

# Base- and Structure-Dependent DNA Dinucleotide–Carbon Nanotube Interactions: Molecular Dynamics Simulations and Thermodynamic Analysis

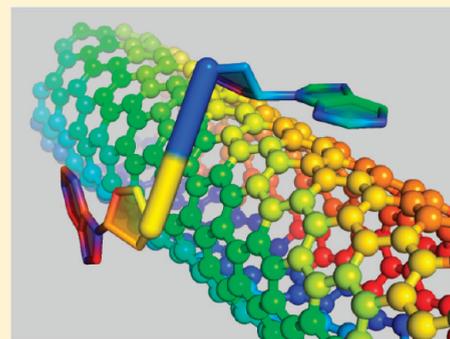
Zhengtao Xiao,<sup>†,‡</sup> Xia Wang,<sup>†,‡</sup> Xue Xu,<sup>†,‡</sup> Hong Zhang,<sup>‡</sup> Yan Li,<sup>§</sup> and Yonghua Wang<sup>\*,†,‡</sup>

<sup>†</sup>Center of Bioinformatics and <sup>‡</sup>College of Life Sciences, Northwest A&F University, Yangling, 712100, Shaanxi, China

<sup>§</sup>Department of Materials Science and Chemical Engineering, Dalian University of Technology, Dalian, 116023, Liaoning, China

**S** Supporting Information

**ABSTRACT:** Wrapping of single-wall carbon nanotubes (SWCNTs) by single-stranded DNA (ssDNA) was found to be sequence-dependent, offering properties such as the facilitation of SWCN sorting, ultrafast DNA sequencing, and construction of chemical sensors. Although the interactions of nucleic acids with SWCNTs have been studied thoroughly, the DNA–CNT hybrid especially for the oligonucleotides containing more than one nucleotide has not yet been fully understood. To address this, we have examined new and unconventional DNA dinucleotides involving all 16 combinations of two DNA nucleotides attached with chiral (8,4) and armchair (6,6) SWCNTs using all-atom molecular dynamics simulations and thermodynamic analyses. The 16 dinucleotides with different sequence compositions are found to readily adsorb onto SWCNTs and display interesting binding behaviors such as base flipping, local dynamic stability of structure, and conformational shifting. Four dinucleotides, i.e., AC, AG, GC, and GT, share similar dynamic properties (base turning and conformation transformation) in (8,4) and (6,6) systems. The different dynamic profiles between the compositional isomers with the reverse sequences such as the AG and GA show that the sequence order also impacts the dynamic recognition and binding energy of the ssDNA–CNT hybrid. Clustering-analysis-derived representative conformations imply that general dinucleotides are inclined to spread on the SWCNT surface, and the adjacent bases tend to stretch away from each other. Dinucleotides like AC, AT, CG, CT, GC, GG, TA, TC, TG, and TT adopt similar geometries on both CNTs, suggesting that their structures are not predominantly influenced by the nanotube chirality but controlled by the identity of the base sequence, sequence order, and the basic cylindrical structure of SWCNT. In addition, the nucleotide bases have a high degree of orientational order on the nanotube surface and the orientations of each base are significantly affected by the sequence of DNA and the chirality of nanotube, emphasizing that the structural order plays an important role in the binding of DNA and CNTs. However, our energy analysis shows that due to small different curvatures of the CNT surface, the binding affinity of most dinucleotides (except AG, CA, CG, and TG) to the chiral and armchair nanotube is not significantly different. Generally, the dinucleotides constituted with purine and thymine exhibit the lowest binding free energy, resulting from the van der Waals interactions and solvent effects. The thymine-based dinucleotides reduce the solvation free energy of the SWCNT in aqueous solution more effectively as compared to other bases. The present work also demonstrates that the total binding free energy is sequence specific but not merely a sum of individual base–SWCNT binding free energies.



## 1. INTRODUCTION

Nowadays, carbon nanotubes (CNTs) have undergone exponential growth with broad applications in nanotechnology and nanoscience due to their extraordinary electronic, thermal, mechanical, and transport properties.<sup>1,2</sup> As one of the building blocks of nanotechnology, the area of CNT–biopolymer composites has been progressing extremely rapidly in recent years. One such material of contemporary interest is the DNA–CNT hybrid, which consists of a single-wall carbon nanotube (SWCNT) coated with a self-assembled monolayer of single-stranded DNA (ssDNA).

In contrast to most other polymers, functionalized CNTs by DNA will improve their biocompatibility, solubility, and selective

binding to the biotargets,<sup>3–5</sup> and such supermolecules hold a remarkable set of technologically useful properties. For example, SWCNTs decorated with ssDNA display remarkable chemical sensing capabilities<sup>6,7</sup> with sequence dependent chemical recognition,<sup>8</sup> making them suitable candidates for biosensors. Furthermore, the array of SWCNTs periodically arranged to fit into the major groove of the DNA offers promise as a very sensitive nanoscale electronic device,<sup>9</sup> and such a combined system can be used as a device for ultrafast DNA sequencing. SWCNTs are also

**Received:** April 30, 2011

**Revised:** August 23, 2011

**Published:** September 16, 2011

used for gene delivery since they can control the release of DNA from the SWCNT surface and freely reach their intended biological destinations.<sup>10,11</sup> A deeper understanding of these biopolymer hybrid nanomaterials lies at a powerful advancing frontier of fundamental research in nanoscience and will open an incredible range of applications of DNA–CNTs. Recent work has shown that the adsorption of ssDNA onto a mixture of different types of nanotubes can be used to separate and purify SWCNTs through ion-exchange chromatography.<sup>12–14</sup> The purification is based on the different stabilities of the ssDNAs wrapping around individual SWCNTs.

Due to the important applications of these composite nanomaterials, intense investigations have been performed to understand their structural, physical, and chemical properties. Atomic force microscopy (AFM) measurements show that  $d(\text{GT})_n$ –CNT hybrids have much more uniform periodic structures with a regular pitch of  $\sim 18$  nm, and the structure of  $d(\text{GT})_n$  on CNTs appears to be very sensitive to minute changes in the composition of the nucleotide bases. These features led Zheng et al. to propose that  $d(\text{GT})_n$  self-assembles into a helical structure around individual nanotubes in such a way that the electrostatics of the DNA–CNT hybrid depends on the tube diameter and electronic properties, enabling the nanotube separation by anion exchange chromatography.<sup>15</sup> Other experimental methods such as the measurements of the transmission electron microscopy (TEM) also suggested the helical wrapping of a ssDNA molecule around the SWCNT.<sup>16,17</sup> Takahashi et al. found that disentangled double-stranded DNA (dsDNA) immobilizing at the surface of multiwall carbon nanotube (MWCNT) formed the arch-like structures and proposed that the DNA wrapping was strongly influenced by the curvature of the sidewall of the nanotube.<sup>18</sup>

Although experimental techniques provide valuable structural information on the DNA–CNT hybrid, its resolution is limited and the conclusively detailed interaction mechanism is hard to be determined.<sup>19</sup> Moreover, the molecular-level information on the thermodynamic, structural, and dynamic/kinetic aspects of SWCNT biomolecule interactions is lacking. In comparison with the substantial body of experimental work, computer simulations can provide detailed insights into the fundamental interactions between SWCNTs and ssDNA and help us establish new concepts for controlling/tuning the performance of such systems to facilitate the design and optimization of nanotube-based functional nanoscale devices. Previous theoretical studies showed that the ssDNA consisting of approximate 10 nucleotide bases can wrap around a CNT owing to the van der Waals attraction between them.<sup>20,21</sup> The ssDNA may exhibit a variety of configurations when interacting with a CNT, including right- and left-handed turns, “loops”, or disordered, kinked structures.<sup>22</sup> Unlike ssDNA, dsDNA adsorbs onto SWCNTs weakly in another fashion, which attaches to the surface of CNT via its hydrophobic end groups, and the adsorption process affects the A to B conversion of an A-DNA.<sup>19</sup> To better understand the nature of DNA–CNT self-assembly, calculations have been employed to investigate the single base–CNT interaction. The base molecules exhibit significantly different interaction strengths and follow the hierarchy of  $G > A > T > C$ . The stabilizing factor in the interaction between the base molecule and CNT is dominated by the  $\pi$ – $\pi$  stacking interactions.<sup>23–25</sup>

As mentioned above, the DNA–CNT combined system is a complicated, dynamic structure and the resulting hybrid structure is dependent on both the DNA sequence and the SWCNT structure. Moreover, this interaction can also be perturbed by the

base composition, backbone, and length of the DNA strand. To date the exact mechanisms of sequence selectivity of the DNA–CNT binding are still unclear, including detailed information of the binding modes, base orientation, and dependence on the DNA sequence. Prompted by this consideration, in the present work we examined a new and unconventional DNA called dinucleotide attached with SWCNT using molecular dynamic (MD) simulations for the first time. Dinucleotides represent the smallest unit for describing neighboring relationships between adjacent bases. The sequence alphabet consists of four bases A, T, G, and C which lead to 16 different dinucleotides: AA, AT, AC, AG, CA, CT, CC, CG, GA, GT, GC, GG, TA, TT, TC, and TG. Our study will focus on the following: (1) How do dinucleotides adsorb on a SWCNT? (2) What geometry is formed when dinucleotide binds to the CNT? (3) Do the bound bases have a preferred direction relative to the nanotube axis? (4) How do different dinucleotides produce different binding free energies to the surface of CNT? (5) Additionally, the adsorptions of dinucleotides on the (6,6) (metallic) SWCNT and the (8,4) (semiconducting) SWCNT of similar diameter and length are also investigated to gain insight into the effects of both the metallic and semiconducting SWCNTs on the nucleotide recognition.

