THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

Theoretical Studies of the Base Pair Fidelity of Selenium-Modified DNA

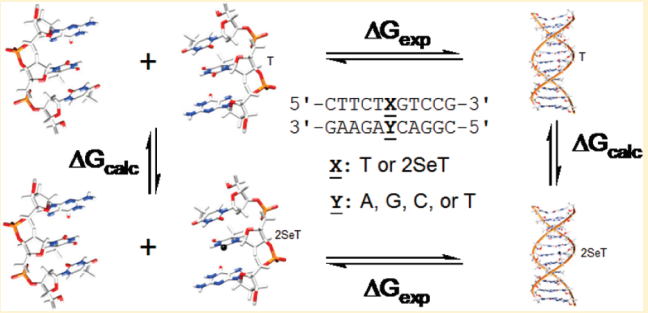
Andrew Christofferson,[†] Lifeng Zhao,[†] Hanzi Sun,[†] Zhen Huang,[‡] and Niu Huang*[,†]

[†]National Institute of Biological Sciences, Beijing, No. 7 Science Park Road, Zhongguancun Life Science Park, Changping District, Beijing 102206, People's Republic of China

[‡]Department of Chemistry, Georgia State University, Atlanta, Georgia 30303, United States

Supporting Information

ABSTRACT: The introduction of selenium into DNA in the place of oxygen provides a unique opportunity for studying the fidelity of DNA replication, as well as providing advantages in the growth of DNA crystals and the greater resolution of their structures. However, the atomic mechanisms of the relative stability and base pair recognition of the selenium-modified DNA are poorly understood. In the present study, quantum mechanics calculations were performed on base pairings, base stacking, and base−water interactions for both unmodified thymine and thymine with the 2-exo-oxygen replaced with selenium, and the results were used to develop and validate CHARMM force field parameters for the 2-Se-thymine. Subsequently, molecular dynamics simulations and free-energy perturbation calculations were performed on 11-base DNA sequences containing native thymine and the 2-Se-thymine. The calculated relative free-energy values are in good agreement with experimentally determined relative stability, where the 2-Se-thymine offers similar stability to T-A in cognate DNA, while it dramatically destabilizes the DNA containing the T-G mismatch base pair when 2-Se-thymine is incorporated. Thus, 2-Se-thymine largely increases the native T-A base pair fidelity by discouraging the T-G wobble pair. Insights into the high pairing specificity and the relative stability of selenium-modified DNA were obtained based on detailed structural and energetic analysis of molecular dynamics trajectories. Our studies move one step further toward an understanding of the high base pair fidelity and thermodynamic properties of Se-DNA in solution and in protein−DNA complexes.



1. INTRODUCTION

DNA plays a central role in biology, and understanding the structural details and energetic properties of both cognate and modified DNA has been actively pursued.[1,2] For the analysis of function and mechanism of nucleic acids and nucleic acid−protein complexes, high-resolution 3D complex structures are of great benefit. X-ray crystallography is an important means by which structural information may be obtained at the atomic level[3,4] but is hampered by difficulties in obtaining phase information and high-quality crystal structures. One solution to this problem is the addition of a selenium atom as an anomalous scattering center for crystal diffraction. Currently, over two-thirds of new protein crystal structures are determined by multiwavelength anomalous dispersion (MAD) phasing of selenomethionyl proteins.[5−7]

Recently, work by Huang and colleagues has resulted in a strategy for replacing any specific nucleic acid oxygen with selenium.[8] Selenium derivatization has the advantage over other modifications (e.g., bromine) in that selenium-modified DNA (Se-DNA) has greater stability, less structure perturbation, and a larger range in choice of incorporation site.[9] Additionally, Se-DNA is easier to synthesize and purify than selenomethionyl protein, making it a useful tool for the study of complicated nucleic acid−protein complex structures.

When the 2-exo-oxygen of thymine is replaced with selenium (denoted as 2SeT), UV denaturation studies determined that the 2SeT-G DNA duplex melted at 6.7 °C lower than the unmodified T-G duplex, while the 2SeT-A duplex only melted at 0.4 °C less than the unmodified T-A duplex.[10] Clearly, the 2SeT modification increases the discrimination of the base pairing, specifically discouraging the T-G wobble mismatch while having little effect on the native T-A pairing. This might be caused by the increased van der Waals (vdW) radius and decreased electronegativity of the selenium atom compared to the oxygen atom, which interferes with the wobble base pair interaction (Figure 1). However, the exact mechanism of the discrimination is not clear. Surprisingly, the 2SeT modification lowered the melting point in the case of the T-C mismatch by 4.1 °C, while it *raised* the melting point in the T-T mismatch by 1 °C. Due to the limitation of experimental methods to explain these melting temperature changes in microscopic detail,

