Decomposition of the Solvation Free Energies of Deoxyribonucleoside Triphosphates Using the Free Energy Perturbation Method

Urban Bren,[†] Václav Martínek,[‡] and Jan Florián*

Department of Chemistry, Loyola University Chicago, Chicago, Illinois 60626

Received: November 16, 2005; In Final Form: May 5, 2006

Free energy perturbation (FEP) calculations using the Amber 95 force field and the TIP3P water model were carried out to evaluate the solvation free energy of deoxyribonucleoside triphosphates in aqueous solution. Solvation free energies of −307.5, −311.5, −314.1, and −317.0 kcal/mol were calculated for the $(Mg{\cdot}dTTP)^{2-}$, $(Mg{\cdot}dATP)^{2-}$, $(Mg{\cdot}dCTP)^{2-}$, and $(Mg{\cdot}dGTP)^{2-}$ complexes, respectively. Structural origins of the relative solvation free energies of deoxyribonucleoside phosphates were examined by calculating the contribution of the interaction of the base moiety with its surroundings. We showed that for each nucleobase the magnitude of this contribution is unaffected by substituting the 5′-OH group of the corresponding nucleoside with the charged mono- or triphosphate groups. This free energy contribution was further decomposed into the sum of free energies originating from the interactions of the base with itself, its substituent, water, and Na^+ ions. Although the sum of these components was nearly constant over a wide range of solutes the individual free energy constituents varied significantly. Furthermore, this decomposition showed a high degree of additivity. Computational conditions necessary for obtaining additive free energy decomposition for the systems studied here within the framework of the FEP method included the use of a single mutation pathway and a subdivision of the FEP protocol into 51 or more windows.

1. Introduction

The partitioning of the free energy into additive contributions originating from groups of atoms or force field terms has the potential to provide relationships between structure and biological activity of molecules.[1−3] Therefore, free energy decomposition is used as a foundation for many established methods of computer-aided drug design such as scoring functions, quantitative structure−activity relationships, or linear response.[4,5] Whether such decomposition is justified has been vigorously debated,[6−10] with thermodynamic integration[4] serving as a key theoretical tool for this discussion. In this paper, we examine the validity of the free energy decomposition scheme for a single perturbation pathway using the free energy perturbation (FEP) formalism and apply this scheme to investigate substituent effects upon the solvation free energy (ΔG_{solv}) of nucleoside phosphates.

A practical advance, which is pursued here from the perspective of free energy decomposition, is the selection of the most reliable methodology for the calculations of binding of nucleoside triphosphates (or other highly charged ligands) to proteins. To quantify the accuracy requirements of these calculations, one has to realize that dissociation constants of 10^{-3} to 10^{-9} M that are typically observed for limiting cases of weakly and tightly bound ligands, respectively, correspond to the binding free energy (ΔG_{bind}) range of −4 to −12 kcal/mol. Thus, to obtain practically meaningful predictions, the accuracy of the calculated ΔG_{bind} should be better than 2 kcal/mol. However, this level of accuracy is very difficult to obtain by FEP calculations that use a direct approach whereby the ligand is forced to diffuse into an active site.[11] The computational accuracy can be significantly increased using a thermodynamic cycle that provides ΔG_{bind} as a difference between free energies for annihilating the ligand in the protein and water.[12,13] Magnitudes of these energies are dominated by contributions of charged groups. Therefore, in the case of multiple charged ligands this procedure involves subtraction of two large numbers thus placing extreme demands on computational accuracy.

Since the same charged triphosphate moiety (tail) is present in all nucleoside triphosphates (NTPs), it is tempting to examine whether the tail contributions cancel out when the relative ΔG_{bind} values between different NTPs ($\Delta\Delta G_{bind}$) are evaluated. In fact, the magnitudes of the $\Delta\Delta G_{bind}$ values of deoxyribonucleoside triphosphate (dNTP) substrates to the T7 DNA polymerase, which were calculated using such an assumption, were found to be stable enough to allow the determination of the binding component of the polymerase fidelity.[14] Nevertheless, a more rigorous evaluation of the validity of the supposition of the zero contribution of the tail to $\Delta\Delta G_{bind}$ is needed. In this paper, we start addressing this issue by determining the substituent effects on the relative solvation free energies ($\Delta\Delta G_{solv}$) of nucleosides and nucleoside phosphates.

Interactions of an NTP molecule with its environment determine an important free energy component that is needed for calculating $\Delta\Delta G_{bind}$ by the way of a thermodynamic cycle. Therefore, we focus our initial analysis on the properties of ΔG_{solv} of nucleoside phosphates in aqueous solution. We start by assessing the statistical errors of the brute force FEP calculations that treat explicitly the tail contributions of dNTPs to ΔG_{solv}. By analyzing a series of 8- and 64-ns FEP calculations involving different trajectories we show that this error is significant even for dNTPs whose total charge is decreased from −4 to −2 by a tightly bound Mg^{2+} ion.

* Author to whom correspondence should be addressed. E-mail: jfloria@luc.edu.

[†] Permanent address: National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia.

[‡] Permanent address: Department of Biochemistry, Faculty of Science, Charles University in Prague, Prague, Czech Republic.