## 2. MATERIALS AND METHODS

**2.1. Molecular Dynamics Simulations.** The AMBER force field and simulation package were employed for the molecular dynamics simulations, which had been successfully applied in the study of the DNA segments and carbon nanotube in solution.<sup>19</sup> The starting structures of 16 ssDNA dimers consisting of two bases, about 10 Å in length, were prepared using the Biopolymer package in Sybyl 6.9 (Tripos, Inc.). We constructed the armchair (6,6) and chiral (8,4) tubes with diameter of 8.2 Å and length of 33.4 Å using the TubeGen online version 3.3.<sup>26</sup> The AMBER99SB force field,<sup>27</sup> which is an all-atom potential including van der Waals (vdW), electrostatic, bond vibration, bond angle, and dihedral distortion energies, was used to model the DNA molecule. The carbon atoms in the SWCNT were modeled as uncharged Lennard-Jones (LJ) particles with benzene-like parameters.<sup>28</sup> The choice of uncharged carbon atoms for the SWCNT was also justifiable as validated by the recent study of Jordan and co-workers.<sup>29,30</sup> They considered extended  $\pi$ -systems as the models for the graphite. All the initial ssDNA dimers assumed a straight conformation with all bases approximately parallel to the CNT wall with the O4\* atoms facing toward the CNT. This process can ensure that the DNA accommodates more favorable vdW interactions with the SWCNT, since preliminary calculations on the binding energies between the DNA bases and the SWCNT suggested that it is energetically favorable for each DNA base to stack on the SWCNT.<sup>21</sup> The DNA–SWCNT systems were solvated in about 5308 to 5529 water molecules, with a typical starting simulation cell of about  $59 \times 72 \times 51$  Å<sup>3</sup>. The TIP3P potential was used for the simulation of waters,<sup>31</sup> and the periodic boundary conditions were applied in all three directions. The solvated box had water buffer layers of at least 15 Å thick between the solute surface and the simulation box boundary in all three directions. One Na<sup>+</sup> counterion was added to neutralize the charged ssDNA backbone for each system.

Molecular dynamics simulations were performed in the isothermal–isobaric ensemble<sup>32</sup> at 1 bar and 300 K using the AMBER 10 suite of programs.<sup>33</sup> The temperature was retained at 300 K using Langevin dynamics. The particle mesh Ewald

method<sup>34</sup> was used for long-range electrostatics with a grid spacing of 1 Å and fourth order interpolation. A cutoff of 10.0 Å was applied to the van der Waals and direct electrostatic interactions. The SHAKE algorithm<sup>35</sup> was used to constrain the bonds containing hydrogen atoms, while other bonds were modeled by the standard harmonic bond potential employed in the AMBER package. Each simulation included 9000 steps for energy minimization: 4000 steps of solvent relaxation and 5000 steps of solute relaxation. The structures were then heated and equilibrated over a period of 500 ps each. The method of Langevin dynamics was used to control the temperature throughout the simulation. All configurations were finally run for 40 ns of production with a time step of 2 fs, and the generated structures were stored in trajectory files every 10 ps. Visualization and analysis of the configurations were performed with the VMD package<sup>36</sup> and the ptraj module included in AMBER.

**2.2. Replica Exchange Molecular Dynamics (REMD) Simulation.** A replica exchange molecular dynamics (REMD) scheme,<sup>37</sup> as implemented in Amber10, was used to improve sampling and observe temperature-dependent effects. The initial conformations of the dimers on the nanotube for the REMD simulation were obtained from the copy of the initial structure in the previous section. First, 200 ps equilibration MD simulations for the complex at NPT ensemble at 1 bar from 298 to 412 K at 6 K/ps steps were performed. Subsequent to equilibration, an additional 20 ns of REMD simulation was run in the NVT ensemble using explicit water solvation. During the exchange phase, target temperatures were conditionally exchanged between pairs of replicas at adjacent temperatures every 1 ps, and each REMD simulation was run for 20000 exchange attempts (20 ns). The REMD simulations were set up under similar conditions as the corresponding standard MD runs.

**2.3. Clustering Analysis.** In order to reduce the vast number of DNA conformations sampled by the MD simulations and to characterize the main binding modes, clustering analysis was conducted by using an average-linkage method.<sup>38</sup> The average-linkage method defined the difference between two clusters as the average of all distances between points of cluster 1 to the points of cluster 2. Those clusters showing the smallest distance were merged. Subsequently, the structure closest to the center of a cluster was selected as the representative structure, which had the smallest rmsd values to all other structures within the cluster.

To determine the optimal number of templates, we increased the number of clusters from one to six. Selecting the appropriate cluster count was based on their metrics values for the obtained clusters: the Davies–Bouldin index (DBI), the pseudo-F-statistic (pSF), and the so-called “elbow criterion” (SSR/SST). These metrics plotted as a function of the cluster count can help locating the optimal cluster count, which had been described by Shao et al.<sup>38</sup> The cluster count, pointed out by the majority of the metrics as the optimal one for the system under study, was chosen and the representative structures of these clusters were analyzed. For further information on the procedures and other detailed analysis, please refer to the Supporting Information.

**2.4. Free Energy Calculation and Decomposition.** The calculation and decomposition of binding free energy for the 32 complexes were evaluated using the MM-GBSA (molecular mechanics general Borned surface area) method<sup>39,40</sup> as implemented in AMBER10. The MM-GBSA approach employed molecular mechanics, the generalized Born model, and solvent accessibility method to elicit the free energy from the structural information circumventing the computational complexity of the free energy simulations. It was parametrized within the additivity

approximation<sup>41</sup> wherein the net free energy change was treated as a sum of a comprehensive set of individual energy components, each with a physical basis. This method was attractive for the binding free energy calculation and decomposition because the pairwise nature of the GB equation allows the decomposition of the free energy into atomic contributions in a straightforward manner.<sup>42</sup> Briefly, in the MM-GBSA approach, the DNA–CNT binding free energy ( $\Delta G_{\text{binding}}$ ) for each snapshot was estimated as

$$\Delta G_{\text{binding}} = [G_{\text{complex}}] - [G_{\text{DNA}}] - [G_{\text{CNT}}] \quad (1)$$

The free energy of each of the above terms was calculated from

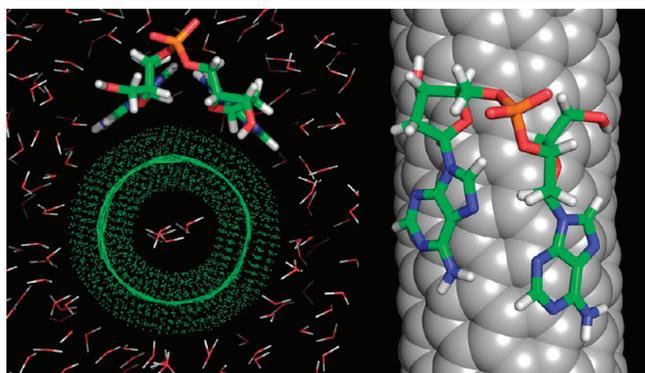
$$\Delta G_{\text{tot}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S \quad (2)$$

where  $E_{\text{MM}}$  is the molecular mechanics energy of the molecule expressed as the sum of the internal energy (bonds, angles and dihedrals) ( $E_{\text{int}}$ ), the electrostatic energy ( $E_{\text{ele}}$ ), and van der Waals ( $E_{\text{vdW}}$ ) terms.  $G_{\text{solv}}$  accounts for the solvation energy which can be divided into the polar and nonpolar parts. Obtaining the solvation free energy ( $G_{\text{solv}}$ ) from an implicit description of the solvent as a continuum is advantageous because it affords a solvation potential that is only a function of the solute's geometry, as discussed and implemented by Srinivasan et al.<sup>43</sup> As reported by other authors, the contribution of the entropy ( $T\Delta S$ ) was negligible because the difference of  $T\Delta S$  was very small considering the similarity of the systems. For detailed information on the calculations, please refer to the Supporting Information.

### 3. RESULTS AND DISCUSSIONS

**3.1. Overall Conformational Changes and rmsd Analysis.** Although the nature of the binding modes, and the preferential geometry assumed by the ssDNA molecules upon their bindings to the CNT are of particular interest, it is still elusive about the correlation of the physicochemical properties of the hybrids with their molecular structures.<sup>44</sup> To address this, we first analyzed the dynamics of the binding process. In the first 1 ns (equilibration part of the run), all bases are stacked onto the CNT surface, showing that the backbone is drawn close to the SWCNT and the nucleobases lie flat on the SWCNT surface efficiently with two adjacent bases staying away from each other because of the steric hindrance. In addition, the distance of individual nucleobases adsorbing to the SWCNT surface is 3.3–3.8 Å, similar to the one that was found for the neighboring planes in graphite.<sup>21</sup> This suitable distance leads to a structural reorganization of the DNA to attach perfectly to the surface of SWCNT. Subsequently, the adsorbed nucleotides are free to slide and rotate along the SWCNT axial and circumferential directions, with no bases desorbing from the surface of the nanotube found. During the simulations, adjacent bases are subjected to steric hindrances depending on their identities that render them to compete for the binding space on the SWCNT surface. Meanwhile, the backbone experiences a rearrangement upon the adsorption, which is responsible for wrapping the initially linear oligonucleotides around the curved surface of the SWCNT. Several distinct conformations characterize the hybrid's structure, indicating that the binding conformation of DNA is dependent on the sequence (with detailed discussions in section 3.3). In all cases, the O4' atom of sugar residues prefers to point toward SWCNT, which is in agreement with previous work.<sup>21</sup>