Received: May 6, 2011
Revised: July 18, 2011
Published: July 19, 2011

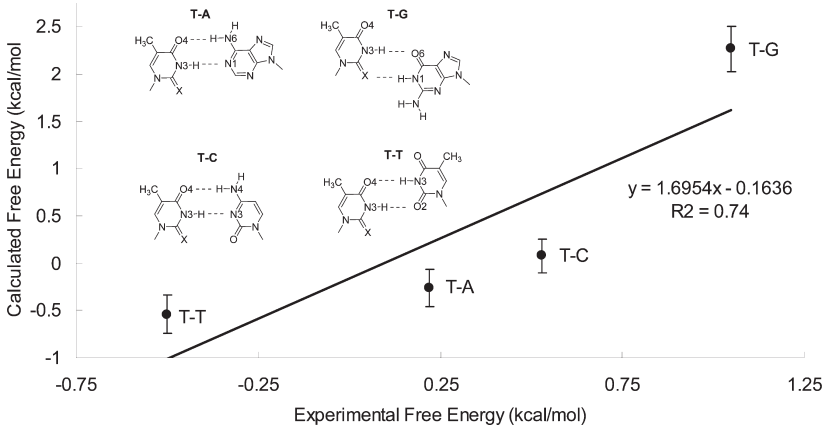


Figure 1. Comparison between experimental and calculated $\Delta\Delta G$ values of 2SeT modifications in DNA duplexes, with a linear regression and schematics of the base pairing interactions. X represents O_2 in the reactant step and Se in the product step. The errors of the free-energy calculations were calculated from the standard deviations from five wham blocks multiplied by the uncertainties of 21 simulations of each free-energy calculations.

computational studies have been undertaken to study the fundamental atomic interactions involved in the modified and unmodified bases.

Previous computational work by Vázquez-Mayagoitia and colleagues found that, for a thymine molecule with the 4-exo-oxygen replaced by selenium (denoted as 4SeT), the T-A base pairing gas-phase interaction energy determined using quantum mechanics (QM) methods at the MP2/6-31G** level is decreased by 3.9 kcal/mol, while the base stacking interaction energy increases by 3.6 kcal/mol. Additionally, they found that the 4SeT substitution reduces the HOMO−LUMO gap and increases the interplane coupling, which suggests that Se-DNA has potential use as nanowire or other nanodevice applications.[11] However, base pairing and stacking interactions of the 2SeT modification have not been investigated to date, nor have computational studies of any type been performed on selenium-modified bases in the context of solvated DNA.

While ab initio calculations with high level theories and large basis sets have been frequently applied to compute small model systems containing dozens of atoms, such calculations are prohibitively expensive for biomolecular simulations in practice. On the other hand, molecular mechanics (MM) methods depend on classical mechanics and calculate structural and energetic characteristics of large biomolecular systems on the basis of empirical force fields (FF).[12] The development and refinement of DNA force fields has allowed the dynamic simulation of fully solvated DNA duplexes to reach a degree of accuracy that consistently reproduces experimental results[13−15] and, therefore, have provided opportunities to study a broad range of DNA, DNA−protein, and RNA−protein interactions using molecular dynamics (MD) simulation techniques, including, but not limited to, "base flipping" studies,[16] binding orientation investigations,[17] stability studies of DNA with modified bases,[18] and hydration analyses.[19]

While preceding experimental and computational projects have begun to shed light onto the nature and potential applications of Se-DNA, there remain unanswered questions, particularly the contributions from different intrinsic and environmental effects to the selenium modification. In this work, we attempt to address some of these questions with ab initio gas-phase calculations of base pairing, base stacking, and base−water interactions for the 2SeT base, along with the development of CHARMM force field parameters for use in MD simulation and free-energy perturbation (FEP) calculations in fully solvated DNA systems. It is evident that base pairing interactions alone are not sufficient to explain the differences between cognate and Se-modified DNA, as both base stacking and interactions with solvent have important effects. The calculated free-energy values that we determined are in reasonable agreement with the experimental data, and insights into the base pair fidelity and the stability of selenium-modified DNA were obtained by examining both enthalpic and entropic contributions in simulation systems. Clearly, our success in predicting relative DNA stabilities using free-energy calculation methods will guarantee further investigations of protein−DNA interactions resulting from selenium modifications.

2. METHODS

2.1. Quantum Mechanics Calculations. Initial molecular geometries for ab initio calculations of the base pairing and stacking interactions were obtained from DNA X-ray crystal structures (Tables 1 and 2). Interaction energies were calculated from optimized structures at the MP2/6-311G** level by GAMESS-US,[20] with the basis set superposition error (BSSE) corrected via the counterpoise method.[21] Full geometry optimizations were carried out for base pairing interactions, while heavy atom coordinates were frozen in all base stacking calculations, as the MP2 method tends to overestimate base stacking interactions.[22] Additional base stacking interaction calculations were performed for geometries from the JCSH-2005 database,[23] also with heavy atom coordinates frozen. For the interactions with water, initial coordinates of thymine and structural water molecules were obtained from the X-ray crystal structure of the native 8-base DNA duplex (PDB ID: 1D78).[24] Each thymine−water interaction was calculated separately, with the angle and dihedral of the water oxygen frozen and all other coordinates optimized. For each water molecule, interaction energies for both the native thymine and 2SeT molecule were calculated.

2.2. Parameterization of 2SeT. The parametrization of the selenium-modified thymine base 2SeT for the CHARMM27 DNA force field[13,25] follows the general strategy of CHARMM force field development,[26] as specifically outlined by Sarzyńska