10.1021/jp056623m CCC: $33.50
Published on Web 06/02/2006

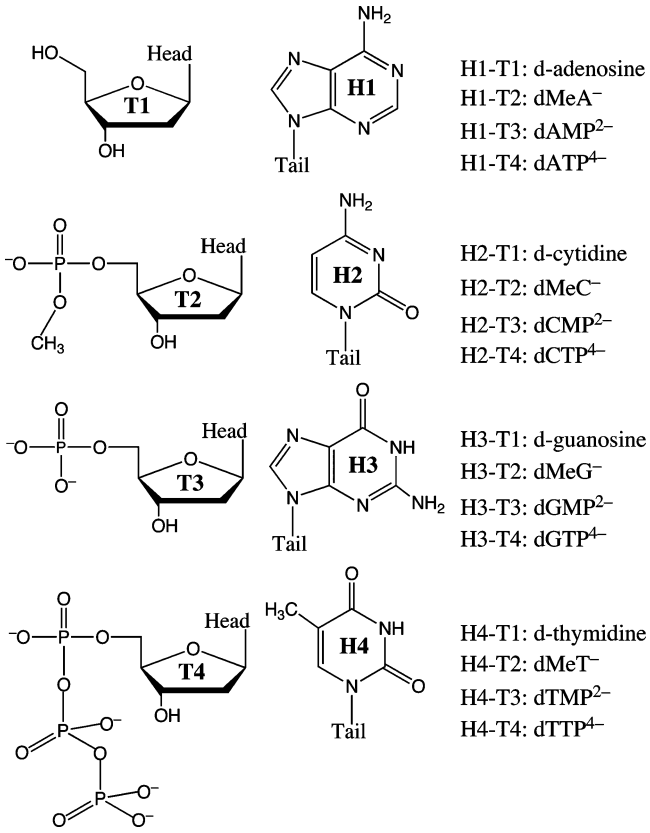

Figure 1. Chemical structures and abbreviations of the solutes studied in the present work. The solutes were generated by combining four nucleic acid bases (H1−H4) with four different substituents (T1−T4).

Next, we focus on the electrostatic contribution of the nucleobase moiety (head) to the $\Delta\Delta G_{solv}$ values of nucleoside phosphates (Figure 1) and examine the additivity of the decomposition of this free energy ($\Delta G_w^{H,ES}$) into head−head, head−solvent, head−counterions, and head−substituent (tail) contributions. These simulations address the question of how much $\Delta G_w^{H,ES}$ is affected by the structure and the charge of the tail. Answering this question should reveal the importance of the coupling between the head and the tail portions of the ligand in aqueous solution from the viewpoint of $\Delta\Delta G_{solv}$. A particularly interesting aspect of such coupling is the actual magnitude of the through-solvent coupling. The mechanism of this indirect coupling involves changes of the orientational polarizability of water in the vicinity of the nucleobase moiety due to interactions with charges located on the tail. As a result, the relative hydrophilicity of the four unsubstituted nucleobases and nucleobase moieties present in DNA or dNTPs could vary. Establishing the absence of such indirect solvation effects using microscopic simulations would be important for justification of the computational approaches utilizing dielectric-continuum solvation models for binding calculations.[15−17]

2. Methods

2.1. Free Energy Decomposition using FEP. The starting point is the standard expression for calculations of free energy differences by FEP[18]

$$\Delta G_{i\to i+1} = -\beta^{-1}\ln\langle\exp(-\beta\Delta E)\rangle_i \quad (1)$$

where $\Delta G_{i\to i+1}$ denotes the free energy difference between states i and $i+1$. $\Delta E = E_{i+1} - E_i$ represents the corresponding potential energy difference. $\beta^{-1} = k_B T$, where k_B stands for the Boltzmann constant and T is the thermodynamic temperature. The notation $\langle\cdots\rangle_i$ indicates averaging over the ensemble of configurations generated by a molecular dynamics (MD) simulation on the potential energy surface of state i. Equation 1 rigorously gives excess Helmholz free energy, but the difference between Helmholz and Gibbs free energy is negligible in condensed systems. If we expand the exponential function into a Taylor series, we obtain

$$\Delta G_{i\to i+1} = -\beta^{-1}\ln\left(1-\beta\langle\Delta E\rangle_i + \frac{\beta^2}{2}\langle\Delta E^2\rangle_i - \cdots\right) \quad (2)$$

The series expansion of the logarithmic function gives

$$\Delta G_{i\to i+1} = -\beta^{-1}\left[\left(-\beta\langle\Delta E\rangle_i + \frac{\beta^2}{2}\langle\Delta E^2\rangle_i - \cdots\right) - \frac{1}{2}(\beta^2\langle\Delta E\rangle_i^{\,2} + \cdots) + \cdots\right] \quad (3)$$

After some rearranging we obtain

$$\Delta G_{i\to i+1} = \langle\Delta E\rangle_i + \frac{\beta}{2}[\langle\Delta E\rangle_i^{\,2} - \langle\Delta E^2\rangle_i] + \cdots \quad (4)$$

If we consider, without loss of generality, the potential energy difference to arise from two components

$$\Delta E = \Delta E_1 + \Delta E_2 \quad (5)$$

we can define the free energy difference originating from the first component (and from the second component) using eqs 1−4 as

$$\Delta G^1_{i\to i+1} = -\beta^{-1}\ln\langle\exp(-\beta\Delta E_1)\rangle_i = \langle\Delta E_1\rangle_i + \frac{\beta}{2}[\langle\Delta E_1\rangle_i^{\,2} - \langle\Delta E_1^{\,2}\rangle_i] + \cdots \quad (6a)$$

$$\Delta G^2_{i\to i+1} = -\beta^{-1}\ln\langle\exp(-\beta\Delta E_2)\rangle_i = \langle\Delta E_2\rangle_i + \frac{\beta}{2}[\langle\Delta E_2\rangle_i^{\,2} - \langle\Delta E_2^{\,2}\rangle_i] + \cdots \quad (6b)$$

By combining eq 4 and 5 we obtain

$$\Delta G_{i\to i+1} = \langle\Delta E_1\rangle_i + \langle\Delta E_2\rangle_i + \frac{\beta}{2}[\langle\Delta E_1\rangle_i^{\,2} + 2\langle\Delta E_1\rangle_i\langle\Delta E_2\rangle_i + \langle\Delta E_2\rangle_i^{\,2} - \langle\Delta E_1^{\,2}\rangle_i - 2\langle\Delta E_1\Delta E_2\rangle_i - \langle\Delta E_2^{\,2}\rangle_i] + \cdots \quad (7)$$

where the free energy difference can be attributed to individual components as

$$\Delta G_{i\to i+1} = \Delta G^1_{i\to i+1} + \Delta G^2_{i\to i+1} + \beta[\langle\Delta E_1\rangle_i\langle\Delta E_2\rangle_i - \langle\Delta E_1\Delta E_2\rangle_i] + \cdots \quad (8)$$