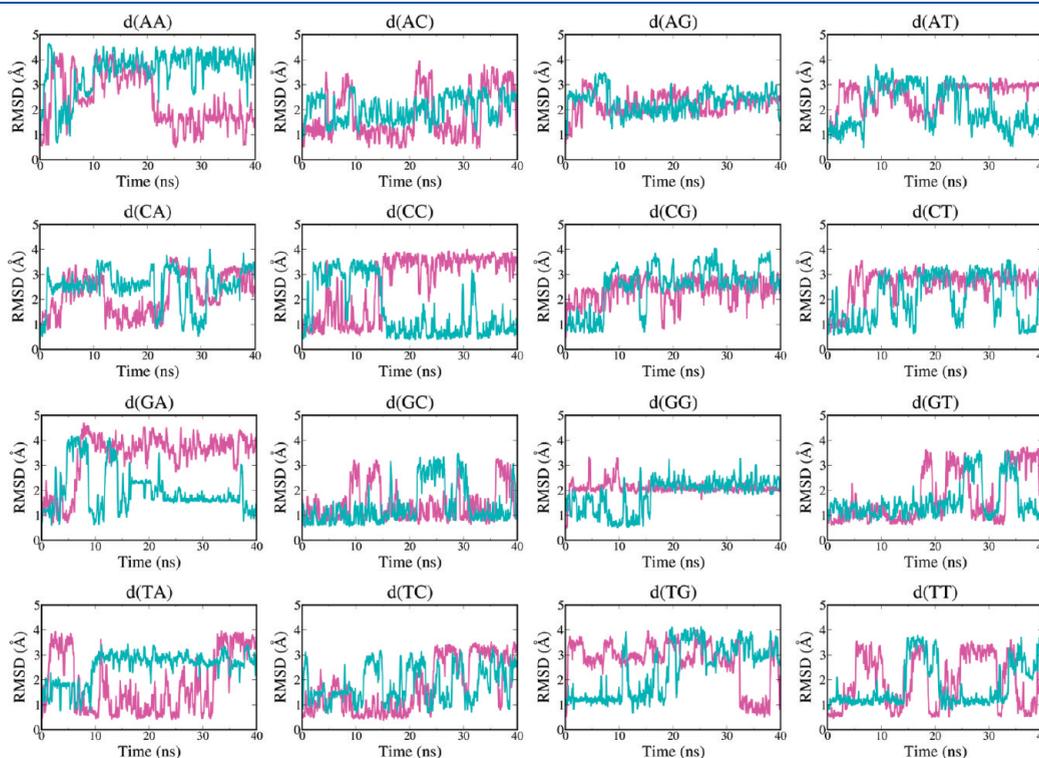
In order to get more intuitionistic scenes, the binding conformation of the d(AA) and CNT is given as a characteristic representation of the ensemble systems (see Figure 1). Upon equilibration in water at 300 K, the two bases of the d(AA) straddle the nanotube like a “saddle” and the sugar ring plane is parallel to the surface of the SWCNT. Because the base has a planar structure, the bound bases rest on the convex (outer) CNT surface, while polar groups on the periphery remain well exposed to the solvent, leading to more hydrations of phosphate groups. The stacking distance between the d(AA) base and the CNT surface is  $\sim 3.6$  Å, and the water envelope starts at a radial distance of  $\sim 6.41$  Å from the CNT axis. The presence of a bound base reduces the amount of the hydrophobic CNT surface area accessible to the solvent ( $2500\text{--}2477$  Å<sup>2</sup>), thereby increasing the



**Figure 1.** The simulation snapshots of the d(AA) interacting with (8,4) SWCNT for the top view (left image) and side view (right image) at the end of equilibration phase. The green dots denote the solvent-accessible surface of SWCNT. The red lines represent water molecules.

solubility of the SWCNT. Figure 1 also shows that the phosphorus (P) atoms are located at a distance of  $\sim 7.0$  Å away from the surface of the CNT and remain solvated in waters.

To investigate the detailed structural behavior of the ssDNA and test the stability of the simulations, we calculated the rmsd of the ssDNA polymer from the starting structure as a function of time to assess quantitatively the degree of the conformational change (shown in Figure 2). The rmsd variations of 16 oligonucleotides in the (8,4) hybrid systems are first investigated. In the MD trajectory, different individual ssDNA dimers exhibit different behaviors depending on their base sequences. It can be seen that the d(CC), d(GA), and d(TA) have high rmsd values (3.6, 4.0, 3.5 Å, respectively) relative to other dimers. From visualization and analysis of the DNA structure as a function of time, it appears that the relatively large rmsd is due to a shift of the bases. To explore the relative motions of different bases, the rotation angle of the base with respect to the starting structure with fitting of successive simulation frames on the backbone has been analyzed. The results are shown in Figure S1 (see Supporting Information). For the d(CC), the rotation angle of the base has a similar profile to the rmsd of the dimers: at the 1.5 ns, the angle of 3' terminal base of d(CC) shows an increase from  $\sim 20^\circ$  to  $\sim 100^\circ$ . This result is in agreement with the increase in the rmsd, indicating that the flipping of the base is the major contribution to the motion of the DNA during the simulation. These findings are also detected in the d(GA) and d(TA). The RMSDs of d(GC), d(GT), and d(TG) do not simply reach a plateau or stabilize but exhibit large-scale fluctuations due to the motions in the bases. This suggests the relative flexibility of these DNAs on the (8,4) CNT. Most intriguingly, following the initial adsorption the d(TT) exhibits a multistage process, whose rmsd is found to oscillate periodically between 0.3 and 3.8 Å, suggesting



**Figure 2.** Evolution of rmsd of 16 different ssDNA dimers in (8,4) (pink) and (6,6) (cyan) systems, compared to the initial structure. In order to enhance the graphical cleanliness, 100 ps running averages were performed on the rmsd curves.

that several fairly stable binding conformations are present. As for the rest of dimers, the rmsd values well behave in the last 5 ns of simulation (with the oscillation around 0.5 Å), thereamong, the d(AG), d(AT), d(CC), d(CT), and d(GG) seem quite stable with an average rmsd of 2.4, 2.8, 3.5, 2.9, and 2.1 Å, respectively. In order to clearly describe the rmsd changes of these dimers, several typical systems (d(GA), d(GT), d(GG)) are chosen for further analysis. The rmsd of d(GA) increases dramatically (from 1.54 to 3.84 Å) within the first 10 ns and then remains stable for the remaining simulation time. Structural analysis reveals that the adenine alternates from one side to the other side of the backbone relative to the initial state for which the bases are placed on the same side of the backbone. Such geometry transformation can minimize the steric hindrance between the bases and, therefore, is helpful for the bases to occupy more binding spaces on the surface of SWCNT.<sup>14,22</sup> In the case of d(GT), the rmsd values in the first 16 ns do not show noteworthy changes but experience a pronounced increase (from 1 to 3 Å) during 17–19 ns. This drastic increase in rmsd stems from the directional change of guanine relative to the axis of SWCNT (from  $\sim 16^\circ$  to  $\sim 90^\circ$  compared to initial structure), which results in the backbone more elongated. Subsequently, the rmsd drops down to  $\sim 1$  Å and the guanine goes back to the previous direction. A further discussion of the orientation change is given in section 4. In the following 20 ns simulation, the dimer switches between two conformational states in the complex: saddle configuration (valley around 1 Å) or stretched configuration (peak around 3 Å). The conformation transformation is mainly dominated by the flipping of the bases and the detailed conformational features are described in the clustering analysis of section 3.2, whereas d(GG) shows a stable rmsd with very small oscillations around  $\sim 2$  Å, indicating that it assumes a stabilized conformation when binding onto the SWCNT. The structure analysis shows that the furanose sugar in the 5' side flips out and is parallel to the CNT referenced to the initial structure in which the O4\* atom of the sugar group points radially inward to CNT, while slight change is observed in the base and the rest with the phosphodiester backbone, emphasizing relatively strong binding of the d(GG) to the CNT. In addition, we also investigated the rmsd variations of the compositional isomers with reverse sequences, as exemplified by the d(AG) and d(GA). It is noted that the rmsd behaviors of the two compositional isomers are significantly different: the rmsd of d(GA) undergoes a noticeable elevation and stabilizes at high values as illustrated above, but d(AG) has relatively small rmsd values ( $\sim 2.5$  Å). The bases of d(GA) have a relatively large shift compared to d(AG) with respect to each initial conformation, in which two bases are positioned on either side of the backbone and shaped like "S". Thus we speculate that the sequence order of the nucleotides also affects the dynamic binding of the ssDNA to CNT.

For the (6,6) hybridizations, the RMSDs of various dimers are also different. Among these dimers, the d(AA) and d(TG) have relatively large rmsd values (4.0 and 3.18 Å respectively). The d(GG) and d(TA) are much more stable during the simulations with an average rmsd of 2.3 and 3.0 Å, respectively. In contrast, the d(CT), d(TC), and d(TT) dimers seem unstable, whose rmsd values undergo large fluctuations (2.4, 2.0, and 2.5 Å, respectively). After comparison of the dynamics profiles for the (8,4) and (6,6) hybrids, we found that the nanotube chirality plays a minor role in modulating the dynamic structures of several noncovalent associated DNAs, as evidenced by their rmsd fluctuations. For instance, the trajectories of d(AG) have very similar rmsd

variation profiles in the two systems. They both pass through a jump (rmsd around 3 Å) in the beginning 10 ns and then gradually level off for the rest of the simulation time (with an average rmsd about 2.2 Å for the last 30 ns). By viewing the trajectories, it is found that the dynamic structures sampled in d(AG)-(6,6) trajectories with d(AG)-(8,4) show excellent agreement in dimer geometries at many time points. The most obvious is that both the 5'-terminal adenines of d(AG) shift to the other side of the backbone around 10 ns. Additionally, the rmsd's of d(AC), d(GC), and d(GT) also present similar tendency, and the visualization of the dynamics simulation demonstrates that most of the intermediate conformations are comparable in the (8,4) and (6,6) systems. From these analyses, a conclusion can be drawn that the conformations of the ssDNA dimers, such as AG, AC, GC, and GT are insensitive to the SWCNT chirality. While this is not the case for other dimers such as CC, GA, TA, and TT, among which the most notable is d(GA), whose rmsd undergoes slight oscillation within the last 30 ns of the simulations when bound onto the (8,4) CNT ( $3.72 \pm 0.36$  Å), but fluctuates largely at the first 20 ns in the (6,6) hybrid ( $2.26 \pm 1.09$  Å) and then stabilizes at  $\sim 1.5$  Å in the last 20 ns.

