2001).
[42]P. Kumar, S. V. Buldyrev, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **106**, 22130 (2009).
[43]A. Mudi and C. Chakravarty, J. Phys. Chem. B **110**, 8422 (2006).
[44]M. Agarwal, H. R. Kushwaha, and C. Chakravarty, J. Phys. Chem. B **114**, 651 (2010).
[45]C. Song, Y. Xia, M. Zhao, X. Liu, F. Li, Y. Ji, B. Huang, and Y. Yin, J. Mol. Model. **12**, 249 (2006).
[46]H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, Proc. Natl. Acad. Sci. U.S.A. **109**, 799 (2012).
[47]A. Y. L. Sim, J. Lipfert, D. Herschlag, and S. Doniach, Phys. Rev. E **86**, 021901 (2012).
[48]F. Wang, Y. Wu, and Z. Tan, Biopolymers **99**, 370 (2013).