It can be seen that the nonadditivity of the free energy $\Delta G^{NAD}_{i\to i+1}$ originates from the mixing potential energy terms of second and higher order, which increasingly loose importance as $|\beta\Delta E|$ becomes smaller and smaller

$$\Delta G^{NAD}_{i\to i+1} = \Delta G_{i\to i+1} - \Delta G^1_{i\to i+1} - \Delta G^2_{i\to i+1} = \beta[\langle\Delta E_1\rangle_i\langle\Delta E_2\rangle_i - \langle\Delta E_1\Delta E_2\rangle_i] + \cdots \quad (9)$$

Usually, the overall change from the initial ($i = 1$) to the final ($i = N + 1$) state is divided into a series of N

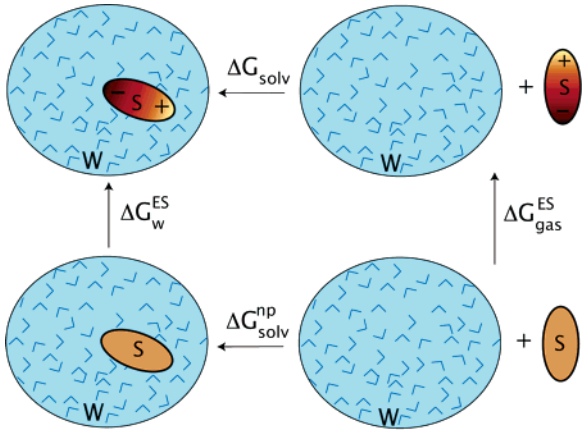

Figure 2. Thermodynamic cycle for the calculation of the solvation free energy in aqueous solution (ΔG_{solv}) using the free energies for charging a solute (S) in the gas phase and water (w) and the solvation free energy of the uncharged solute (ΔG^{np}_{solv}).

small steps (windows)[4]

$$\Delta G = \sum_{i=1}^{N} \Delta G_{i \rightarrow i+1} \tag{10}$$

The overall nonadditivity of the free energy ΔG^{NAD} can be therefore expressed as

$$\Delta G^{NAD} = \sum_{i=1}^{N} \Delta G^{NAD}_{i \rightarrow i+1} \tag{11}$$

If we want to decrease the overall nonadditivity, then we only have to increase the number of FEP windows, which will consequently lower the value of $|\beta \Delta E|$ for each individual step.

The change in eq 10 occurs along a perturbation path that gradually morphs the initial state into the final state. As the free energy is a state function its difference depends only on the choice of initial and final states. The free energy difference is independent of the selection of the perturbation path. However, free energy components are not state functions as only their sum is a state function. Hence, their differences are path-dependent, and additional care must be taken. This fact underscores the demand for a unique perturbation path, which retains as much as possible the natural behavior of the system throughout the simulation.

2.2. Decomposition of the Solvation Free Energy. The solvation free energy, ΔG_{solv}, is defined as the free energy for the transfer of a 1 M solute in the gas phase to solution to form a 1 M solution.[19] For neutral solutes in aqueous solution, ΔG_{solv} values can be measured directly from their partial pressures above the water surface.[20] For charged solutes, ΔG_{solv} values must be determined indirectly using a thermodynamic cycle.[21] The use of a thermodynamic cycle is also necessary for rigorous calculations of ΔG_{solv} by thermodynamic integration or FEP approaches. Thermodynamic cycles that involve uncharging of the solute and shrinking its size to nothing have been applied.[22] Since these algorithms are numerically quite unstable for larger solutes, it is advantageous to use FEP for the thermodynamic cycle of Figure 2 that yields ΔG_{solv} as

$$\Delta G_{solv} = \Delta G^{ES}_{w} - \Delta G^{ES}_{gas} + \Delta G^{np}_{solv} \tag{12}$$

where ΔG^{ES}_{w} and ΔG^{ES}_{gas} are free energies for charging a solute (S) in water and the gas phase, respectively, and the solvation free energy of the uncharged solute, ΔG^{np}_{solv}, is often approximated as[23−25]

$$\Delta G^{np}_{solv} = \Delta G^{vdW} + \Delta G^{cav} \tag{13}$$

where ΔG^{vdW} represents the free energy contribution of van der Waals interactions between the solute and the solvent and ΔG^{cav} is the free energy required to form the solute cavity within the solvent. The ΔG^{vdW} and ΔG^{cav} terms are of a different sign, therefore together yielding a relatively small constant (in comparison with the electrostatic component).

Let us now focus on a group of structurally similar solutes that could be divided into two regions. The first region (referred to as tail or T) is the same for all solutes. The second region (referred to as head or H) is unique for each solute. The charging process may be accomplished by first charging the tail in the presence of the uncharged head, leading to free energy change $\Delta G^{T,ES}$. The second step, which is associated with free energy change $\Delta G^{H,ES}$, is for charging the head region in the presence of charged tail. Thus, we can write

$$\Delta G^{ES} = \Delta G^{T,ES} + \Delta G^{H,ES} \tag{14}$$

Supposing the additivity of the free energy contributions, the free energy for charging the head in water containing Na^+ ions, $\Delta G^{H,ES}_{w}$, originates from the interactions within the head itself (HH) and interactions between the head and its surroundings (Hsurr)

$$\Delta G^{H,ES}_{w} = \Delta G^{HH,ES}_{w} + \Delta G^{Hsurr,ES}_{w} \tag{15}$$

where the interactions with the surroundings can be furtrher decomposed into head−tail (HT), head−water (HW), and head−sodium (HNa) interactions

$$\Delta G^{Hsurr,ES}_{w} = \Delta G^{HT,ES}_{w} + \Delta G^{HW,ES}_{w} + \Delta G^{HNa,ES}_{w} \tag{16}$$

Similarly, free energy of charging the head in the gas phase can be decomposed as

$$\Delta G^{H,ES}_{gas} = \Delta G^{HH,ES}_{gas} + \Delta G^{HT,ES}_{gas} \tag{17}$$