To sum up, most dimers can achieve the stable arrangement on the nanotube surface, indicating that the inherent one-dimensional (1D), cylindrical shape of the SWCNTs limits the conformational disorder of ssDNA upon the adsorption and plays an important templating role in forming the  $\pi$ - $\pi$  conjugation with the DNA base independent of the SWCNT chirality. However, several dimers (such as d(GT)-(8,4) and d(TT)-(8,4)) are more dynamic than others. Further analysis shows that the nucleotide compositions as well as the sequence order of the nucleotides affect the dynamic binding of the ssDNA to the CNT, and the structures of certain ssDNAs (AG, AC, GC, GT) are less dependent on the chirality of the tubes. During 40 ns simulation the O4\* atom of the sugar group prefers to orient toward the surface of the tube, suggesting that our initial structure model is reasonable and energetically favorable. However, it should be noted that ssDNA is a flexible polymer with many degrees of freedom and the potential energy may contain some local energy minima. In order to further probe the conformational space of ssDNA, to verify that the ssDNAs have indeed reached the "native" conformation under these simulation conditions, and to calculate the free energy landscape employing the principal component analysis (PCA), more extensive samplings are generated using the Replica exchange molecular dynamics (REMD) method. With this method a set of simulations were performed independently at different target temperatures, permitting random walks in the temperature space and escape from the local energy traps, thus allowing more conformations to be sampled and overcoming the free energy barrier.

**3.2. REMD analysis.** Two typical DNA dimers, the homooligonucleotide d(AA) and the heterooligonucleotide d(TG), are chosen for this study because they have a relatively high binding energy and the lowest binding energy, respectively (see section 5.2), and besides TG is the most widely studied sequence. We intend to check whether the standard MD simulation sampled the sufficient conformations compared to REMD. An explicit solvent model has been substantiated to represent the solvation effect, and the temperature of the 20 replicas ranged from 298 to 404 K. In the 20 ns REMD simulation, the adsorbed nucleotides slide and rotate freely along the CNT. Numerous transitions of the binding states occur and several stable states are visited multiple times, indicating that the bound state ensembles

are well sampled in our simulations. The most probable oligonucleotide configurations are the loops having an extended backbone. The saddle-shaped configurations are relatively few.

Then we processed the 298 K REMD simulation as the most preferable trajectory to be analyzed in depth. The PCA was used to obtain the dimer free-energy landscape, and two-dimensional profiles were constructed from the simulations using the first two principal components as the reaction coordinates. In the Cartesian PCA, the motions along the first two eigenvectors  $v_1$ ,  $v_2$  contain 73.3% and 74% of the fluctuations of the  $d(\text{AA})$  and  $d(\text{TG})$ , respectively. This means that the two-dimensional PCA space is a good approximation to express the free energy landscape. The free energy surface along these vectors is given by

$$\Delta G(v_1, v_2) = -k_B T [\ln \rho(v_1, v_2)]$$

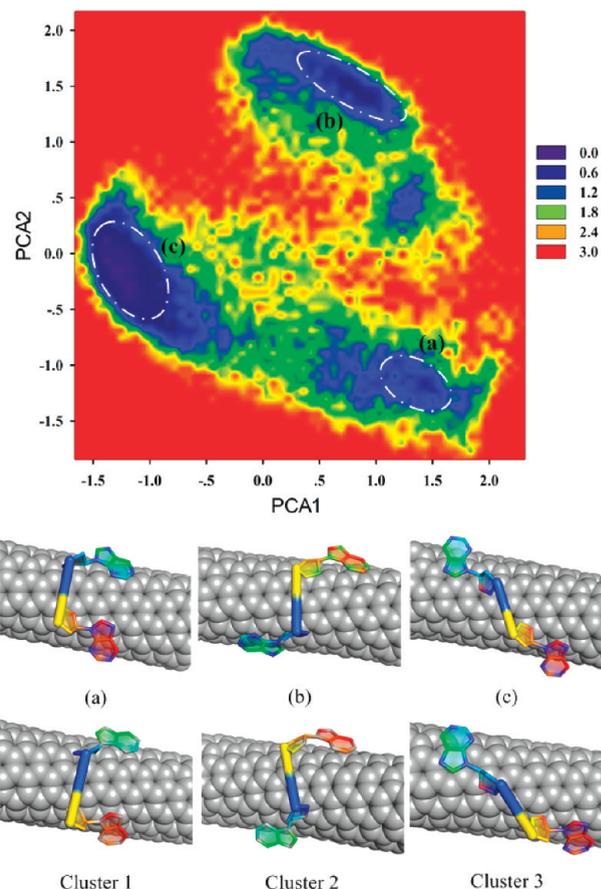
Here  $\rho(v_1, v_2)$  is the density at position  $(v_1, v_2)$  in subspace. We set the highest density  $\rho$  equal to 1.0, to ensure the  $\Delta G(v_1, v_2) = 0$  for the lowest free energy minimum.

For the  $d(\text{AA})$ , three free-energy basins are found that correspond to region (a) at (1.5, -1.2), region (b) at (1.0, 1.5), and region (c) at (-1.2, -0.2) (Figure 3). The corresponding representative structures are shown below. Of all these conformations, (c) wraps around the tube more extendedly, and this provides the maximum vdW interaction between oligonucleotide to the tube. (b) and (c) are the energetically more favorable conformations on the nanotube. We further compared these three kinds of configurations with the representative structure derived from the clustering analysis (as discussed in section 3.3) of conventional MD, and the results suggest REMD samples and conventional MD bearing close resemblance (as shown in Figure 3). Then we compared the snapshots of ssDNA generated by the REMD method to those from the conventional MD and found that the conformations resulting from the REMD are also sampled in the previous conventional MD simulation. Thus, for the conventional MD, there are enough statistics for the latter analysis. Similar conclusions can be drawn from the analysis of the  $d(\text{GT})$ , for which the results of the REMD are in accord with the conventional MD. We show the free energy landscape and the representative conformation from the energy landscape and clustering analysis in Figure S2 in the Supporting Information.

Because of the enormous reduction in the dimensionality when a free-energy profile is constructed from the REMD simulation, a two-dimensional map presents an incomplete picture of the binding scenario. Alternatively, clustering analyses can potentially provide complementary information to represent this multidimensional process.

**3.3. Clustering Analysis.** The cluster analysis based on the average-linkage algorithm<sup>45</sup> was applied to the full set of sampled configurations for each of the 32 hybrids for the purpose of: (i) identifying the transition of DNA structure along the bound process; (ii) comparing and analyzing the diversity and coherence of the geometries for each dinucleotide, and (iii) investigating the impact of the nanotube chirality on the different combinations of two nucleotides. Table S1 (Supporting Information) lists the major clusters in the order of their probability of existence in the whole trajectory, which also includes the values of the clustering metrics. For all clusters, the nearest-neighbor distances are distributed between 2.44 and 3.30 Å, mostly around 2.80 Å.

In the (8,4) hybrid systems, most dimers (10/16, CA, CC, CG, GA, GC, GG, GT, TA, TC, TT) are dominated by two clusters



**Figure 3.** Free-energy landscapes from  $d(\text{AA})$ -(8,4) hybrid at 298 K with typically sampled conformations (a–c). Unit of the potential of mean force is kcal/mol. Axes PCA1 and PCA2 are the first and second eigenvectors computed from the corresponding ensemble. The representative structures of the cluster 1–3 of the  $d(\text{AA})$  obtained from the conventional MD are also shown on the figure.

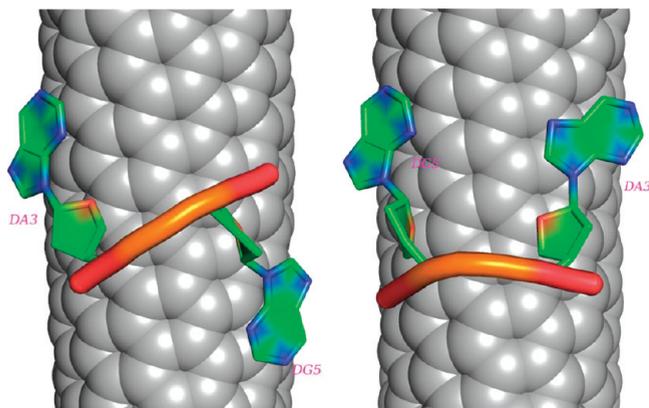
except the AA, AG, AT, CT (three clusters), AC, TG (four clusters). Likewise, in the (6,6) hybrid systems, the majority of DNAs (9/16) are divided into two clusters (AA, AC, CC, CT, GA, GC, GG, GT, and TA), and the remains of the AG, AT, CA, CG, TC, TG, and TT consist of three clusters. In all systems, the average distance to the centroid varies between 0.62 and 1.39 Å. For each identified cluster, the central member which has the smallest rmsd to all other structures within the cluster is chosen as the cluster representative structure. For comparing purposes, this appropriate conformation should be accessible to all structures.