Table 1. Base Pairing Interaction Energies and Hydrogen Bond Distances for Full Geometry Minimization Using MP2/6-311G Theory and CHARMM Force Field with Our Derived 2SeT Parameters (Hydrogen Bond Distances in Crystal Structures, Where Available, Are Listed in Parentheses)**

base pair	MP2/6-311G**			CHARMM			difference[f]
T-A pair	$T_{O4}-A_{N6}$ (Å)	$T_{N3}-A_{N1}$ (Å)	E_{int} (kcal/mol)	$T_{O4}-A_{N6}$ (Å)	$T_{N3}-A_{N1}$ (Å)	E_{int} (kcal/mol)	E_{diff} (kcal/mol)
T-A[a]	2.98 (2.92)	2.85 (2.83)	−12.76	2.88 (2.92)	2.88 (2.83)	−12.82	0.06
2SeT-A[b]	2.93 (2.88)	2.88 (2.91)	−13.22	2.81 (2.88)	2.90 (2.91)	−12.87	−0.35
T-G pair	$T_{O2}-G_{N1}$ (Å)	$T_{N3}-G_{O6}$ (Å)	E_{int} (kcal/mol)	$T_{O2}-G_{N1}$ (Å)	$T_{N3}-G_{O6}$ (Å)	E_{int} (kcal/mol)	E_{diff} (kcal/mol)
T-G[c]	2.81 (2.94)	2.88 (3.06)	−14.22	2.79 (2.94)	2.83 (3.06)	−13.82	−0.04
2SeT-G[c]	3.43	2.78	−13.36	3.43	2.80	−9.53	−3.83
T-C pair	$T_{O4}-C_{N4}$ (Å)	$T_{N3}-C_{N3}$ (Å)	E_{int} (kcal/mol)	$T_{O4}-C_{N4}$ (Å)	$T_{N3}-C_{N3}$ (Å)	E_{int} (kcal/mol)	E_{diff} (kcal/mol)
T-C[d]	2.89 (2.85)	2.92 (3.04)	−11.96	2.86 (2.85)	3.00 (3.04)	−11.33	−0.63
2SeT-C[d]	2.94	2.83	−12.00	2.86	2.97	−12.46	0.46
T-T pair	$T_{N3}-T_{O2}$ (Å)	$T_{O4}-T_{N3}$ (Å)	E_{int} (kcal/mol)	$T_{N3}-T_{O2}$ (Å)	$T_{O4}-T_{N3}$ (Å)	E_{int} (kcal/mol)	E_{diff} (kcal/mol)
T-T[e]	2.89 (2.98)	2.88 (2.82)	−10.17	2.82 (2.98)	2.85 (2.82)	−10.12	−0.05
2SeT-T[e]	2.91	2.84	−10.07	2.82	2.83	−10.63	0.56

[a] Initial coordinates obtained from the 1D78[24] crystal structure. [b] Initial coordinates from the 3HGD[10] crystal structure. [c] Initial coordinates from the 1X2X[42] crystal structure, with oxygen substituted manually by selenium to construct the 2SeT mutant. [d] Initial coordinates from the 1FKY[43] crystal structure, with oxygen substituted manually by selenium to construct the 2SeT mutant. [e] Initial coordinates from the 1DSD[44] crystal structure, with oxygen substituted manually by selenium to construct the 2SeT mutant. [f] Interaction energy difference of MP2/6-311G** E_{int} − CHARMM E_{int}.

Table 2. Base Stacking Interaction Energies for Geometries Obtained from X-ray Crystal Structures and JSCH-2005 Database at the MP2/6-311G Level of Theory and CHARMM27 Force Field Using Our Derived 2SeT Parameters**

	crystal structure[a]	CHARMM	JSCH-2005[b]	CHARMM
stack	E_{int} (kcal/mol)	E_{int} (kcal/mol)	E_{int} (kcal/mol)	E_{int} (kcal/mol)
T/A	−5.95	−8.62	−9.12	−8.32
2SeT/A	−6.66	−8.73	−10.55	−9.98
T/G	−4.49	−7.53	−9.57	−11.20
2SeT/G	−5.44	−8.75	−10.41	−13.37
T/C	−4.53	−6.08	−7.35	−9.67
2SeT/C	−4.13	−5.15	−7.99	−11.24
T/T	−2.81	−4.79	−6.20	−6.09
2SeT/T	−2.78	−5.22	−6.50	−7.86

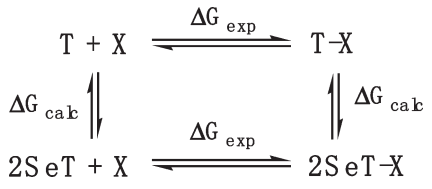
[a] Initial coordinates were obtained from the 1D78[24] crystal structure for T/A and T/G and the 3HGD[10] crystal structure for 2SeT/A and 2SeT/G. The remaining coordinates were obtained from the 1AGH[28] crystal structure, with oxygen substituted manually by selenium to construct the 2SeT mutants. [b] Initial coordinates were obtained from the JSCH-2005 database,[23] with oxygen substituted manually by selenium to construct the 2SeT mutants.

and colleagues in their work on the parametrization of 2- and 4-sulfur thymine.[27] Initial Lennard-Jones (LJ) 6−12 parameters ε_{ij} (the LJ well depth) and $R_{min,ij}$ (the interatomic distance at the LJ energy minimum) for selenium were obtained by linearly extrapolating the ab initio interactions between water and thymine, water and 2-sulfur thymine, and water and 2-selenium thymine with respect to the LJ parameters of the well-established oxygen and sulfur atoms, with the assumption that the $R_{min,ij}$ parameter should scale linearly in relation to the internuclear distance, while the ε_{ij} parameter should scale linearly in relation to the QM energy (Table S1, Supporting Information). Initial LJ parameters for selenium were subsequently optimized through an iterative process along with the charge parameter optimization. Target data for the 2SeT atomic partial charges were minimum interaction energies and distances of QM calculations of the base−water interactions at the MP2/6-31* level. Equilibrium bond lengths and angles were obtained from crystal structure geometries. More details of the base parametrization can be found in the Supporting Information.