If the head is rigid, then it should sample very similar configurations while being charged in water and in gas, i.e.,

$$\Delta G^{HH,ES}_{gas} = \Delta G^{HH,ES}_{w} \tag{18}$$

Using this approximation and eqs 12−17 gives

$$\Delta G_{solv} = \Delta G^{Hsurr,ES}_{w} - \Delta G^{HT,ES}_{gas} + \Delta G^{T,ES}_{w} - \Delta G^{T,ES}_{gas} + \Delta G^{np}_{solv} \tag{19}$$

Since the tail is the same for all solutes and the change in the nonpolar term, $\Delta \Delta G^{np}_{solv}$, is negligible for solutes with similar heads, the relative solvation free energy can be approximated as

$$\Delta \Delta G_{solv} = \Delta \Delta G^{Hsurr,ES}_{w} - \Delta \Delta G^{HT,ES}_{gas} \tag{20}$$

This approach can be generalized to determine the electrostatic component of the relative protein−ligand binding free energy, $\Delta \Delta G^{ES}_{bind}$, for ligands with the same tail as

$$\Delta \Delta G^{ES}_{bind} = \Delta \Delta G^{Hsurr,ES}_{p} - \Delta \Delta G^{Hsurr,ES}_{w} \tag{21}$$

where the subscript p denotes the protein environment.

TABLE 1: Calculated Solvation Free Energy (kcal/mol) in Aqueous Solution for 2′-Deoxynucleoside Triphosphates Complexed with Mg^{2+}

system composition[a]	ΔG_w^{ES}		ΔG_{solv}^{np} [d]	ΔG_{gas}^{ES} [e]	ΔG_{solv}[f]
	16×4 ns [b]	2×32 ns [c]			
$dATP^{4-}\cdot Mg^{2+} + 2Na^+$	-1206.5 ± 3.3	-1206.0 ± 1.5	7.2	-887.6 ± 0.3	-311.5 ± 1.7
$dCTP^{4-}\cdot Mg^{2+} + 2Na^+$	-1291.3 ± 2.3	-1291.7 ± 1.5	7.0	-970.6 ± 0.1	-314.1 ± 1.6
$dGTP^{4-}\cdot Mg^{2+} + 2Na^+$	-1255.9 ± 4.5	-1256.5 ± 2.2	7.9	-931.6 ± 0.1	-317.0 ± 2.3
$dTTP^{4-}\cdot Mg^{2+} + 2Na^+$	-1169.4 ± 2.6	-1169.4 ± 1.8	7.4	-854.4 ± 0.1	-307.5 ± 1.9

[a] The first coordination shell of Mg^{2+} contained one oxygen from each phosphate group and three water molecules. These ligands were not exchanged during the simulation. Sodium ions were considered to be a part of the solvent. [b] Free energy perturbation calculations for charging the solute in water. Average and standard deviations that were determined from eight forward and eight reverse 4-ns trajectories are reported. [c] A single 64-ns trajectory that consisted of 32-ns forward and 32-ns reverse simulations. [d] Calculated using the Langevin dipoles solvation model implemented in the Chemsol 2.1 program.[53] [e] A single 200-ns trajectory that consisted of 100-ns forward and 100-ns reverse simulations. [f] Calculated using eq 12 and ΔG_w^{ES} results and standard deviations obtained from the two 32-ns simulations.

2.3. Computational Methods. The configurational ensembles for the evaluation of free energies were generated from MD trajectories using the Amber force field[26] implemented in the program Q.[27] Force field parameters for deoxynucleoside triphosphates and Mg^{2+} ions were reported previously.[28,29] The charges used for the base and sugar moieties of methyl esters of nucleoside monophosphates were identical to those used for the corresponding triphosphates. The charges in a.u. of the methyl phosphate moiety were chosen as follows: P, 1.17; O(ionic), −0.7761; O(bridging), −0.48; C(methyl), 0.13; H(methyl), 0.012; O5′, −0.4954. The charges of the base moieties were constant throughout all nucleoside phosphate series; i.e., no charge transfer to or from the base (other than the charge transfer present in nucleosides of the original Amber 94 force field) was introduced.

The simulated solute molecules were immersed in a 24 Å radius sphere of TIP3P water molecules subjected to the surface-constraint all-atom solvent (SCAAS)-type boundary conditions.[27,30,31] These constraints were designed to mimic infinite aqueous solution. The solute was positioned near the center of this sphere. A restraining harmonic potential with the force constant of 50 kcal/(mol Å^2) was applied on the position of the C1′ atom of the solute. If sodium atoms were added, then their positions were restrained by a 50 kcal/(mol Å^2) flat-bottom harmonic potential that was zero for the distances less than 20 Å from the center of the simulation sphere. These potentials were applied to prevent the diffusion of the solute and sodium ions toward the edge of the simulation sphere.

The simulations of $dNTP^{4-}$ molecules complexed with a single Mg^{2+} ion were initiated using the geometry of dNTP from the crystal structure of polymerase β ternary complex (PDB code 1BPY),[32] in which the structural Mg^{2+} ion is ligated by three oxygen atoms of α-, β-, and γ-phosphates of dNTP. The distances to these oxygen atoms were restrained by a weak (5 kcal/(mol Å^2)) flat-bottom harmonic potential that was zero at distances shorter than 3.7 Å. This restraining potential was applied in both solution and gas-phase simulations. In addition, restraining harmonic potentials were applied on the C1′, C2′, and O4′ atoms of the solute in the gas-phase simulations to prevent the "flying ice cube" artifact[54] of Berendsen's thermostat.[55] Two Na^+ ions were added into the aqueous solution to obtain electroneutrality of the simulated system.