All 79 representative structures are inspected individually (40 for (8,4) and 39 for (6,6) complexes), and in order to characterize the conformations of the nucleic acid biopolymers, seven parameters ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\xi$ ,  $\chi$ ) are calculated (Table S2, Supporting Information). The definition of these dihedral angles is given by Dickerson et al.<sup>46</sup> For the 16 kinds of dinucleotides in the (8,4) hybrid systems, the structure parameters show distinct backbone configurations and orientations of the nucleoside side chains among the subclusters of each dimer. For instance, significant difference in the backbone geometry is observed between two subclusters of the  $d(\text{GA})$  as shown in Table 1. The main difference is in torsions  $\xi$ ,  $\alpha$ ,  $\beta$ , and  $\gamma$ , confirming that the transformation in backbone allows the dimer to swing into a dissimilar stacking conformation. Visual inspection of the representative

**Table 1. Backbone Torsion Angles<sup>a</sup> and Glycosyl Angle  $\chi$  for the d(GA) in (8,4) Hybrid**

representative structure	$\chi$ (S')	$\delta$ (S')	$\epsilon$	$\xi$	$\alpha$	$\beta$	$\gamma$	$\delta$ (3')	$\chi$ (3')
cluster 1	38.28	126.1	-70.6	76.3	64.1	99.7	-157.5	98.5	134.9
cluster 2	39.5	152.2	-85.3	140.3	-84.8	173.2	56.3	101.2	18.02

<sup>a</sup> Backbone torsion angles are  $O5'-\delta(S')-\epsilon-\xi-\alpha-\beta-\gamma-\delta(3')-O3'$ .



**Figure 4.** The representative structures from cluster 1 (left) and cluster 2 (right) of the d(GA) in (8,4) systems. The configuration of cluster 1 is more stretched than that of cluster 2. Water molecules are not displayed for clarity.

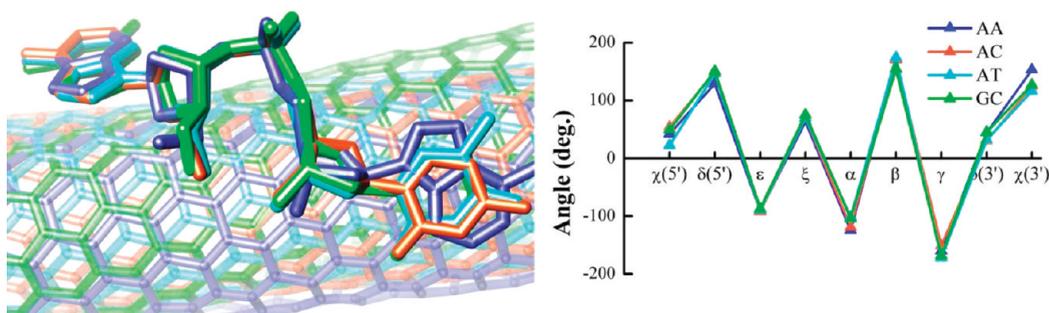
structures for these two subclusters indicates that the chain of cluster 1 is more stretched (Figure 4). In addition, the dihedral angle  $\chi$  which describes the orientation of the base with respect to the sugars is also different for guanine and adenine in these clusters: in cluster 1, the two bases are apart away from each other and located in the opposed side of the chain of backbone. All four angles reflect the different conformations of the two clusters. Moreover, inspection of the representative structures for other dimers implies that the preferred conformations corresponding to large clusters are generally stretched and the adjacent bases tend to be away from each other. For example, the cluster 1 of d(GA) has many more members than cluster 2 (3284 and 716, respectively), in which the DNA is more elongated as stated above. Moreover, the subclusters of GT, CC, AG, GA, etc., are also consistent with this observation. This result is in excellent agreement with our original conclusion drawn from the REMD simulations.

Even though the combinations of the dinucleotides are distinct, several dimers share a considerably similar conformational pattern. For example, the representative structures of cluster 3 of AA and cluster 2 of AC, AT, GC, TC and cluster 1 of CG, CT have very close structure parameters, as shown in Figure 5. Moreover, the seven dimers are well superimposed on the d(AA) with an rmsd of 0.00, 0.98, 1.05, 1.04, 1.45, 1.05, 1.5 Å, respectively. For clarity, here, only the superimposed representative conformations and the torsion angles of d(AA), d(AC), d(AT), and d(GC) are shown in Figure 5. Further analysis suggests that for each of the groups including cluster 2 of CG, GA, cluster 1 of AA, TC, as well as cluster 3 of CT, cluster 2 of CC, TT, and cluster 1 of CA, TA, etc., the representative DNA assumes similar conformations (data not shown). These findings suggest that the morphology of ssDNA is not solely determined by the base sequence but may also be influenced by the surrounding solvent and SWCNT. Since the backbone of ssDNA is quite flexible,<sup>47</sup>

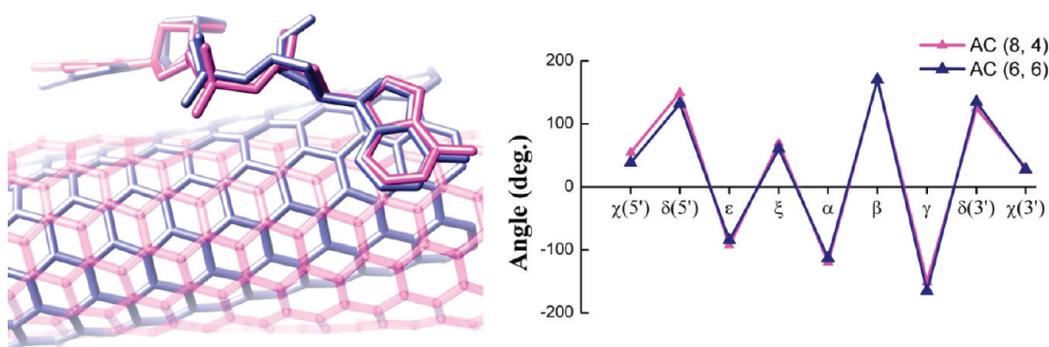
allowing it to stretch to accommodate the binding of the bases onto the relatively rigid nanotube: the bases orient approximately parallel to hexagons on the nanotube surface and twist to compete for the binding space. The transformations of the backbone and the base configuration proceeding via a rearrangement of the seven torsional angles also lead the dimers to find suitable low-energy conformations. For example, the principal binding mode of d(GA) shows that the oligonucleotide chain straddles the tube with a more extended geometry and the two bases spread out at the opposite sides of the chain. This geometry will afford good stacking of the bases to the nanotube surface by increasing the contact area and weakening the solvation free energy of dimers submerged in water, due to the reduced hydrophobic area of the SWCNT surface exposed to the waters. Taking all of these factors into account, it appears reasonable to suggest that the general ssDNA can attach to a SWCNT in a similar structure.

With regard to the (6,6) hybrid systems, similar findings are observed: the subclusters of each dimer yields characteristic backbone and base configurations as described by the torsion angles (Table S2b, Supporting Information). The geometries with the stretched chains are prevalent in most of the binding modes. Besides, the adjacent bases depart away from each other due to the competition for binding space on the SWCNT surface. It is very similar to the aforementioned conditions that such an arrangement of conformation will enhance favorable  $\pi$ -stacking interactions and reduce the solvation free energy. Semblable conformations are also observed among different ssDNA dimers (nine groups), such as cluster 3 of CG, cluster 2 of AG, TC, cluster 1 of AC, GC, cluster 3 of AT, TC, cluster 2 of GT, TG, as well as the cluster 1 of AA, AG, CA, CG, TA, etc. For these groups, the best-fit superposition of the DNA structures yields atomic rms difference of 0.74–1.05 Å. Meanwhile, minor variations in the geometry are observed as indicated by the torsion angle parameters (5.93–14°). These results strongly suggest that the ssDNA dimers are inclined to form an optimal bound geometry related to the nanotube surface curvature and solvation effects.

Finally, it is instructive to compare the structures of the (6,6) hybridizations with those of the (8,4) systems. Interestingly, we find that d(AC) in both systems presents uniform conformational features (as shown in Figure 6). The dimers attach and spread on the surface of CNT and the two bases separate away from each other (with a distance of 13 Å). The rmsd for the optimal superposition of the DNAs is 0.58 Å. The nine dihedral angles are in good agreement with the relatively minor discrepancies between them (2–17°). The same situation is also observed in the AT, CG, CT, GC, GG, TA, TC, TG, and TT with rmsd values below 2.0 Å. From this finding, which is in agreement with the conclusions derived from section 3.1, we confirm that the structure of the dinucleotides is not sensitive to the chirality of the SWCNT, and the general ssDNA dimers are thus expected to wrap the SWCNT in a similar manner. However, it is important to note that these structures differ in the orientation of the bases with respect to the axis of the CNT. Such



**Figure 5.** (Left) The representative structures from cluster 2 of d(AC) (orange), cluster 2 of d(CG) (green), and cluster 1 of d(AT) (cyan), superimposed onto cluster 3 of d(AA) (blue) by the average-linkage clustering algorithm. The corresponding (8,4) CNT of each dinucleotide is also labeled with the same color. Adsorbed bases in DNA lie flat on the SWCNT surface and can interact via vdW force. The hydrogen atoms and the water molecules are not shown for clarity. The images on the right show the structure parameters for corresponding dinucleotides.



**Figure 6.** A superposition of the d(AC) in (8,4) (magenta) and (6,6) (blue) SWCNTs. Adsorbed bases in DNA lie flat on the SWCNT surface and can interact via vdW force. Water molecules have been removed for clarity. The right part depicts the structure parameters for d(AC) in both systems.

directional difference of the bases may be the result of the  $\pi-\pi$  interaction between the base ring and the hexagons of the nanotube that the DNA will adopt specific orientation to stack on the hexagons of the CNT surface depending on how the carbon hexagons arrange relative to the tube axis. Consequently, our following discussions will concentrate on the orientation of the DNA base on the SWCNT. Of note, some oligonucleotides can wrap about these two chirality tubes in disparate fashions, such as the d(GG) and d(TG).

