Constant pH Molecular Dynamics Simulations of Nucleic Acids in Explicit Solvent

Garrett B. Goh,[†] Jennifer L. Knight,[†] and Charles L. Brooks[*,†,‡]

[†]Department of Chemistry and [‡]Biophysics Program, University of Michigan, 930 N. University, Ann Arbor, Michigan 48109, United States

Ⓢ *Supporting Information*

ABSTRACT: The nucleosides of adenine and cytosine have pK_a values of 3.50 and 4.08, respectively, and are assumed to be unprotonated under physiological conditions. However, evidence from recent NMR and X-ray crystallography studies has revealed the prevalence of protonated adenine and cytosine in RNA macromolecules. Such nucleotides with elevated pK_a values may play a role in stabilizing RNA structure and participate in the mechanism of ribozyme catalysis. With the work presented here, we establish the framework and demonstrate the first constant pH MD simulations (CPHMD) for nucleic acids in explicit solvent, in which the protonation state is coupled to the dynamical evolution of the RNA system via λ-dynamics. We adopt the new functional form $\lambda^{N_{exp}}$ for λ that was recently developed for multisite λ-dynamics (MSλD) and demonstrate good sampling characteristics in which rapid and frequent transitions between the protonated and unprotonated states at pH = pK_a are achieved. Our calculated pK_a values of simple nucleotides are in a good agreement with experimentally measured values, with a mean absolute error of 0.24 pK_a units. This work demonstrates that CPHMD can be used as a powerful tool to investigate pH-dependent biological properties of RNA macromolecules.

1. INTRODUCTION

An increasing number of experimental studies in recent years have recognized the role of protonated nucleotides, particularly adenine and cytosine, in RNA structure and function. Experimental pK_a values of the nucleosides have been measured to be 3.50 for adenosine and 4.08 for cytidine, which become protonated at the N1 and N3 atoms, respectively (Figure 1).[1] These findings suggest that adenine and cytosine should typically be unprotonated at physiological pH and that their contributions to RNA structure and function were initially assumed to be minimal. However, recent studies have revealed the prevalence of protonated adenine and cytosine in a wide variety of nucleic acid structures ranging from DNA triple helices to the anticodon stem loop of tRNA,[2−6] indicating that the pK_a value of these residues may be shifted upward to near physiological conditions of pH 7. Protonated bases have been reported to be responsible for a number of noncanonical base pair configurations, which suggests their ability to influence RNA structure.[7,8] A key example is the wobble $A^+ \cdot C$ base pair that has been implicated in stabilizing RNA loop structures[9,10] and in the pH-dependent conformational flexibility of the ribosomal peptidyl transferase center.[11] Wobble $A^+ \cdot C$ base pairs may also form in DNA under biologically relevant conditions,[12] where they can have mutagenic and carcinogenic effects.[13,14] Apart from these structural influences, it has been suggested that the elevated pK_a of these nucleic acids may play a significant role in ribozyme catalysis, as these protonated residues may be involved in general acid−base catalysis, playing an analogous role to histidine residues in proteins.[15−19] Some examples of its role in catalysis include hepatitis δ virus ribozyme[20−24] and hairpin ribozyme,[25−30] where experimental studies have demonstrated that a loss-of-function mutation of key residues that have elevated pK_a values led to a significant drop in catalytic activity.

Despite the copious amount of biochemical and structural studies that are available for these RNA structures, there still remains some ambiguity as to the exact function of these protonated residues. For example, while experimental studies strongly suggest that adenine 38 (A_{38}) participates in the cleavage and ligation reaction that is catalyzed by the hairpin ribozyme, its specific role in the catalytic mechanism and the structural dynamics of the ribozyme remains disputed.[25−30] In such situations, in silico modeling of RNA structures may shed some light on the existing controversy. Walter and co-workers have demonstrated the usefulness of using molecular dynamics (MD) simulations to clarify the role of the protonated A_{38} in the hairpin ribozyme by suggesting that it serves as a general acid in aligning reactive groups and stabilizing the negative charge.[31,32] However, such traditional MD simulations are limited in the sense that prior knowledge obtained from experiment about the identity of key catalytic residue(s) and its protonation state(s) is required. In terms of in silico prediction of pK_a values, Honig and co-workers have recently demonstrated the ability to accurately calculate the pK_a values of nucleotides using numerical solutions to the Poisson−Boltzmann equation from a series of representative static snapshots obtained from RNA NMR structures.[33] While these calculated pK_a values may identify the correct protonation state to be used in a traditional MD simulation, the latter still lacks the ability to incorporate protonation state information on-the-fly. The ability to perform pH-coupled MD is clearly desirable since it would model realistic pH-dependent responses to structural fluctuations and provide mechanistic insight to RNA-catalyzed reactions.
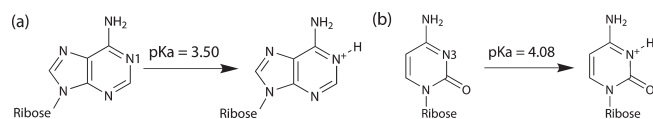
Figure 1. Protonation site of (a) adenosine and (b) cytidine and their respective pK_a values.