The simulated systems were equilibrated by series of MD simulations that included 25 ps dynamics at a temperature of 5 K, gradual heating to 298 K in 50 K increments over total period of 50 ps, and a 20 ps trajectory with the step 1 fs at 298 K. The free energy differences ($\Delta G°$) were calculated as an arithmetic average of the "forward" and "reverse" FEP calculations[33,34] at 298 K. Charges of the atoms of the head portion of the solute (see below) or the whole solute were simultaneously changed to zero in the forward simulations and from zero to their original values in the reverse simulations. Bonding and van der Waals parameters were kept constant in all trajectories. Both forward and reverse trajectories consisted of a series of 51 independent single-topology FEP calculations (windows) that differed in the value of the coupling parameter λ. The number of windows was increased to 101 for calculations of the dNTP·Mg complexes presented in Table 1. λ was varied from 1 to 0 and from 0 to 1 in the forward and reverse calculations, respectively. Forward trajectories of total lengths of 2 ns were followed by a 40 ps simulation at zero solute charges, which served as a starting point of the reverse trajectories with 2 ns of total simulation time. The simulation time was doubled for calculations of dNTP·Mg complexes. The integration step was 2 fs. The SHAKE algorithm was applied to constrain the bond lengths of hydrogen atoms in the solute and solvent molecules. The energy of the system was sampled every 10th step. The first 10 energies (200 fs) were discarded for each window before the free energy change was evaluated. All interactions involving mutated atoms were explicitly evaluated, whereas nonbonded interactions between other atoms were subjected to a 10 Å cutoff. The local-reaction field (LRF) method[35] was used to treat long-range electrostatic interactions for distances beyond the 10 Å cutoff.

3. Results and Discussion

3.1. Solvation Free Energies of dNTPs. As shown by crystal structures[36] and pre-steady-state kinetic experiments,[37] dNTP substrates bind to DNA polymerases as Mg^{2+} complexes, in which the Mg^{2+} cation forms ionic bonds with one nonbridging oxygen atom on each phosphate and oxygen atoms of three water molecules in an octahedral arrangement. The complexation of the triphosphate moiety with Mg^{2+} has also been observed for ATP binding by F_1-ATPase.[38,39] Thus, dNTPs complexed with Mg^{2+} appear to be very biologically relevant systems that possess the computational advantage of having less negative total charge than $dNTP^{4-}$.

The average lifetime of the first water coordination shell of Mg^{2+} in aqueous solution is about 1 μs.[40] Due to the facts that phosphate ligands are covalently linked and phosphate oxygens have higher charges than water oxygens, triphosphates complexed with Mg^{2+} are expected to have significantly longer lifetimes. Therefore, the phosphate ligand exchange is unlikely to occur during our simulations, which last several nanoseconds. Indeed, Mg^{2+} spontaneously retains its initial coordination during the entire trajectory run on the full and nearly full solute charges, while the flat-bottom constraining potentials described in the Methods section assured that Mg^{2+} retained the same ligands at the low- and zero-charge simulations, which are

essential parts of the FEP protocol. Stable ionic bonds to the metal made the triphosphate parts of the dNTPs relatively rigid. In contrast, the nucleoside moiety sampled both anti and syn conformations as well as a variety of sugar puckers. The two sodium atoms, which were added to the solvent to keep the entire simulated system electroneutral, showed no clear structural preferences in coordinating with either the phosphate, the nucleobase moieties, or the water molecules.

The calculated magnitudes of ΔG_{solv} for Mg^{2+} complexes of dNTPs in aqueous solution are presented in Table 1. In addition, Table 1 shows free energy differences calculated for each leg of the thermodynamic cycle of Figure 2. The calculated ΔG_{solv} values are narrowly spread from -317.0 to -307.5 kcal/mol because the total ΔG_{solv} of $(dNTP{\cdot}Mg)^{2-}$ is dominated by the charged tail region, which is identical in all dNTPs. To assess the stability of the calculated averages, ΔG_w^{ES} was determined using two independent sets of simulations. The statistical sample of 16 4-ns simulations has a larger statistical error than the sample consisting of 2 32-ns simulations. The 32-ns sample was actually used to determine magnitudes of ΔG_{solv} and their estimated errors presented in Table 1. However, both samples provide essentially identical averages.

The statistical errors of the calculated ΔG_{solv} are small enough to conclude that the affinity of dNTPs for aqueous solution decreases in the order dGTP > dCTP > dATP > dTTP. The calculated order of ΔG_{solv} agrees with the order of free energies measured for the transfer of methylated nucleobases between water and chloroform[41] or the transfer of tetrahydrofuryl-substituted nucleobases between water and cyclohexane.[42]

Magnitudes of ΔG_{solv} for nucleic acid bases in aqueous solution were previously determined using thermodynamic integration[22] and FEP[43] simulations. Among the large body of data on this subject available in the literature (see, for example, ref 44 and references therein), these two studies employed the force field that was either identical or very similar to the force field used in our calculations. Both of these studies found ΔG_{solv} to decrease (in absolute value) in the order Gua > Cyt > Thy > Ade. Except for the reversal of the positions of the two least hydrophilic bases, this order is the same as the order calculated here for dNTPs. Also the ΔG_{solv} difference of 9.5 kcal/mol between the most and the least hydrophilic dNTP (Table 1) agrees well with the corresponding span of about 10 kcal/mol obtained for nucleobases.[22,43]

3.2. Substituent Effects. To examine solvation effects of the charged tail we carried out a systematic study of the compounds shown in Figure 1. Since the magnitude of the free energy for charging the head portion of nucleoside triphosphates in water, $\Delta G_w^{Hsurr,ES}$, is important for understanding the relative ΔG_{solv} of substituted nucleobases (eq 20), we focused on the analysis of the substituent effects on this term (Table 2). We found that, regardless of the tail structure, the magnitudes of $\Delta G_w^{Hsurr,ES}$ follow the same order as ΔG_{solv} for nucleic acid bases (see above). This order is retained even for dNTPs.

In addition, $\Delta G_w^{Hsurr,ES}$ of each head was found to be practically unaffected by the structure of the tail. The largest but still relatively small (~1 kcal/mol) substituent effect on $\Delta G_w^{Hsurr,ES}$ was obtained upon going from a base attached to deoxyribonucleoside to the same base in deoxyribonucleoside methyl phosphate in the absence of counterions. This effect may be caused by the negatively charged phosphate substituent that reorients water dipoles around the base toward its own charge and thus decreases the ability of these molecules to form stabilizing electrostatic interactions with the base. Consequently, the charging free energies of bases with charged substituents become smaller (in absolute value).