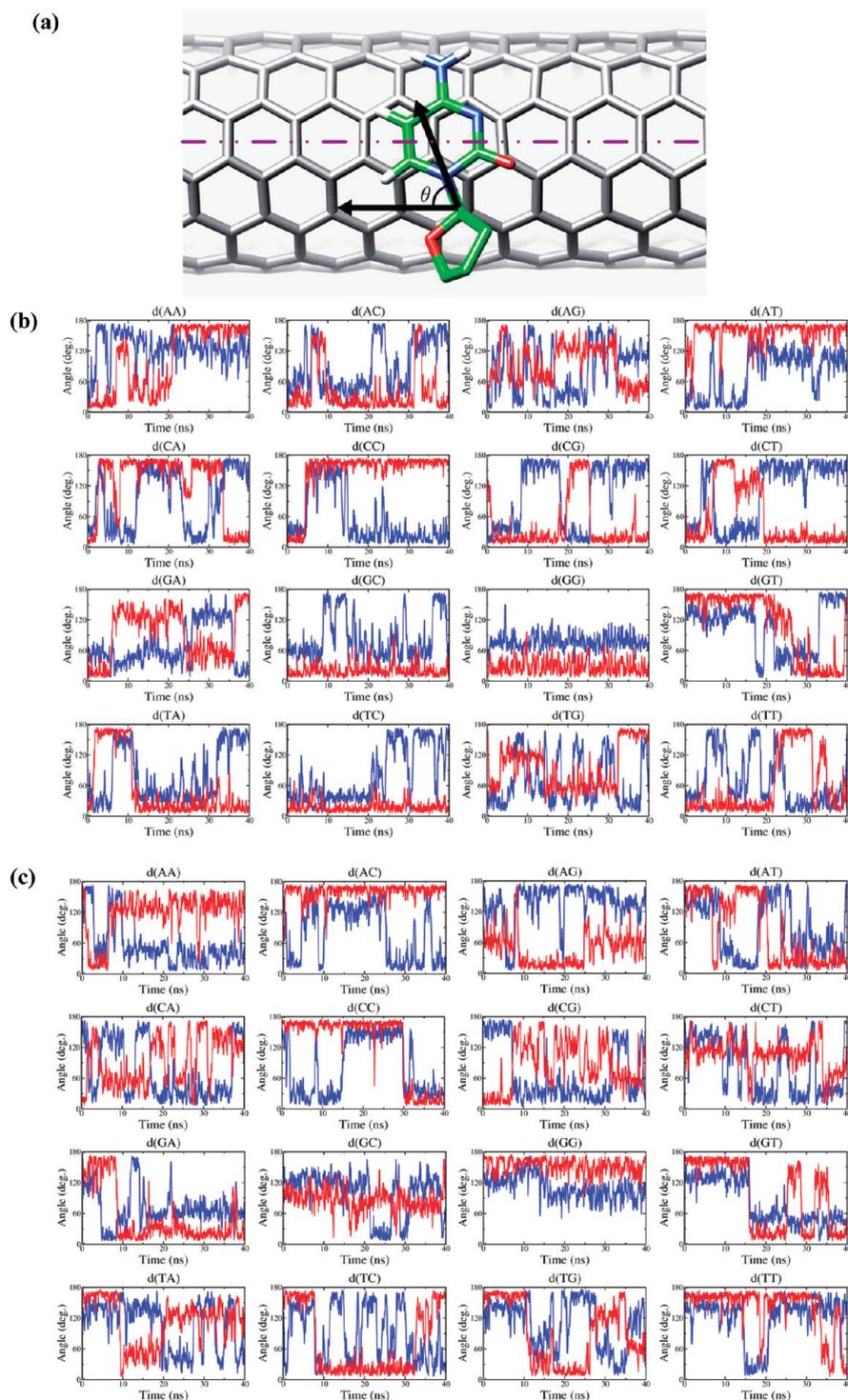
#### 4. ORIENTATION OF THE NUCLEOTIDE BASES ON THE SWCNT

Because the main purpose of this study is to explore the interaction between ssDNA dimers and SWCNT, it is inevitable to identify the direction of the nucleic acid base relative to the CNT. To check this, we quantify the tendency of the bases to bind onto the CNT by tracking a base orientation angle  $\theta$  which is computed as the angle between the glycosidic bond and the central axis of CNT,<sup>22</sup> as illustrated in Figure 7a.

Panels b and c of Figure 7 depict how the base orientation angle for each base in the ssDNA dimers bound on the SWCNT varies during the simulation. Also, 100 ps running average was performed on all angle plots in order to enhance the graphical clearness. The observed behavior implies that for most systems, the bases on the CNT have a preferred orientation; namely, the bound bases strongly prefer orientations that align themselves to a given direction along the nanotube axis. For instance, in the

(8,4) system, the 5' terminal cytosine residue of d(CC) dimers favors angle  $\theta$  from 150 to 170° and occupies 83.5% in the whole simulation. Meanwhile, the glycosidic bond of C base at 3' terminal aligns relative to the nanotube axis with an angle of 10–20° in the first 5 ns and during the 15–40 ns. Other bases also display orientation preference, reflected by the curves showing characteristics of the angle  $\theta$  in Figure 7. A wide angle  $\theta$  distribution is observed with respect to the adenine (10–70° and 100–180°). Interestingly, the 5' terminal base of AA, AT, CC, CG, CT, GC, GG, GT, TA, TC, and the 3' terminal base of AA, AT, CC, CG, CT, GG are found to have a propensity to rotate within a constant range of angles with the CNT axis. Depending on the variations of angle  $\theta$ , many more preferred nucleic acid base orientations can be identified. Besides, the same bases exhibit distinct preferred adsorption directions in different sequence dinucleotides; as for the case of the d(GA), d(GC), d(GG), d(GT) in the (8,4) system, the guanine shows distinct preferred adsorption directions (~50°, ~60°, ~75°, ~130°, respectively). On the contrary, the bases of some hybrids show no obvious bias for the orientation angle  $\theta$ , such as the adenine of the d(AG) in (8,4) CNT systems, in which  $\theta$  does not dwell in any intermediate range for an appreciable time as shown in Figure 7b.

Beyond that, the bases on the (6,6) CNT also have these characters, but the preference angles of individual bases are different in these two nanotubes. The large discrepancy occurs in the case of d(AA) of two systems: in the (8,4) system d(AA) is most mobile where the angles of both bases significantly oscillate



**Figure 7.** The DNA base on the CNT has a preferred orientation. (a) Orientation is measured by the angle ( $\theta$ ) between the glycosidic bond and the SWCNT axis. The dash line specifies the long axis of the SWCNT. The evolutions of orientation angle for bases of every dimer in (b) (8,4) systems and (c) (6,6) systems versus time during the simulations (blue, means the 5' terminal base; red, the 3' terminal base).

in a broader range, while for the d(AA) at (6,6) CNT, the angles  $\theta$  of two bases are inclinable at angles of  $\sim 40^\circ$  and  $\sim 125^\circ$ , respectively. Similar results are also observed in AC, AG, GA, CC, etc. We then infer that the chirality of nanotube facilitates the base deformations along preferred directions depending on the chirality angle of CNT.

Essentially, the change of angle  $\theta$  is related to the conformation motion of the dimer. We chose d(AC) from the (8,4) system as an example to illustrate. During the 0–5.8 ns simulation, the  $\theta$  of the cytosine swings between 0 and  $40^\circ$  and then subjects to a sharp rise toward  $100\text{--}160^\circ$ . Visual inspection of the DNA translocation events with VMD reveals that in the first 5.8 ns this

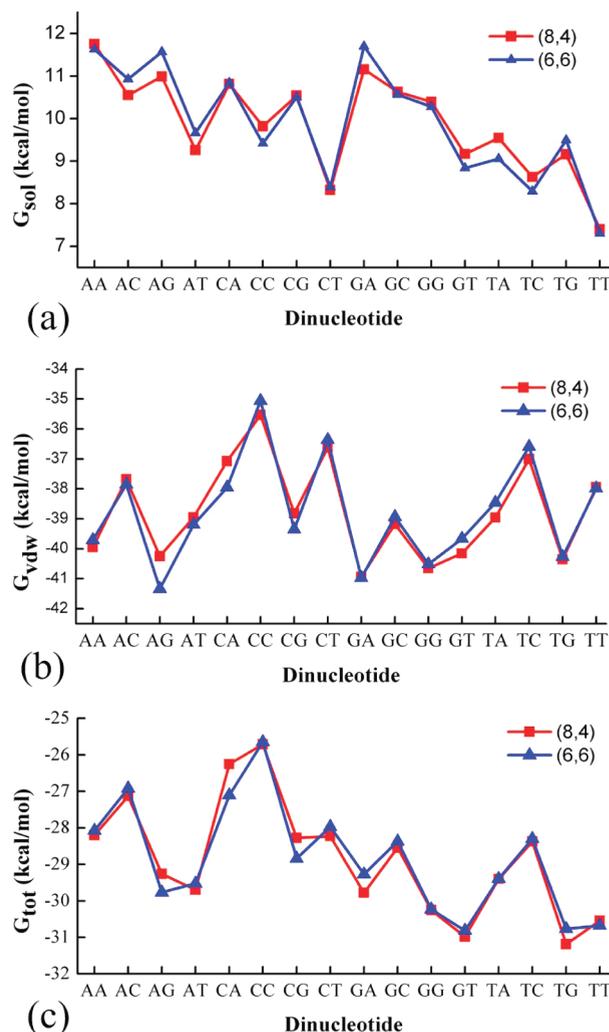
motion adjusts the conformation of ssDNA from the “U” (both bases at the same side of the backbone) to “S” shape with the C base rotating to the other side of the chain. When angle  $\theta$  becomes  $<40^\circ$ , the DNA returns back to the “U” conformation similarly to that of the original phase. For the A base, the most preferred angle is  $40\text{--}80^\circ$ . These findings are consistent with the clustering result that most snapshots adopt the “U” conformation in the trajectory of d(AC) with (8,4) hybrid. We also notice that the angle  $\theta$  of several adjacent bases in the same dimer exhibits the same change tendency during the simulation, such as d-(GG)-(8,4) where both bases fluctuate within a constant range ( $\sim 25^\circ$  and  $\sim 70^\circ$ , respectively). The visual observation displays that d(GG) has a relatively stable structure which agrees well with the rmsd result that its fluctuation is small as illustrated in section 3.1. From this analysis, we find that the base orientation has a good correlation with the dimer’s conformation, that is, the translocation of the bases in one of the orientations is correlated well to the conformational variation of the DNA.