In the development of MD simulations, there has been considerable success in calculating pK_a values of protein residues. Warshel and co-workers first demonstrated the feasibility of using microscopic free energy calculations to determine the pK_a values of protein residues.[34−37] Variations of this approach have been developed that couple the protonation state of a titratable residue with the protein conformation; in these strategies, the atomic coordinates and the protonation state itself evolve according to the dynamics of the system. Two distinct classes of implementation for this methodology exist and differ in the manner in which the titration coordinates are treated—either discretely or continuously. The discrete titration variant is typically implemented by combining MD sampling of the atomic coordinates with Monte Carlo (MC) sampling of protonation states. At regular intervals during a typical MD simulation, a MC step is performed to determine the change of the protonation state. Implementation of discrete constant pH MD (CPHMD) in explicit solvent was first reported by Bürgi et al.[38] and Baptista and co-workers,[39] and a number of methodological improvements were made by Baptista and co-workers[40−42] and Stern.[43] Discrete CPHMD has also been implemented with implicit solvent by Dlugosz and Antosiewicz[44,45] and Mongan et al.,[46] with improvements to achieve better sampling by Meng et al.[47] and Williams et al.[48] More recently, Warshel and co-workers developed a more physically realistic form of pH-dependent MD[49] based on the time-dependent MC sampling of the proton transfer process,[50] which uses the empirical valence bond (EVB) framework to simulate a single proton transfer from the protein to a surrounding water molecule.[51] In contrast, in the continuous titration variant of CPHMD, the titration coordinate is propagated continuously between the protonated and unprotonated states. Brooks and co-workers developed CPHMD using implicit solvent[52−54] which utilized the λ dynamics approach[55−57] to treat the titration coordinates. Recent work by Shen and co-workers has improved the prediction accuracy of continuous CPHMD,[58−60] and it has been extended to explicit solvent by Grubmüller and co-workers.[61] Yang and co-workers have also reported using orthogonal space random walk, an enhanced sampling technique compatible with λ-dynamics, that provided accurate pK_a predictions for buried protein residues in explicit solvent simulations.[62] CPHMD has been used by Brooks and co-workers to investigate numerous pH-dependent conformational changes in proteins,[63−65] and other investigators in the field have reported similar successes as well.[66−70] For a more comprehensive overview of CPHMD, we refer our readers to the following review.[71]

In this article, we will adopt the newer functional forms of λ developed by Knight and Brooks that have been implemented in multisite λ-dynamics (MSλD)[72,73] as the basis of a new MSλD-based constant pH MD simulation framework (CPHMD$^{MS\lambda D}$) and parametrize CPHMD$^{MS\lambda D}$ to investigate protonation events of nucleic acids in explicit solvent. We will demonstrate the quality of this new CPHMD$^{MS\lambda D}$ model by its ability to accurately reproduce experimental pK_a values of simple mononucleotide

systems to a mean absolute error of 0.24 pK_a units. To the best of our knowledge, this is the first constant pH MD simulation for nucleic acids to be reported in literature.