2.3. Molecular Dynamics Simulations. The focus of the molecular dynamics simulations was on the 11-base heterodimer 5′-CTTCTTGTCCG-3′ (target T is underlined), for which UV denaturation experiments by Huang and colleagues have been performed for the DNA duplexes containing target base pairs T-A, T-G, T-C, and T-T and corresponding selenium-modified thymine base pairs 2SeT-A, 2SeT-G, 2SeT-C, and 2SeT-T.[10] Initial coordinates for the 11-base DNA duplex were obtained from the X-ray crystal structure of the native DNA duplex (PDB ID: 1AGH).[28] The T-G, T-C, and T-T mismatched duplexes were generated by replacing the adenine with guanine, cytosine, and thymine, respectively.

All simulations were performed using CHARMM version c35b1[29] with the CHARMM27 DNA force field.[13,25] Hydrogens

Scheme 1. Thermodynamic Cycle for the Alchemical Transformation of Thymine to 2-Se-Thymine in DNA, Where X Represents A, G, C, or T



were added to the DNA using CHARMM, and DNA structures were overlaid and centered with respect to a TIP3P[30] water box of 60 Å^3. All solvent molecules having a non-hydrogen atom within 1.8 Å of DNA were deleted, and sodium ions were added to obtain electrical neutrality. Periodic boundary conditions were applied, and the particle-mesh Ewald (PME)[31,32] method was utilized for treating long-range electrostatics. All calculations used an atom-based truncation scheme updated heuristically with a list cutoff of 14 Å, a nonbond cutoff of 12 Å, and the Lennard-Jones (LJ) smoothing function initiated at 10 Å. SHAKE[33] was applied to covalent bonds involving hydrogen, and a 2 fs time step was used.

In each case, the resulting system was minimized for 4000 steps, followed by a 400 ps MD equilibration with a 2 kcal/mol/Å^2 force restraint on all heavy atoms of DNA at a constant temperature of 300 K. This was followed by an additional 400 ps MD simulation with a 2 kcal/mol/Å^2 force restraints only on the O5′ or O3′ oxygen atoms of terminal bases. For each system, the equilibration step was repeated until base pairing interactions resembled crystal structure geometries (Table 1). Subsequently, 10 ns production runs were performed with the same restraints as the final equilibration.

Trajectory analysis was performed mainly using the CHARMM program. Analysis of the MD simulations included structural and energetic properties averaged over the 10 ns production simulation. Initial analysis showed the structural change to be localized in the vicinity of the modified base. Accordingly, analysis focused on the base pairs defined as the central three base pairs centered on the target base T or 2SeT. Structural analysis mainly included standard root-mean-square differences (rmsd values) and hydrogen bonding. Hydrogen bonds were defined as acceptor-to-donor distances of ≤3.5 Å. Energetic analysis included intrinsic energies (e.g., base pairing, base stacking) and intermolecular energies (DNA−solvent). Solvent entropy contribution was investigated by measuring the number of water molecules as the function of distance between the water oxygen atom and the 2-oxygen atom of the target T or the 2-selenium atom of the 2SeT and calculating both the occupancy and the stability of stable water molecules identified using an in-house program analogous to Schrodinger WaterMap software,[34] which was applied to cluster the water distribution from 1000 simulation frames to identify positions with a high density of water neighbors within a 1 Å radius. Occupancy is defined as the total number of water molecules observed within the radius in all of the frames divided by the total number of frames, and stability is defined as the total number of water molecules observed within the radius in all of the frames divided by the total number of individual water molecules observed at any time within the 1 Å radius.

2.4. Free-Energy Perturbation Simulations. In order to calculate the free-energy difference between the native and selenium-modified DNA, the free-energy perturbation[35,36] method was utilized. For the FEP calculations, stable conformations of the DNA were extracted from MD trajectories for use as initial structures for the perturbation simulations.

The experimental free-energy values determined by Huang and colleagues represent the free energy of destabilization of the 11-base DNA duplex, that is, the ΔG_{exp} in Scheme 1. In order to complete the thermodynamic cycle, FEP calculations for the alchemical transformation of thymine to Se-modified thymine for both the DNA duplex and for the single-strand DNA are required (i.e., the ΔG_{calc}). For the simulation of the single strand, a 3-base sequence 5′-TTG-3′ centered at the target thymine was used.We expect that such an approximation is capable of capturing the localized effects of a single-strand DNA.

For each simulation, the unmodified and selenium-modified DNA represent the two thermodynamic states 0 and 1, and the reaction path between them is divided into N nonphysical intermediate windows according to the coupling parameter λ, which varies from 0 to 1 accordingly:

$$\Delta G_{0\rightarrow 1} = G(\lambda_1) - G(\lambda_0) = -k_B T \sum_{k=1}^{N-1} \ln\left\langle \exp\left[-\frac{U(x,\lambda_{k+1}) - U(x,\lambda_k)}{k_B T}\right]\right\rangle_{\lambda_k} \quad (1)$$

where ΔG is the free-energy difference between the two states, k_B is the Boltzmann constant, T is the temperature (300 K), and U is the potential energy of the system, which is dependent on the Cartesian coordinates of each window and the coupling parameter.

In each FEP simulation, the reaction path was divided into 21 windows, with the first and last of length 0.025 and the intermediate windows of length 0.05. An initial equilibration simulation of 400 ps was followed by five 1 ns production runs, for a total of 5 ns of data collection. Simulations were run in both forward and reverse directions. All other simulation conditions were identical to the MD simulations. Postprocessing was performed by WHAM.[37−40] The uncertainties were propagated from the standard errors of each error of the free-energy calculations, thus, the errors of the free-energy calculations were calculated from the standard deviations from five wham blocks multiplied by the uncertainties of 21 simulations of each free-energy calculations.