Briefly, we find that the bases in the ssDNA–CNT complexes prefer to have a definite orientation relative to the nanotube axis, although the global orientation preference of a DNA base depends on its own properties, adjacent bases and the chirality angle of the CNT. The unique orientations of the DNA bases on CNT might originate from the van der Waals interaction between the bases and the curved CNT walls, which would tend to maximize the base CNT contact region. The sugar and phosphate groups in the DNA backbone might also contribute to this interaction by providing further geometrical constraints.

## 5. ENERGY ANALYSIS

DNA–CNT is usually assembled and employed in aqueous solution, and it is expected that DNA–CNT interactions include a complex interplay of the molecular vacuum energy and solvent-mediated effects. Here, the MM-GBSA analysis allows us to separate the total free energy of binding into the electrostatic, van der Waals, and solute–solvent interactions, and thereby provides additional information about the DNA–CNT adhesion strength. Since the CNT in our model carries no charge, the contribution of the electrostatic energy is neglected. Hence, we focus on the two components, i.e., the contributions from the solvation free energy and the van der Waals (vdW) interactions.

**5.1. Solvation Free Energy.** As it has been reported that ssDNA can effectively solubilize highly hydrophobic CNTs,<sup>14</sup> solvent interaction is considered to play an important role in the association of DNA and CNT. The ensemble average solvation free energies for the (8,4) and (6,6) hybrids are given in Tables S3 (Supporting Information), along with the averages computed for each of the principle clusters identified in the former analysis. In all cases, the binding is accompanied by the reduction of SASA due to the burial of large portions of nanotube surface through the stacking of DNA bases,<sup>24</sup> and thereby, the solvation free energies ( $G_{\text{sol}}$ ) provide a comparatively large, negative contribution to the binding free energy. The data indicate that the subcluster with the slightly larger solvation free energy is expected to favor the stretched structures, which reduces the expose of the highly hydrophobic CNT but increases the delta solvation energy of the complex. For example, the  $G_{\text{sol}}$  values of clusters 1 and 2 of d(GA) in the (8,4) system are 11.39 and 10.93 kcal/mol, respectively, with corresponding representative conformations shown in Figure 4. From this figure, it is clearly seen that the representative structure of cluster 1 is more “extended”



**Figure 8.** The contributions of solvation effects (a) and van der Waals interactions (b), as well as total binding free energy (c) of dinucleotides with (8,4) and (6,6) CNTs.

than that of cluster 2. Such an arrangement is propitious to increase the contact surface between ssDNA and CNT (from  $439 \text{ \AA}^2$  (cluster 2) to  $450 \text{ \AA}^2$  (cluster1)) and thus reduces the SASA of the nanotube surface. This feature is also well reflected in d(CC), for which, cluster 1 has the larger solvation free energy (10.64 kcal/mol) than cluster 2 (9.00 kcal/mol). Investigations of their representative structures show that cluster 1 is indeed more stretched than cluster 2. In order to directly account for their stretching degree, we estimated the average distance between two bases. The average distances between two bases for the two clusters are 11.2 and 6.9  $\text{\AA}$ , following the trend of cluster 1 > cluster 2 and are positively correlated with the energy tendency. It is thus germane to consider that the ssDNA adhesion on the SWCNT will cause significant deformation (tending to stretch) to decrease the solvation free energy of CNT and increase the area of contact. Overall, the desolvation penalty is the largest for d(AA) with a loss of 11.75 kcal/mol and the smallest for d(TT) where 7.40 kcal/mol is lost. Evidently, this behavior can be explained by the larger molecular volumes of the purine residues compared with the pyrimidine ones. Interestingly, analysis shows that the dinucleotides including the thymine usually have lower

solvation energies (shown in Figure 8a). For the sake of explaining this phenomena, we performed the energy decomposition, and found that the solvation energy of the thymine is smallest (in the range of 3.38–5.13 kcal/mol) among the four bases. It thus demonstrates that the introduction of thymine into the ssDNA would increase the water solubility of the DNA–CNT hybrids. Further comparison reveals that there is little variation between (8,4) and (6,6) systems except for the d(AG), d(GA), and d(TA) (0.57, 0.54, –0.48 kcal/mol, respectively), possibly due to small different curvatures of the CNT surface. For d(AG) and d(GA), conformational comparison reveals that they are more stretched in the (6,6) systems, and their corresponding solvation free energies are higher than those in (8,4) systems.

**5.2. vdW Interaction Energy.** Since the vdW interaction is the dominant interaction between ssDNA and nanotubes,<sup>24</sup> we have carried out classical force field calculations to calculate the vdW energy between the nanotube and dinucleotides. In Table S3 (Supporting Information) we report, for the (8,4) and (6,6) hybrids, the ensemble average of DNA–CNT interaction energies and corresponding total binding energies. All bases of dinucleotides stack well on the surface of the SWCNT, stabilizing the complexes with the vdW energy ranging from –34.01 to –41.84 kcal/mol. The discrepancies on the vdW energy among subclusters reveal that the hybridization interaction is related to the conformation of the DNA. The “extending” dinucleotides and two far-apart bases lead to more spaces for the DNA binding onto the surface of the SWCNT, therefore aggrandizing their vdW interaction. Additionally, the dinucleotide with two purines exhibits more favorable interactions with CNT. The vdW energies also show that the dimers containing a purine tend to have lower vdW interaction energies, implying that the purine has greater preference for stacking with CNT (Figure 8b).

To delineate the underlying mechanism responsible for the observed result, we calculated the vdW energy between the DNA bases and the surface of the CNT using MM-GBSA decomposition. As a result of comparison, relatively smaller vdW values are obtained for pyrimidine. The strength of this interaction varies among the four bases and follows the trend  $G > A > T > C$  ( $\sim -10.2$ ,  $\sim -9.7$ ,  $\sim -9.0$ ,  $\sim -8.3$  kcal/mol, respectively), whose values compare well with the reported values of graphite–base interaction using molecular dynamics and ab initio calculations.<sup>48,49</sup> This trend is understandable from the geometric considerations that the purines, G and A, which contain two aromatic rings have a stronger interaction with CNT than pyrimidines, C and T, which contain only a single ring. The decomposition analysis also confirms that the base ring contributes greater than the backbone of DNA to the binding affinity in the sidewall of SWCNTs (data not shown).

After all, the hybrid's binding free energy includes contributions from both the base–SWCNT stacking and the solvation energy (shown in Figure 8c). Among the 16 dinucleotides in the (8,4) hybrid system, d(TG) has a minimum binding energy (31.19 kcal/mol) with CNT. In the (6,6) system the binding energy of d(GT) (–30.82 kcal/mol) is the lowest of other dinucleotides, for which vdW and solvation free energy contribute favorably to the total free energy. These findings explain the experiment<sup>14,50</sup> that the oligonucleotides made of thymine bases are more effective in dispersing the SWCNT in aqueous solution as compared to poly (A) and poly (C). On the whole, we notice that the d(CC), in both the (8,4) and (6,6) systems, has the largest binding free energy as show in Figure 8c. The AG, CA, CG, and TG exhibit relatively larger difference in binding free

energy between (8,4) and (6,6) systems (0.50, 0.85, 0.57, –0.42 kcal/mol, respectively). Analysis of these free energy profiles and the variability also suggests that the total DNA–CNT binding free energy is sequence specific and not merely a sum of individual base–SWCNT binding free energies.

## 6. SUMMARY

While a number of experimental studies have shown that certain short ssDNA sequences can recognize specific SWCNTs, thus allowing the separation of low resolution metal/semiconductor CNTs from a mixture by ion exchange chromatography (IEX),<sup>15</sup> relatively little else is known about the details of the process on an atomic level.<sup>14,15,21,51</sup> To better explore this issue, we have conducted a series of atomistic MD simulations to explore the structural and energetic properties of 16 dinucleotides with different combinations associated with two types of nanotubes (chiral and armchair) in water. In all systems, stable hybrids were found to form through noncovalent adsorption of DNA onto the nanotube surface, with base planes stacking onto the CNT surface, leaving the backbone exposed to the solvent. This arrangement maximizes the favorable  $\pi$ -stacking interactions and is prevalent in all of the binding modes. The rmsd profiles show that several dimers are more dynamic than others and the nucleotide compositions as well as the sequence order of the nucleotide affect the dynamic binding of the ssDNA on the CNT. Certain ssDNAs (AG, AC, GC, GT) are less dependent on the chirality of the tubes. In order to further verify that the ssDNAs have indeed reached the “native” conformation under these simulation conditions, we conducted the REMD simulations for two typical dinucleotides and all of the results are in accord with our standard MD simulation results. In the subsequent clustering analyses, we find that some dinucleotides exhibit several distinct geometries with the stretched configuration being the most probable and which is in excellent agreement with the REMD results. The conformations of oligonucleotides are found to be dependent on the identity of the nucleotide, on the solvation effects, and also, to a lesser extent, on the nanotube chirality.<sup>44</sup> More strikingly, in addition to favoring surface-stacked conformations, each nucleotide base has a preference for a certain range of orientations along the CNT axis. This behavior was also observed through the ab initio time dependent density functional theory and via optical spectroscopy measurements,<sup>52</sup> and replica exchange molecular dynamics of the DNA–CNT hybrid.<sup>22</sup> The orientational preference of the DNA bases on the CNT might originate from the physisorption of base plan on the lattices of CNT through  $\pi$ – $\pi$  stacking interaction and possibly depends on the base character and chiral nature of the nanotubes.