TABLE 2: Substituent and Ionic Strength Effects on the Charging Free Energy of Nucleobase Moieties in Aqueous Solution[a]

	in kcal/mol				
system composition[b]	$\Delta G_w^{Hsurr,ES}$ [c]	$\Delta G_w^{HT,ES}$	$\Delta G_w^{HW,ES}$	$\Delta G_w^{HNa,ES}$	ΔG^{NAD} [d]
d-adenosine	−13.3±0.1	−3.7	−9.6		
$dMeA^-$	−11.8±0.2	0.1	−11.9		
$dMeA^- + Na^+$	−13.5±0.2	−1.4	−9.3	−2.8	0.001
$dAMP^{2-} + 2Na^+$	−13.4±0.1	2.4	−9.5	−6.3	0.005
$dATP^{4-}{\cdot}Mg^{2+} + 2Na^+$	−13.6±0.1	0.4	−8.6	−5.4	−0.001
$dATP^{4-} + 4Na^+$	−13.8±0.2	7.6	−10.3	−11.1	0.001
d-cytidine	−18.7±0.1	−2.4	−16.3		
$dMeC^-$	−17.9±0.1	−4.4	−13.5		
$dMeC^- + Na^+$	−19.2±0.3	−5.8	−11.5	−1.9	0.002
$dCMP^{2-} + 2Na^+$	−18.7±0.1	−4.9	−13.4	−0.4	−0.009
$dCTP^{4-}{\cdot}Mg^{2+} + 2Na^+$	−19.6±0.1	−8.6	−10.7	−0.3	−0.002
$dCTP^{4-} + 4Na^+$	−19.1±0.1	−3.9	−12.1	−3.0	−0.015
d-guanosine	−21.0±0.1	−0.8	−20.2		
$dMeG^-$	−20.0±0.3	1.4	−21.4		
$dMeG^- + Na^+$	−21.1±0.1	1.3	−19.1	−3.3	0.003
$dGMP^{2-} + 2Na^+$	−20.4±0.2	4.9	−20.4	−4.8	0.002
$dGTP^{4-}{\cdot}Mg^{2+} + 2Na^+$	−21.3±0.1	3.1	−19.3	−5.1	−0.010
$dGTP^{4-} + 4Na^+$	−21.6±0.4	9.6	−19.3	−11.9	−0.005
d-thymidine	−16.1±0.1	−5.8	−10.3		
$dMeT^-$	−14.9±0.1	−8.3	−6.6		
$dMeT^- + Na^+$	−16.4±0.1	−6.1	−6.6	−3.6	−0.001
$dTMP^{2-} + 2Na^+$	−16.5±0.1	−4.8	−8.3	−3.4	0.006
$dTTP^{4-}{\cdot}Mg^{2+} + 2Na^+$	−16.9±0.1	−7.4	−6.7	−2.8	0.004
$dTTP^{4-} + 4Na^+$	−16.6±0.1	0.8	−7.7	−9.8	0.014

[a] The calculated magnitudes of $\Delta G_w^{HH,ES}$ amounted to −85,4, −166.0, −127.1, and −45.6 kcal/mol for Ade, Cyt, Gua, and Thy, respectively, and were unaffected (within 0.1 kcal/mol) by the tail structure. [b] See Figure 1. [c] Equation 16. Reported error values were calculated as one-half of the difference between the free energies calculated using the independent forward and reverse FEP trajectories. [d] Nonadditivity of the free energy, $\Delta G^{NAD} = \Delta G_w^{H,ES} - (\Delta G_w^{HH+HT,ES} + \Delta G_w^{HNa,ES} + \Delta G_w^{HW,ES})$, where HH, HT, HNa, and HW denote head−head, head−tail, head−sodium, and head−water components of the total charging free energy of the head, $\Delta G_w^{H,ES}$ (see also eqs 9 and 11).

However, this orientational polarization effect completely disappears when the total charge of the simulated system is kept at the neutral value by adding sodium counterions. In this case, we observed a nearly complete compensation of the changes in the base−water and base−sodium interactions by the base−tail interactions. This compensation leads to a charging free energy of the base moiety that is independent of the charge or geometry of the tail. This result is consistent with the empirical observation that free energies of solvation behave in most cases as additive functions of the substituent groups that are present in solutes.[45−47]

3.3. Additivity of ΔG Contributions. The total charging free energies of the head, $\Delta G_w^{H,ES}$, reflect the sum of interactions of the atomic charges of the base with other atoms on the same base, with its deoxyribophosphate substituent, with the surrounding water molecules and Na^+ ions (eqs 15 and 16). The free energies evaluated by considering each of these interactions separately are presented in Table 2. The magnitude of the difference of the total free energy for the charging process and the sum of its contributions allows us to examine the additivity of the free energy components. For all studied systems, the calculated nonadditivity is smaller than 0.02 kcal/mol (Table 2). These results demonstrate the additivity of the free energy components along a single FEP trajectory, which is implied by eq 9.

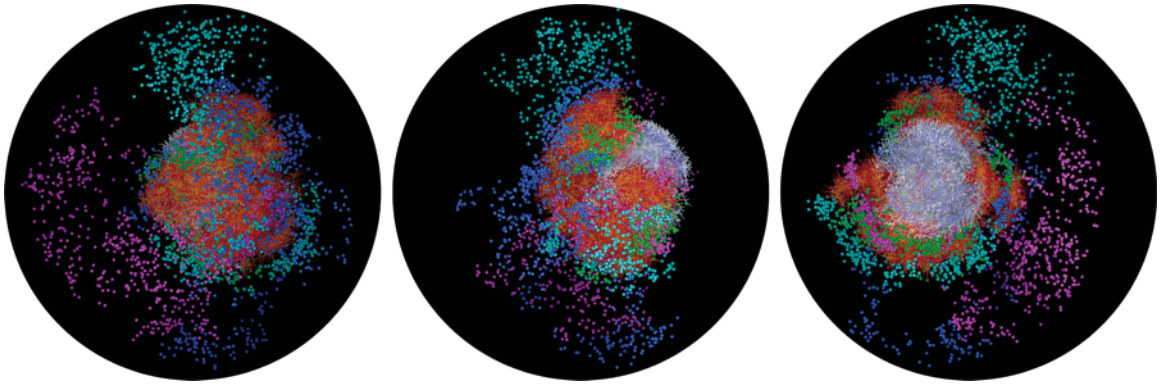

Figure 3. Left picture: Superposition of 1000 snapshots distributed uniformly along a 2-ns MD trajectory for the FEP calculation for charging the adenine moiety of $dATP^{4-}$ in a water droplet (black) containing four Na^+ ions represented by blue, green, purple, and turquoise dots. The thick bundle of lines in the center shows the dATP molecule, which consists of triphosphate (orange) and nucleoside (light blue and white) moieties. The middle and right pictures represent the same system as the one depicted on the left but rotated by 120° and 240° around the vertical axis, respectively.