2. THEORY

We briefly review the theory behind CPHMD and highlight the relevant modification in our implementation of CPHMD$^{MS\lambda D}$. In the original CPHMD model, the protonation/deprotonation process is simulated as a special case of λ-dynamics where the λ variables are used to define titration coordinates.[55−57] In λ-dynamics, the simulation is under the influence of a hybrid Hamiltonian, and its potential energy is defined by

$$U_{tot}(X, \{x\}, \{\lambda\}) = U_{env}(X) + \sum_{\alpha=1}^{N_{sites}} [\lambda_{\alpha,1}(U(X, x_{\alpha,1})) + \lambda_{\alpha,2}(U(X, x_{\alpha,2}))] \quad (1)$$

where N_{sites} is the total number of titrating residues, X represents the coordinates of the environment atoms, and $x_{\alpha,1}$ and $x_{\alpha,2}$ represent the coordinates of atoms in residue α that are associated with the protonated and unprotonated states, respectively. The titrating proton and the other atoms whose charges vary according to the protonation state of the residue (usually atoms within 2−3 bonds from the titrating proton) are included in both $x_{\alpha,1}$ and $x_{\alpha,2}$ and are defined as a part of the "titrating fragment." The scaling factor that is associated with the titrating residue α changes dynamically throughout the simulation and is described by a set of continuous coordinates that are governed by the following equations:

$$\lambda_{\alpha,1} = \sin^2 \theta_\alpha \text{ and } \lambda_{\alpha,2} = 1 - \sin^2 \theta_\alpha \quad (2)$$

The end points define the physically relevant protonated ($\lambda_{\alpha,1} = 1$, $\lambda_{\alpha,2} = 0$) and unprotonated ($\lambda_{\alpha,1} = 0$, $\lambda_{\alpha,2} = 1$) states. In recent work, Knight and Brooks developed the alternative $\lambda^{N_{exp}}$ functional form of λ:

$$\lambda_{\alpha,i}^{N_{exp}} = \frac{e^{c \sin \theta_{\alpha,i}}}{\sum_{j=1}^{N} e^{c \sin \theta_{\alpha,j}}} \quad (3)$$

When applied to the two-state system representing the protonated and unprotonated forms this functional form becomes

$$\lambda_{\alpha,1} = \frac{e^{c \sin \theta_{\alpha,1}}}{e^{c \sin \theta_{\alpha,1}} + e^{c \sin \theta_{\alpha,2}}} \text{ and } \lambda_{\alpha,2} = \frac{e^{c \sin \theta_{\alpha,2}}}{e^{c \sin \theta_{\alpha,1}} + e^{c \sin \theta_{\alpha,2}}} \quad (4)$$

This new form implicitly satisfies the constraints as required by λ-dynamics:

$$0 \leq \lambda_i \leq 1 \text{ and } \sum_{i=1}^{N} \lambda_i = 1 \quad (5)$$

The use of the $\lambda^{N_{exp}}$ functional form also expands the future functionality of our CPHMD$^{MS\lambda D}$ model to titrate between more than two states, such as the tautomeric forms of nucleic acids.

In CPHMD simulations, the overall free energy of deprotonation of a given residue, $\Delta G_{exp}(RNA)$, is obtained by calculating the difference between the free energy of deprotonation in the RNA environment, $\Delta G_{sim}(RNA)$, compared to that of a model compound in solvent, $\Delta G_{sim}(model)$. By equating this difference

37

dx.doi.org/10.1021/ct2006314 |*J. Chem. Theory Comput.* 2012, 8, 36–46

of free energies between the simulated system to that of the experimental system, we obtain

$$\Delta G_{exp}(\text{RNA}) - \Delta G_{exp}(\text{model}) = \Delta G_{sim}(\text{RNA}) - \Delta G_{sim}(\text{model}) \tag{6}$$

which can be rearranged to estimate the experimental free energy of deprotonation of the RNA:

$$\Delta G_{exp}(\text{RNA}) = \Delta G_{sim}(\text{RNA}) - \Delta G_{sim}(\text{model}) + \Delta G_{exp}(\text{model}) \tag{7}$$

The free energy of deprotonation of the model compound may also be expressed as

$$\Delta G_{sim}(\text{model}) = \ln(10)k_B T(pK_a - pH) \tag{8}$$

From this perspective, titratable groups in the RNA can be viewed as model compounds that are perturbed by the introduction of the RNA environment via nonbonded interactions, and this is the fundamental expression that needs to be calibrated for each titratable residue in our model, in the present study, adenosine and cytidine. For the initial calibration, the free energy of deprotonation of each isolated model compound calculated using traditional λ-dynamics provided the $\Delta G_{sim}(\text{model})$ value. When $\Delta G_{sim}(\text{model})$ is applied to the simulation as a bias, it results in a zero free energy difference between the protonated and unprotonated states, and this condition is equivalent to pH = pK_a. To simulate the system under different pH environments, eq 8 is used to derive the equivalent $\Delta G_{sim}(\text{model})$ value that should be applied to the simulation. The reference pK_a used in eq 8 was obtained from experimental pK_a values that were measured at zero ionic strength.[1]