To investigate the origin of the binding of DNA with CNT, we have employed the MM-GBSA method to compute the ssDNA–CNT binding free energy, which shows that the hybridization is driven by the attractive noncovalent interactions, primarily through the stacking of the DNA bases with the nanotube surface. Solvation effects play a relatively minor role. Nonetheless, the binding free energy is rugged which reveals that the binding free energy is sequence-dependent. Moreover, the dinucleotides are composed of two purines well stacked onto the nanotube surface, which provides a richer stacking contact with the CNT, resulting in relatively lower vdW interaction energy. Conversely, d(TT) exhibits the lowest solvation energies when compared with other dimers. It is not surprising that the d(TT)

has small SASA and tends to be the most efficient in solubilizing the nanotubes during the comparisons with other dimers. Further analysis implies that thymine can efficiently increase the solubility of dinucleotide–CNT hybrids and reduce their total free energy.

In conclusion, we have explored the characteristics of two-base ssDNA hybrids with metal/semiconductor SWCNTs. This study provides an important insight for understanding of these hybrids, though there are still many questions, like the sequence-dependent effects including the solvation patterns or intra-ssDNA nucleotide–nucleotide interactions,<sup>21</sup> remain to be explored. It is possible that longer specific polynucleotide chains would prefer to adopt high order structures.<sup>44</sup>

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Detailed REMD simulation, clustering procedures, MM-GBSA method, and additional figures and tables as referred to in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [yh\\_wang@nwsuaf.edu.cn](mailto:yh_wang@nwsuaf.edu.cn).

## ■ ACKNOWLEDGMENT

We are grateful to Professor Ling Yang for access of Sybyl software. This work is supported by high-performance computing platform of Northwest A & F University and is financially supported by the National Natural Science Foundation of China (Grant No. 10801025) and also the Fund of Northwest A & F University.

## ■ REFERENCES

- (1) Hu, J.; Shi, J.; Li, S.; Qin, Y.; Guo, Z.-X.; Song, Y.; Zhu, D. *Chem. Phys. Lett.* **2005**, *401*, 352–356.
- (2) Khabashesku, V. N.; Margrave, J. L.; Barrera, E. V. *Diamond Relat. Mater.* **2005**, *14*, 859–866.
- (3) Lin, Y.; Taylor, S.; Li, H.; Fernando, K. A. S.; Qu, L.; Wang, W.; Gu, L.; Zhou, B.; Sun, Y.-P. *J. Mater. Chem.* **2004**, *14*, 527–541.
- (4) Singh, R.; Pantarotto, D.; McCarthy, D.; Chaloin, O.; Hoebeke, J.; Partidos, C. D.; Briand, J.-P.; Prato, M.; Bianco, A.; Kostarelos, K. *J. Am. Chem. Soc.* **2005**, *127*, 4388–4396.
- (5) Zhang, L.; Kiny, V. U.; Peng, H.; Zhu, J.; Lobo, R. F. M.; Margrave, J. L.; Khabashesku, V. N. *Chem. Mater.* **2004**, *16*, 2055–2061.
- (6) Dong, X.; Fu, D.; Xu, Y.; Wei, J.; Shi, Y.; Chen, P.; Li, L.-J. *J. Phys. Chem. C* **2008**, *112*, 9891–9895.
- (7) Heller, D. A.; Jeng, E. S.; Yeung, T. K.; Martinez, B. M.; Moll, A. E.; Gastala, J. B.; Strano, M. S. *Science* **2006**, *311*, 508–511.
- (8) Chen, C. L.; Yang, C. F.; Agarwal, V.; Kim, T.; Sonkusale, S.; Busnaina, A.; Chen, M.; Dokmeci, M. R. *Nanotechnology* **2010**, *21*, 095504.
- (9) Lu, G.; Maragakis, P.; Kaxiras, E. *Nano Lett.* **2005**, *5*, 897–900.
- (10) Cheung, W.; Pontoriero, F.; Taratula, O.; Chen, A. M.; He, H. X. *Adv. Drug Delivery Rev.* **2010**, *62*, 633–649.
- (11) Star, A.; Tu, E.; Niemann, J.; Gabriel, J.-C. P.; Joiner, C. S.; Valcke, C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 921–926.
- (12) Huang, X.; McLean, R. S.; Zheng, M. *Anal. Chem.* **2005**, *77*, 6225–6228.
- (13) Tu, X. M.; Manohar, S.; Jagota, A.; Zheng, M. *Nature* **2009**, *460*, 250–253.
- (14) Zheng, M.; Jagota, A.; Semke, E. D.; Diner, B. A.; McLean, R. S.; Lustig, S. R.; Richardson, R. E.; Tassi, N. G. *Nat. Mater.* **2003**, *2*, 338–342.
- (15) Zheng, M.; Jagota, A.; Strano, M. S.; Santos, A. P.; Barone, P.; Chou, S. G.; Diner, B. A.; Dresselhaus, M. S.; Mclean, R. S.; Onoa, G. B.; et al. *Science* **2003**, *302*, 1545–1548.
- (16) Gigliotti, B.; Sakizzie, B.; Bethune, D. S.; Shelby, R. M.; Cha, J. N. *Nano Lett.* **2006**, *6*, 159–164.
- (17) Staii, C.; Johnson, A. T.; Chen, M.; Gelperin, A. *Nano Lett.* **2005**, *5*, 1774–1778.
- (18) Takahashi, H.; Numao, S.; Bandow, S.; Iijima, S. *Chem. Phys. Lett.* **2006**, *418*, 535–539.
- (19) Zhao, X.; Johnson, J. K. *J. Am. Chem. Soc.* **2007**, *129*, 10438–10445.
- (20) Gao, H.; Kong, Y.; Cui, D.; Ozkan, C. S. *Nano Lett.* **2003**, *3*, 471–473.
- (21) Johnson, R. R.; Johnson, A. T. C.; Klein, M. L. *Nano Lett.* **2007**, *8*, 69–75.
- (22) Johnson, R. R.; Kohlmeyer, A.; Johnson, A. T. C.; Klein, M. L. *Nano Lett.* **2009**, *9*, 537–541.
- (23) Gowtham, S.; Scheicher, R. H.; Pandey, R.; Karna, S. P.; Ahuja, R. *Nanotechnology* **2008**, *19*, 125701.
- (24) Johnson, R. R.; Johnson, A. T.; Klein, M. L. *Small* **2010**, *6*, 31–34.
- (25) Rajarajeswari, M.; Iyakutti, K.; Kawazoe, Y. *J. Mol. Model.* **2011**, *1–8*.
- (26) Frey, J. T.; Doren, D. J. *TubeGen 3.3*; University of Delaware: Newark, DE, 2005.
- (27) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.
- (28) Hummer, G.; Rasaiah, J. C.; Noworyta, J. P. *Nature* **2001**, *414*, 188–190.
- (29) Jenness, G. R.; Jordan, K. D. *J. Phys. Chem. C* **2009**, *113*, 10242–10248.
- (30) Subramanian, V.; Balamurugan, K.; Gopalakrishnan, R.; Raman, S. S. *J. Phys. Chem. B* **2010**, *114*, 14048–14058.
- (31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (32) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (33) Case, D. A. D., T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. *Amber 10*; University of California: San Francisco, CA, 2008.
- (34) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (35) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (36) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14* (33–38), 27–38.
- (37) Okamoto, Y.; Sugita, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (38) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.
- (39) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
- (40) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824–19839.
- (41) Dill, K. A. *J. Biol. Chem.* **1997**, *272*, 701–704.
- (42) Tsui, V.; Case, D. A. *Biopolymers* **2000**, *56*, 275–291.
- (43) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.
- (44) Martin, W.; Zhu, W.; Krilov, G. *J. Phys. Chem. B* **2008**, *112*, 16076–16089.
- (45) Jain, A. K.; Murty, M. N.; Flynn, P. J. *ACM Comput. Surv.* **1999**, *31*, 264–323.
- (46) Dickerson, R. E. *J. Biomol. Struct. Dyn.* **1989**, *6*, 627–634.
- (47) Parish, J. H. *Biochem. Educ.* **1985**, *13*, 92–92.
- (48) Gowtham, S.; Scheicher, R. H.; Ahuja, R.; Pandey, R.; Karna, S. P. *Phys. Rev. B* **2007**, *76*, 033401–033403.

- (49) Sowerby, S. J.; Cohn, C. A.; Heckl, W. M.; Holm, N. G. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 820–822.
- (50) Sood, A. K.; Das, A.; Maiti, P. K.; Das, M.; Varadarajan, R.; Rao, C. N. R. *Chem. Phys. Lett.* **2008**, *453*, 266–273.
- (51) Cathcart, H.; Nicolosi, V.; Hughes, J. M.; Blau, W. J.; Kelly, J. M.; Quinn, S. J.; Coleman, J. N. J. *Am. Chem. Soc.* **2008**, *130*, 12734–12744.
- (52) Meng, S.; Wang, W. L.; Maragakis, P.; Kaxiras, E. *Nano Lett.* **2007**, *7*, 2312–2316.