To understand the origin of such a high degree of additivity we need to consider the second- and higher-order terms of the Taylor expansion of the free energy as a function of its potential energy components (eq 4). These terms will become zero if the free energy surfaces of the states i and $i + 1$, which differ in the magnitude of the charge, can be approximated by parabolas of equal curvature. This condition is satisfied in condensed systems, as shown by the success of the linear response (LRA) [28,48] and linear interaction energy (LIE) [49−51] approximations, which both rely on its validity. Thus, it appears that the additivity of eq 9 is generally valid for nucleoside phosphates in aqueous solution provided a single FEP pathway subdivided into many windows is used in the calculations. Of course, different pathways will generally yield different partitioning into the components. Nevertheless, the present results show that, if any of these trajectories can be singled out as the most relevant for the studied property, then the additivity will hold.

3.4. Head−Tail Interactions in the Gas Phase. Since $\Delta G_w^{Hsurr,ES}$ is independent of the nature of the tail substituent (Table 2), this term is unlikely to cause the reversal of the relative solvation free energy of adenine and thymine upon their substitution by deoxyribonucleoside triphosphates complexed by Mg^{2+} ions. Indeed, the order of $\Delta G_w^{Hsurr,ES}$ is the same as the order of ΔG_{solv} for nucleic acid bases (see above). Thus, it is of interest to check the behavior of the head−tail interactions in the gas phase, which constitute the second term in the equation for $\Delta\Delta G_{solv}$ (eq 20).

The magnitudes of $\Delta G_{gas}^{HT,ES}$ calculated by a 100-ns FEP calculation for $dATP^{4-}\cdot Mg^{2+}$, $dCTP^{4-}\cdot Mg^{2+}$, $dGTP^{4-}\cdot Mg^{2+}$, and $dTTP^{4-}\cdot Mg^{2+}$ amount to 1.9 ± 0.8, -3.5 ± 0.2, -0.5 ± 0.5, and -0.5 ± 0.1 kcal/mol, respectively. These results mean that the interaction with its own tail makes the adenine moiety more hydrophilic, whereas the opposite is true for the cytosine, thymine, and guanine moieties. This gas-phase tail effect explains the reversal of the order of ΔG_{solv} calculated by full FEP for dATP and dTTP (Table 1) from the order for the corresponding nucleic acid bases.[22]

However, we would like to caution that eq 20, apart from being less rigorous, may not represent a more efficient way for calculating $\Delta\Delta G_{solv}$ values of charged solutes than the straightforward FEP calculation utilizing the thermodynamic cycle of Figure 1. This is because the FEP calculation of the $\Delta G_{gas}^{HT,ES}$ term is associated with significantly larger statistical error than the calculation of the ΔG_{gas}^{ES} term. In fact, both of these calculations are rather demanding because the large electrostatic energies involved make it possible for the tail to be locked in a single conformation that is separated from other low-energy conformations by a large energetic barrier. The tail is fully charged during the calculation of $\Delta G_{gas}^{HT,ES}$ whereas it is fully charged only in the last FEP window in the calculation of ΔG_{gas}^{ES}. Thus, significantly larger barriers need to be overcome during the calculation of $\Delta G_{gas}^{HT,ES}$.

3.5. Distribution of Na^+ Ions. The $\Delta G_w^{Hsurr,ES}$ invariant can be calculated with the least effort by embedding the base in a nucleoside because calculations with charged tails and/or counterions require longer simulation times to fully converge. For example, the phosphate tail in the $dNTP^{4-}\cdot 4Na^+$ system has two or three sodium atoms in its immediate vicinity most of the time, but the identities of these ions change during the simulation (Figures 1S−4S in the Supporting Information and Figure 3). The remaining ions diffuse through the simulation sphere and at times engage in direct interactions with the base moiety. Interestingly, while the ions diffuse uniformly in the radial direction with respect to the solute, the C1′ atom of which is constrained in the center of the simulation sphere, they remain most of the time in the hemisphere containing the charged triphosphate (Figure 3) or monophosphate (Figure 5S in the Supporting Information) substituents. Proper averaging of the resulting configuration space is computationally demanding, but the stability of the $\Delta G_w^{Hsurr,ES}$ free energy calculated from independent forward and reverse trajectories, in which Na^+ ions occupied different positions, indicates that our simulation length was sufficient for obtaining properly converged results.

4. Conclusions

Quantitative evaluation of solvation free energies of nucleoside phosphates is important for understanding the energetics of important biological reactions, for example, the hydrolysis of ATP. In this paper, we determined solvation free energies of magnesium complexes of deoxyribonucleoside triphosphates. For each base, we extracted a contribution of the interaction of the base moiety with its surroundings and showed that this contribution is constant with respect to a wide range of its substituents. This result indicates that the coupling between the solvation of the phosphate and the base regions of the solute is negligible.

Using the FEP formulas we established that every dissection of the free energy into its components possesses an inherent error (termed the nonadditivity error) arising from coupling between corresponding potential energy components. This effect is due to the appearance of mixing terms in the equation for the free energy difference. Fortunately, these mixing terms

increasingly loose importance as the change in the potential energy decreases. Therefore, it is possible to decrease the nonadditivity error to an arbitrary value just by increasing the number of FEP windows. In particular, for the solutes studied here, a solvation free energy nonadditivity of less than 0.02 kcal/mol was achieved by using only 51 windows. However, one has to be aware that the free energy components are not state functions and should consequently be calculated on the most natural trajectory possible. The free energy decomposition approach presented in this paper can be combined with the recent suggestion of such a trajectory[52] to form a powerful theoretical tool for rational drug design.