In our implementation of CPHMD$^{MS\lambda D}$, two biases (F^{fixed} and F^{var}) are incorporated into the potential energy function, and the resulting total potential energy function in our CPHMD simulation may be written as

$$U_{tot}(X, \{x\}, \{\lambda\}) = U_{env}(X) + \sum_{\alpha=1}^{N_{sites}} [\lambda_{\alpha,1}(U(X, x_{\alpha,1}) - F^{fixed}_{\alpha,1})$$
$$+ \lambda_{\alpha,2}(U(X, x_{\alpha,2}) - F^{fixed}_{\alpha,2})$$
$$+ F^{var}_{\alpha,1}(\lambda_{\alpha,1}) + F^{var}_{\alpha,1}(\lambda_{\alpha,2})] \tag{9}$$

In this formalism, the fixed biasing potential that is applied to the unprotonated state ($F^{fixed}_{\alpha,2}$) represents the calibrated $\Delta G_{sim}(\text{model})$ value. The other fixed biasing potential applied to the protonated state ($F^{fixed}_{\alpha,1}$) is kept at zero. Using this setup, when the titration coordinates are allowed to propagate dynamically, the two end points that correspond to physical states may not be well-sampled. Thus, we included the variable biasing potential (F^{var}) which applies an additional bias to encourage sampling of physical states. Identical variable biases are applied to both protonation states.

$$F^{var}_{\alpha,i} = \begin{cases} k_{bias}(\lambda_{\alpha,i} - 0.8)^2; & \text{if } \lambda_i < 0.8 \\ 0; & \text{otherwise} \end{cases} \tag{10}$$

The populations of unprotonated (N^{unprot}) and protonated (N^{prot}) states are extracted from the λ trajectory at each pH value, which are used to derive the unprotonated fraction (S^{unprot}):

$$S^{unprot}(pH) = \frac{N^{unprot}(pH)}{N^{unprot}(pH) + N^{prot}(pH)} \tag{11}$$

Overall pK_a values can be calculated by running simulations at different pH values and fitting the series of S^{unprot} values to a more generalized version of the Hendenson−Hasselbach formula:

$$S^{unprot}(pH) = \frac{1}{1 + 10^{-n(pH - pK_a)}} \tag{12}$$

where n is the Hill coefficient. In this formalism, n has a theoretical value of one, and deviations from this value indicate the degree of cooperativity ($n > 1$) or anticooperativity ($n < 1$) between strongly interacting titratable groups.[74,75]

3. METHODS

3.1. Generating Input Structures. The input structures of the nucleic acids that were used for the simulations were generated from CHARMM topology files using the IC facility in CHARMM, while hydrogen atoms were added using the HBUILD facility.[76] Model compounds, adenosine and cytidine, were solvated in a cubic box of explicit TIP3P water molecules[77] of length ∼20 Å using the convpdb.pl tool from the MMTSB toolset.[78] The test compounds, adenosine monophosphate (AMP) and cytidine monophosphate (CMP) and dinucleotide sequences of CYT-CYT, ADE-ADE, and CYT-ADE were solvated in a cubic box of explicit water molecules of length ∼50 Å using the convpdb.pl tool from the MMTSB toolset. The ionic strength was simulated by adding the appropriate number of Na$^+$ and Cl$^-$ ions to match experimentally reported salt concentrations using convpdb.pl. For the mononucleotides, two isomers in the form of 5′-phosphate and 3′-phosphate were constructed using the patch keywords 5PHO and 3PHO, respectively, in CHARMM. All other nucleic acid structures had hydroxyl groups patched to the terminal ends via patch keywords 5TER and 3TER. Additional patches were constructed to represent the protonated forms of adenine and cytosine, and all of the associated bonds, angles and dihedrals were explicitly defined in the patch. Each titratable residue was simulated as a hybrid model that explicitly included atomic components of both the protonated and unprotonated forms. The titratable fragment included the nitrogen atom that is protonated, the protonated hydrogen, and adjacent atoms whose partial charge differed according to the protonation state (see Table 1 and corresponding Tables S1−S3 in the Supporting Information). The environment atoms were defined as all atoms that were not included in the titratable fragments.