3. RESULTS AND DISCUSSION

3.1. QM Interaction Energies. Gas-phase QM calculations alone may not adequately explain the experimental free-energy values. For the full geometry minimization at the MP2/6-311G** level, the 2-selenium substitution appears to stabilize the T-A base pair interaction by −0.46 kcal/mol; −13.22 versus −12.76 kcal/mol, respectively (Table 1). A frequency analysis reveals that the 2SeT$_{N3}$−A$_{N1}$ hydrogen bond is stabilized relative to the T-A pair (2983 versus 3074 cm^{-1}), while the 2SeT$_{O4}$−A$_{N6}$ hydrogen bond is slightly destabilized (3486 versus 3457 cm^{-1}). Interestingly, the 2-selenium modification has very small impact on the geometries and interaction energies for mismatched base pairs T-C and T-T, while it significantly destabilizes the T-G mismatch by 0.86 kcal/mol. Clearly, QM gas-phase base pairing results do not correlate well to experimental free-energy data except in the case of the T-G mismatch. Still, these calculations

Table 3. Interactions between Bases and Structural Water from the 1D78 X-ray Crystal Structure[24] Calculated at the MP2/6-311G and CHARMM27 Levels of Theory Using Our Derived Force Field Parameters for Selenium Modifications**

water	interaction distance (Å)	QM E_{int} (kcal/mol)	CHARMM E_{int} (kcal/mol)	E_{diff} (kcal/mol)
T_{O2}—HOH26	2.96	−3.89	−4.63	0.74
$2SeT_{SE2}$—HOH26	3.38	−0.88	−0.87	−0.01
T_{O4}—HOH23	2.95	−3.83	−4.81	0.98
$2SeT_{O4}$—HOH23	2.97	−3.47	−4.84	1.37
T_{O4}—HOH24	2.94	−4.31	−5.31	1.00
$2SeT_{O4}$—HOH24	2.94	−4.19	−5.50	1.31

provide a useful basis for assessing CHARMM parameters for the 2SeT.

For the base stacking interaction from crystal structure geometry, the selenium substitution stabilizes the stacking interaction in 2SeT/A and 2SeT/G and slightly destabilizes the 2SeT/C and 2SeT/T (Table 2). For the JSCH-2005 geometries, the selenium substitution stabilizes the base stacking interaction in every single case. This suggests that, in general, the selenium modification indeed has an overall stabilizing effect on the stacking interaction regardless of which bases are present, but there is some dependence on local geometry.

Calculations on interactions with crystal structure water molecules, summarized in Table 3, reveal that the selenium substitution decreases the interaction energy by ~3 kcal/mol when the water interacts directly with the selenium and also slightly decreases the interaction energy of the hydrogen bond when the water interacts with the nonsubstituted oxygen atom on the selenium-modified thymine. This suggests that while intrinsic interaction energies (base pairing and base stacking) are generally favorable in 2SeT-modified DNA, its interaction with environmental waters is largely unfavorable. While these results indicate some general trends in interaction energies in static configurations, they also illustrate the importance of the environment and dynamic properties of the molecular orientations. Nevertheless, it is difficult to predict overall DNA stabilities based on gas-phase QM energies without consideration of thermodynamic properties in solution.

3.2. 2SeT Force Field Parameters. To more thoroughly investigate the structural and energetic properties of fully solvated DNA duplexes, the only currently feasible means is the use of molecular mechanics-based calculations. Therefore, we developed force field parameters for the 2SeT-modified base using the target data from QM calculation results mentioned above, in combination with high-resolution crystal structures containing the native T-A pairing, as well as the 2SeT base. We aimed to develop a reasonably good set of parameters for simulating the DNA containing 2SeT base; further development of high-quality parameters for selenium-modified thymine is desirable but not within the scope of this current work. Partial atomic charges for all atoms, LJ ε_{ij} and $R_{min,ij}$ values, and equilibrium bond, angle, dihedral, and force constant parameters for the 2SeT base are reported in the Supporting Information.

Table 1 shows a comparison between the base pair interaction energies and hydrogen bond distances of the developed CHARMM27 parameters for the 2SeT base and ab initio interaction energies at the MP2/6-311G** level of theory. It is encouraging that agreement between CHARMM and MP2 values is reasonably good, and slight differences between gas-phase energies are less important so long as the CHARMM parameters are able to reproduce bulk phase properties. One notable exception is the 2SeT-G base pairing interaction, where the difference between the CHARRM and MP2 energies is 3.83 kcal/mol. While the reason is not immediately clear, this large difference may explain the lack of close agreement between the calculated and experimental free-energy values for the 2SeT-G system (discussed further in section 3.3). The qualitative trends in the CHARMM and QM stacking energies also agree reasonably well (Table 2), considering that both QM calculated base pairing and base stacking energies were not used for parametrizing 2SeT.

Table 4. Comparison of Experimental $\Delta\Delta G$ to Calculated $\Delta\Delta G$[a]

	$\Delta\Delta G_{exp}$ (kcal/mol)	$\Delta\Delta G_{calc}$ (kcal/mol)
T-A vs 2SeT-A	0.22	−0.27 ± 0.20
T-G vs 2SeT-G	1.05	2.27 ± 0.24
T-C vs 2SeT-C	0.53	0.08 ± 0.18
T-T vs 2SeT-T	−0.50	−0.54 ± 0.20

[a] The errors of the free-energy calculations were calculated from the standard deviations from five wham blocks multiplied by the uncertainties of 21 simulations of each free-energy calculations.