Acknowledgment. This work was supported by the National Institutes of Health (Grant No. 1U19CA105010). The authors thank Janez Mavri from the National Institute of Chemistry, Slovenia, for helpful discussions.

Supporting Information Available: Distances of each of the four Na^+ ions from the β-phosphorus atom of the phosphate of $dATP^{4-}$, $dCTP^{4-}$, $dGTP^{4-}$, and $dTTP^{4-}$ and a superposition of 1000 snapshots distributed uniformly along a 2-ns MD trajectory for the FEP calculation for charging the guanine moiety of $dGMP^{2-}$ in a water droplet. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Amidon, G. L.; Pearlman, R. S.; Anik, S. T. *J. Theor. Biol.* **1979**, *77*, 161.
(2) Warshel, A. *Acc. Chem. Res.* **1981**, *14*, 284.
(3) Bohm, H. J.; Klebe, G. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2589.
(4) Leach, A. R. *Molecular Modelling. Principles and Applications*; Prentice Hall: Harlow, U. K., 2001.
(5) Muegge, I.; Tao, H.; Warshel, A. *Protein Eng.* **1998**, *10*, 1363.
(6) Boresch, S.; Archontis, G.; Karplus, M. *Proteins* **1994**, *20*, 25.
(7) Boresch, S.; Karplus, M. *J. Mol. Biol.* **1995**, *254*, 801.
(8) Mark, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1994**, *240*, 167.
(9) Smith, P. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1994**, *98*, 13735.
(10) Brady, G. P.; Sharp, K. A. *J. Mol. Biol.* **1995**, *254*, 77.
(11) Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 5716.
(12) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742.
(13) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395..
(14) Florián, J.; Goodman, M. F.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6819.
(15) Jorgensen, W. L. *Science* **2004**, *303*, 1813.
(16) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430.
(17) Mavri, J.; Hadzi, D. *J. Mol. Struct. (THEOCHEM)* **2001**, *540*, 251.
(18) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.
(19) Ben-Naim, A. *J. Phys. Chem.* **1978**, *82*, 792.
(20) Wolfenden, R. *Science* **1983**, *222*, 1087.
(21) Pearson, R. G. *J. Am. Chem. Soc.* **1986**, *108*, 6109.
(22) Miller, J. L.; Kollman, P. A. *J. Phys. Chem.* **1996**, *100*, 8587.
(23) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(24) Luque, F. J.; Bachs, M.; Aleman, C.; Orozco, M. *J. Comput. Chem.* **1996**, *17*, 806.
(25) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.
(26) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(27) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *J. Mol. Graphics Modell.* **1999**, *16*, 213.
(28) Florián, J.; Goodman, M. F.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 5739.
(29) Florián, J.; Goodman, M. F.; Warshel, A. *J. Am. Chem. Soc.* **2003**, 8163.
(30) King, G., Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647.
(31) Sham, Y. Y.; Warshel, A. *J. Chem. Phys.* **1998**, *109*, 7940.
(32) Sawaya, M. R.; Prasad, R.; Wilson, S. H.; Kraut, J.; Pelletier, H. *Biochemistry* **1997**, *36*, 11205.
(33) Valleau, J. P.; Torrie, G. M. A guide to Monte Carlo for statistical mechanics. 2. Byways. In *Statistical Mechanics*; Berne, B. J., Ed.; Modern Theoretical Chemistry 5; Plenum: New York, 1977; p 169.
(34) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1991.
(35) Lee, F. S., Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100.
(36) Pelletier, H.; Sawaya, M. R.; Kumar, A.; Wilson, S. H.; Kraut, J. *Science* **1994**, *264*, 1891.
(37) Bakhtina, M.; Lee, S.; Wang, Y.; Dunlap, C.; Lamarche, B.; Tsai, M. D. *Biochemistry* **2005**, *44*, 5177.
(38) Hochman, Y.; Lanir, A.; Carmeli, C. *FEBS Lett.* **1976**, *61*, 255.
(39) Okimoto, N.; Yamanaka, K.; Ueno, J.; Hata, M.; Hoshino, T.; Tsuda, M. *Biophys. J.* **2001**, *81*, 2786.
(40) Helm, L.; Merbach, A. E. *Coord. Chem. Rev.* **1999**, *187*, 151.
(41) Cullis, P. M.; Wolfenden, R. *Biochemistry* **1981**, *20*, 3024.
(42) Shih, P.; Pedersen, L. G.; Gibbs, P. R.; Wolfenden, R. *J. Mol. Biol.* **1998**, *280*, 421.
(43) Elcock, A. H.; Richards, W. G. *J. Am. Chem. Soc.* **1993**, *115*, 7930.
(44) Bentzien, J.; Florián, J.; Glennon, T. M.; Warshel, A. QM/MM Approaches for Studying Chemical Reactions in Proteins and Solution. In *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M. A., Eds.; ACS Symposium Series 712; American Chemical Society: Washington, DC, 1998; p 16.
(45) Leo, A.; Hantsch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525.
(46) Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1975**, *40*, 292.
(47) Wolfenden, R.; Liang, Y.-L.; Matthews, M.; Williams, R. *J. Am. Chem. Soc.* **1987**, *109*, 9.
(48) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Protein Eng.* **1992**, *5*, 215.
(49) Åqvist, J.; Medina, C.; Samuelson, J. E. *Protein Eng.* **1994**, *7*, 385.
(50) Åqvist, J.; Hansson, T. *J. Phys. Chem.* **1996**, *100*, 9512.
(51) Jones-Hertzog, D. K.; Jorgensen, W. L. *J. Med. Chem.* **1997**, *41*, 55272.
(52) Bren, U.; Martínek, V.; Florián, J. *J. Phys. Chem. B* **2006**, *110*, 10557.
(53) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 10282.
(54) Harvey, S. C.; Tan, R. K.-Z.; Cheatham, T. E. *J. Comput. Chem.* **1998**, *19*, 726.
(55) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.