3.2. MD Simulations. MD simulations were performed within the CHARMM macromolecular modeling program (version c36a6)[76] using the CHARMM36 all-atom force field for RNA and TIP3P water.[79] The simulation set up for λ dynamics is similar to that reported by Knight and Brooks.[72,73] The SHAKE algorithm[80] was used to constrain the hydrogen heavy-atom bond lengths. The Leapfrog Verlet integrator was used with an integration time step of 2 fs. A nonbonded cutoff of 15 Å was used with an electrostatic force shifting function, and a vdW switching function between 10 Å and 12 Å. The λ dynamics was performed within the BLOCK facility using the MSλD framework (MSLD) and selecting the $\lambda^{N_{exp}}$ functional form for λ (FNEX). Linear scaling by λ was applied to all energy terms except bond, angle, and dihedral terms, which were treated at full strength regardless of λ value to retain physically reasonable geometries. Each θ_α was assigned a fictitious mass of 12 amu·Å^2, and λ values were saved every 10 steps. The threshold value for assigning $\lambda_{\alpha,i} = 1$ was $\lambda_{\alpha,i} \geq 0.8$. Variable biases ($F^{var}$) were added

Table 1. Charges and Atom Types Assigned to the Protonated and Unprotonated States of Titratable Nucleic Acids

name	atom	unprotonated		protonated	
		atom type	charge	atom type	charge
ADE	H1	—	—	HN2	0.527
	N1	NN3A	−0.74	NN2G	−0.489
	C2	CN4	0.50	CN4	0.611
	C6	CN2	0.46	CN2	0.571
CYT	H3	—	—	HN2	0.52
	C5	CN3	−0.13	CN3	−0.174
	C2	CN1	0.52	CN1	0.75
	N3	NN3	−0.66	NN2C	−0.874
	C4	CN2	0.65	CN2	0.962
	N4	NN1	−0.75	NN1	−0.654
	H41	HN1	0.37	HN1	0.42
	H42	HN1	0.33	HN1	0.38

to the hybrid potential energy function, and the associated force constant (k_{bias}) was optimized to enhance transition rates between the two protonation states. Since identical k_{bias} values were applied to both protonated and unprotonated states, the PMF at the end-points were not altered, and no reweighting scheme was required. The temperature was maintained at 298 K by coupling to a Langevin heatbath using a frictional coefficient of 10 ps^{-1}. Prior to the simulation, each system was minimized using 300 steps of steepest descents (SD), followed by 200 steps of adopted basis Newton−Raphson. After an initial heating of 4 ps and equilibration of 4 ps, a production run of 1 ns was performed, unless otherwise stated.

3.3. Calculation of pK_a Value. In our protocol, a single S^{unprot} value was estimated by combining the populations of N^{prot} and N^{unprot} from three independent simulations that used different initial seed values. These combined S^{unprot} ratios that were computed at different pH values were then fitted to eq 12 to obtain a single pK_a value. Unless otherwise specified, the reported pK_a value and its error correspond to the mean and standard deviation calculated from three sets of pK_a calculations.

4. RESULTS AND DISCUSSION

4.1. New CHARMM Parameters for Protonated Adenine and Cytosine. We calculated the partial charges for the adenine and cytosine nucleobases in their neutral (unprotonated) and charged (protonated) states using the MMFF94 force field.[81] The change in the partial charge was added to the existing partial charge parameters for neutral adenine and cytosine in CHARMM to assign the charge distribution for the protonated residues. A summary of the differences for charges and atom types between the protonated and unprotonated nucleic acids is reported in Table 1. Parameters for the bond, angle, and dihedral energy terms for the protonated nucleic acid were adapted from existing nucleic acid structures in CHARMM (see Supporting Information). For the protonated adenine, the respective bonded parameters were obtained from guanine, specifically from the six-membered ring component that has atoms analogous to that of adenine (N1, H1, C2, and C6). For the protonated cytosine, the respective bonded parameters were obtained from a tautomeric form of neutral cytosine (obtained from patch CYT1).