3.3. FEP Simulations. Experimental free-energy values were derived from a curve fitting of melting point data.[10] A comparison of the $\Delta\Delta G$ values for the experimental and calculated reactions for the 11-base DNA sequence is shown in Table 4, and the free-energy values calculated from both forward and reverse directions are reported in Table S5 (Supporting Information). Overall, the free energy in all simulation systems reaches convergence quickly, judged by its small fluctuation within 5 ns time scale. While the denaturation experiments found the 2SeT substitution to be slightly energetically unfavorable for the T-A pairing, calculations found it to be slightly favorable. Given the uncertainties in both experimental and calculated free-energy values, it is difficult to argue with any certainty whether the 2-selenium modification truly stabilizes or destabilizes the DNA in a T-A base pairing. Regardless, the effect that the modification has on DNA stability is certainly small. For the T-G pairing, both experimental and computational results found the selenium modification to destabilize the DNA duplex, while the calculated value overestimated the destabilization by slightly more than 1 kcal/mol. This may be due to the aforementioned differences in the CHARMM and QM gas-phase pairing energies. For the T-C pairing, the calculation underestimated the free-energy difference by 0.47 kcal/mol. For the T-T pairing, both experiment and calculation gave almost identical values of about −0.50 kcal/mol.

Regardless of the absolute values of the free-energy differences, the trend of the overall stabilization of the various base pairings is the same for both experimental and calculated values (Figure 1). Considering all of the practical issues that potentially lead to large calculation errors, such as the quality of 2SeT parameters, the simplified reference state of the single-strand DNA model and the starting geometries of mismatched base pairs, the semiquantitative agreement with experimental results is encouraging.

3.4. Extended Molecular Dynamics Simulations and Trajectory Analysis. MD-based free-energy determination methods

Table 5. Average Base−Solvent, Base Pairing, and Stacking Interaction Energies and Their Standard Deviations from the 10 ns MD Simulations

	thymine−solvent (kcal/mol)	base pairing (kcal/mol)	stacking with 5′ T base (kcal/mol)	stacking with 3′ G base (kcal/mol)
T-A	−22.99 ± 3.03	−12.92 ± 0.09	−4.14 ± 0.12	−5.53 ± 0.08
2SeT-A	−15.16 ± 3.51	−13.87 ± 0.14	−4.96 ± 0.10	−6.01 ± 0.27
T-G	−20.82 ± 4.07	−14.28 ± 0.14	−5.15 ± 0.13	−5.03 ± 0.14
2SeT-G	−22.67 ± 2.65	−9.21 ± 0.17	−4.72 ± 0.32	−6.24 ± 0.57
T-C	−23.95 ± 2.61	−7.83 ± 0.53	−3.90 ± 0.13	−7.45 ± 0.51
2SeT-C	−10.02 ± 4.14	−10.74 ± 0.12	−5.21 ± 0.18	−8.07 ± 0.46
T-T	−22.49 ± 2.15	−8.67 ± 0.09	−3.74 ± 0.12	−7.48 ± 0.21
2SeT-T	−12.47 ± 2.65	−9.71 ± 0.13	−4.51 ± 0.13	−8.68 ± 0.22

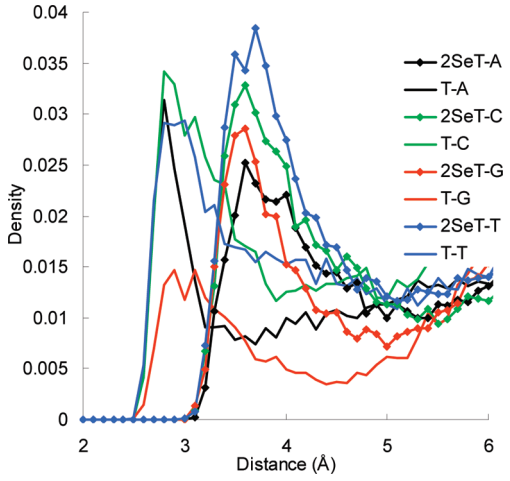


Figure 2. Density of solvent water molecules as the function of distance between the water oxygen atom and the 2-oxygen atom of the target thymine or 2-selenium atom of 2SeT in the 10 ns MD simulations. Density refers to the average number of waters in a shell of 0.1 Å thickness centered around the target oxygen or selenium and a radius that varies according to interatomic distance (*x*-axis), divided by the volume of the shell.

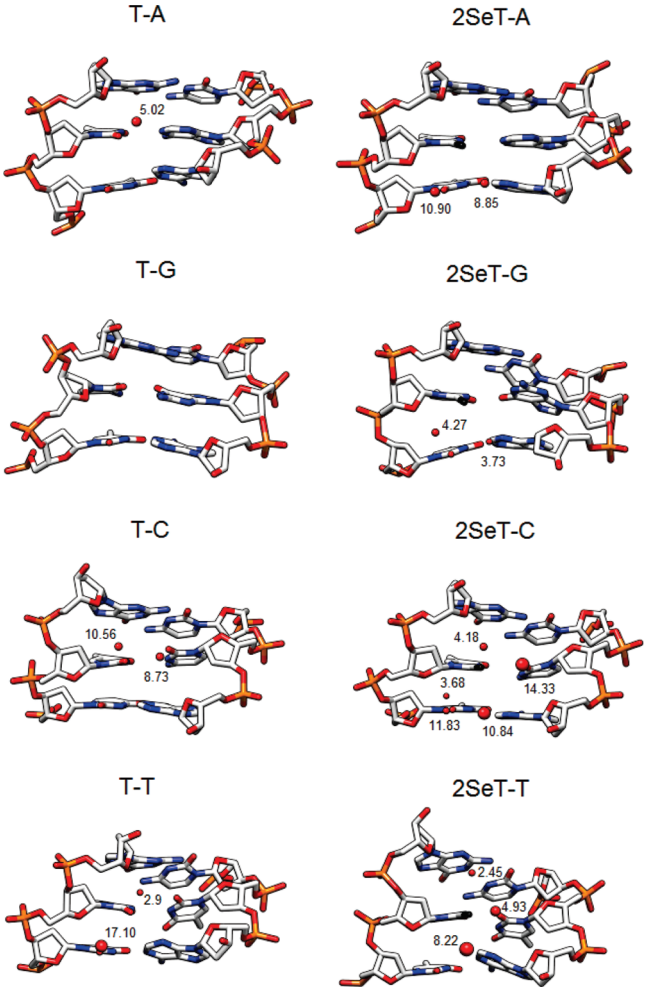


Figure 3. Interactions with stable water molecules for the 2-oxygen atom of the target thymine or 2-selenium atom (black) of 2SeT in the 10 ns MD simulations. The size of the sphere representing the water is proportional to the percent occupancy of any water in that location, and the corresponding number represents the average stability of a single water molecule in that location. Only waters with an occupancy above 30% are shown.

allow calculation of the difference of the thermodynamic behaviors between cognate DNA and modified DNA. However, the free energy depends on a balance of many different intrinsic and environmental contributions; the relative free energy alone is not adequate for understanding the atomic details of these different contributions. Therefore, extensive structural and energetic analysis was undertaken with emphasis on the atomic events that contribute to the free-energy stability. It should be noted that a free energy cannot be formally separated into component contributions. However, such an analysis can yield useful insights into the atomic contribution on the thermodynamic properties.[41] The average solvent interaction energy, base pairing interaction, and base stacking interaction energies for the target T or 2SeT are presented in Table 5, and the rmsd and hydrogen bond analysis are reported in Table S4 and Figure S2 (Supporting Information). Although it is plausible to consider the base pairing and stacking interactions as the dominant players, the contribution from solvent entropy may also be important. The solvent entropy contribution, which arises from a significant loss of conformational freedom of the solvent upon exposure of the target base to the solvent, is extremely difficult to quantify accurately. Thus, we simply measured the number of solvent molecules accessible to the target base as well as considered the occupancy and stability of the identified stable water. We expect that the overall trend in stability of the Se-DNA duplex can be explained by combining enthalpic and entropic factors.

The density of water around the 2-oxygen of thymine and 2-selenium of 2SeT is shown in Figure 2. Water density increases considerably for the 2SeT-G pairing and slightly for the 2SeT-T, while it decreases for the 2SeT-A and 2SeT-C. The specific location and orientation of high-density water occupancy around the 2-oxygen and 2-selenium are shown in Figure 3 and Table S6 (Supporting Information).

A qualitative explanation for the observed free-energy trends is straightforward, while an explicit quantitative analysis is problematic. In every case except for the T-G base pairing, the selenium modification destabilizes the base−solvent interaction, while stabilizing the base pairing interaction and base stacking interaction. Compared to the native T-A pairing, the 2SeT-A pairing system is favored energetically, and while the stability of the solvent increases, the overall density of the solvent decreases. For the wobble T-G pairing, though both the solvent density and stability increase, the 2SeT-G pairing system is energetically disfavored drastically, leading to an overall large unfavorable 2SeT-G pairing. With the T-C mispairing, the 2SeT-C system is favored energetically, and its solvent density decreases slightly, but solvent stability around the 2SeT base increases. Because both experimental and calculated $\Delta\Delta G$ values are positive, it is clear that the entropic effects dominate for the 2SeT-C system. Compared to the T-T mispairing, the 2SeT-T system is energetically favored, and while its water density and water occupancy increase, the stability of these waters is significantly less than that of the native T-T mispairing. It seems that the decreased water stability somewhat leads to a favorable entropic contribution to the system.

4. CONCLUSIONS

While QM calculations for the base pairing and base stacking of native and Se-modified thymine provide useful information about the nature of modified and unmodified DNA base interactions, the fact that the trends in interaction energy do not match the experimentally observed trends in free energy for the various interactions demonstrates that QM calculations alone cannot adequately explain bulk phase properties of Se-DNA. However, these QM calculations do provide good target data for force field parametrization and validation. MD-based FEP simulations demonstrate that the developed parameters from the QM calculations yield free-energy values in a good agreement with experimental data, thereby confirming that the parameters are reasonable in modeling the Se-modified thymine. Furthermore, analysis of MD simulation trajectories reveals an explanation for the observed trend in both experimental and calculated free-energy differences between the Se-modified and unmodified DNA when taking into account both entropic and enthalpic effects. Consistent with the experimental results, the calculated relative free energy indicates that the 2-Se-modified T-A pair and the native T-A pair have a similar stability, while the modified T-G pair is significantly destabilized, compared to the native T-G wobble pair. The free-energy calculation suggests that the source of this high specificity of the 2SeT-A pair is due to both the disfavored hydrogen bonding interaction of the 2SeT-G wobble pairing and the unfavorable solvent entropic contribution. This reveals that the modified T-A pair has higher base pair fidelity than the natural T-A pair. Therefore, these new force field parameters for the 2-Se-thymine have been shown to be in good agreement with both theoretical and experimental data, and such molecular dynamics-based free-energy calculations are a promising approach in future studies of selenium-modified DNA and protein−DNA complexes.

■ ASSOCIATED CONTENT

Supporting Information. The CHARMM force field parameters and topology file for the 2-selenium-modified thymine base, full FEP forward and reverse simulation free energies, rmsd data for the DNAs and water occupancy around the T or 2SeT in the 10 ns MD simulations, and atomic coordinates for the base pairing and stacking QM calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: huangniu@nibs.ac.cn.

■ ACKNOWLEDGMENT

Financial support from the National Chinese Ministry of Science and Technology "863" Grant 2008AA022313 (to N.H.) and the USA NSF grants (CHE-0750235 and MCB-0824837 to Z. H.) is gratefully acknowledged. We thank Q. Pei and X.Y. Lin at NIBS for advice in parametrization and FEP calculations. Computational support was provided by the Supercomputing Center of Chinese Academy of Sciences (SCCAS) and the Beijing Computing Center (BCC). We thank the reviewers for helpful comments.

■ REFERENCES

(1) Srinivasan, A. R.; Sauers, R. R.; Fenley, M. O.; Boschitsch, A. H.; Matsumoto, A.; Colasanti, A. V.; Olson, W. K. *Biophys. Rev.* **2009**, *1*, 13.
(2) Schlick, T. Topics in Nucleic Acids Structure: DNA Interactions and Folding. In *Molecular Modeling and Simulation: An Interdisciplinary Guide*; Marsden, J. E., Sirovich, L., Wiggins, S., Eds.; Springer: New York, 2010; Vol. 21, p 163.
(3) Rambo, R. P.; Tainer, J. A. *Curr. Opin. Struct. Biol.* **2010**, *20*, 128.
(4) Klug, A. *Annu. Rev. Biochem.* **2010**, *79*, 1.
(5) Deacon, A. M.; Ealick, S. E. *Structure* **1999**, *7*, R161.
(6) Hendrickson, W. A. *Trends Biochem. Sci.* **2001**, *25*, 637.
(7) Jiang, J.; Sweet, R. M. *J. Synchrotron Radiat.* **2004**, *11*, 319.
(8) Sheng, J.; Huang, Z. *Chem. Biodiversity* **2010**, *7*, 753.
(9) Jiang, J.; Sheng, J.; Carrasco, N.; Huang, Z. *Nucleic Acids Res.* **2007**, *35*, 477.
(10) Hassan, A. E. A.; Sheng, J.; Zhang, W.; Huang, Z. *J. Am. Chem. Soc.* **2010**, *132*, 2120.
(11) Vazquez-Mayagoitia, A.; Huertas, O.; Brancolini, G.; Migliore, A.; Sumpter, B. G.; Orozco, M.; Luque, F. J.; Felice, R. D.; Fuentes-Cabrera, M. *J. Phys. Chem. B* **2009**, *113*, 14465.
(12) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646.
(13) MacKerell, A. D.; Banavali, N.; Foloppe, N. *Biopolymers* **2000**, *56*, 257.
(14) Cheatham, T. E.; Young, M. A. *Biopolymers* **2000**, *56*, 232.
(15) MacKerell, A. D.; Nilsson, L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 194.
(16) Huang, N.; Banavali, N. K.; MacKerell, A. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 68.
(17) Garcia-Nieto, R.; Manzanares, I.; Cuevas, C.; Gago, F. *J. Am. Chem. Soc.* **2000**, *122*, 7172.
(18) Florian, J.; Goodman, M. F.; Warshel, A. *J. Phys. Chem. B* **2000**, *104*, 10092.
(19) Steinbach, P. J.; Brooks, B. R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9135.
(20) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Kosecki, S.; Matsunaga, N.; Nguyen, K. A.;

Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
(21) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 533.
(22) Riley, K. E.; Pitonak, M.; Cerny, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 66.
(23) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
(24) Thota, N.; Li, X. H.; Bingman, C.; Sundaralingam, M. *Acta Crystallogr., Sect. D* **1993**, *49*, 282.
(25) Foloppe, N.; MacKerell, A. D. *J. Comput. Chem.* **2000**, *21*, 86.
(26) Mackerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584.
(27) Sarzyńska, J.; Kuliński, T. *Comput. Methods Sci. Technol.* **2005**, *11*, 49.
(28) Feng, B.; Stone, M. P. *Chem. Res. Toxicol.* **1995**, *8*, 821.
(29) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.
(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Biol. Chem.* **1983**, *79*, 926.
(31) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(32) Darden, T.; Perera, L.; Li, L. P.; Pedersen, L. *Struct. Fold. Des.* **1999**, *7*, R55.
(33) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(34) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2008**, *130*, 2817.
(35) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.
(36) Boresch, S.; Karplus, M. *J. Phys. Chem. B* **1999**, *103*, 103.
(37) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(38) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339.
(39) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275.
(40) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40.
(41) Boresch, S.; Archontis, G.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 25.
(42) Alvarez-Salgado, F.; Desvaux, H.; Boulard, Y. *Magn. Reson. Chem.* **2006**, *44*, 1081.
(43) Boulard, Y.; Cognet, J. A.; Fazakerley, G. V. *J. Mol. Biol.* **1997**, *268*, 331.
(44) Lian, C.; Robinson, H.; Wang, A. H.-J. *J. Am. Chem. Soc.* **1996**, *118*, 8791.