4.2. Optimization of Model Potential Parameters. Our CPHMD$^{MS\lambda D}$ model was implemented using the recently developed $\lambda^{N_{exp}}$ functional form for λ in MSλD:[73]

$$\lambda_{\alpha,i}^{N_{exp}} = \frac{e^{c\sin\theta_{\alpha,i}}}{\sum\limits_{j=1}^{N} e^{c\sin\theta_{\alpha,j}}}$$

Knight and Brooks reported setting the coefficient to 5.5 for the optimal balance between enhancing transition rate and maintaining numerical stability of the integrator in different environments.[73] An identical setup was used successfully in our CPHMD$^{MS\lambda D}$ model. As with the previous implementation of CPHMD for protein residues,[52−54] we have used the calibrated free energy of deprotonation (G_{bias}) as the fixed biasing potential value in our simulation. The free energy of deprotonation was calibrated for each isolated model compound, i.e., adenosine and cytidine, embedded in explicit solvent using traditional λ-dynamics. In order to facilitate transitions between the two protonation states, we optimized the force constant (k_{bias}) on the variable biasing potential that was applied for each model compound.

It is interesting to note that without the application of the variable bias, no transitions between the protonated and unprotonated states were observed at conditions pH = pK_a, where one should expect equal population of both states and the maximum transition rate between the two states (see Figure 2). At values of k_{bias} < 20 kcal/mol, there were very few transitions in λ phase space between the two states for the entire duration of a 1 ns trajectory. At values k_{bias} > 40 kcal/mol, transitions were rapid, but the end states were not adequately sampled. The optimal value of k_{bias} for each nucleoside was selected by considering the competing needs for a high number of transitions and an adequate sampling of the end points (i.e., maintaining a high fraction of physical ligands (FPL) that were sampled). As illustrated in Figure 3, these two properties were observed to be anticorrelated to each other, and there is a distinct range of k_{bias} values (between 25 and 35 kcal/mol) that yielded good transition rates, and where more than 80% of the simulation is spent at the physically relevant end points. The optimized parameters for the two model potentials are reported in Table 2.

The variable bias with a relatively large force constant of 28−30 kcal/mol that is required to achieve a reasonable number of transitions in our simulation may be rationalized by noting that the appearance of a full charge unit when titrating between the two states is likely to significantly perturb the solvent environment around the nitrogen atom. We suggest that time is required for the solvent to reorganize and fully accommodate the new charge distribution as the system titrates from the unprotonated to the protonated state. Figure 4 provides a comparison of the radial distribution function (RDF) of water molecules surrounding the N1 atom of adenosine in its protonated and unprotonated state and indicates that considerable rearrangement of the first solvent shell upon ionization of the residue does occur. For the RDF that describes the distances between N1 and the TIP3P oxygen atoms, we observed that the charged protonated state had a first solvation shell (2.7 Å) that is slightly closer than the uncharged unprotonated state (2.9 Å). A more significant change, however, was observed for the RDF that describes the distances between N1 and the TIP3P hydrogen atoms in which the protonated state first solvation shell was pushed back (3.4 Å) compared to that of the unprotonated state (2.0 Å). These observations are consistent with the expectation that water
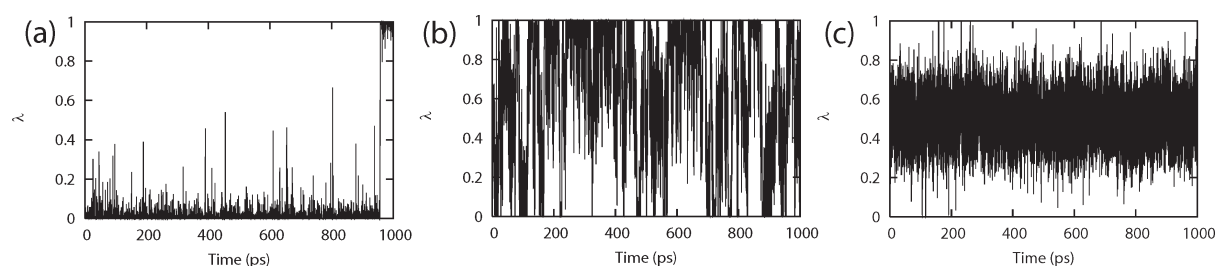
Figure 2. Transitions between the protonated and unprotonated state of adenosine in λ phase space at pH = pK_a for a 1 ns trajectory with varying k_{bias} values of (a) 20, (b) 30 and (c) 40.
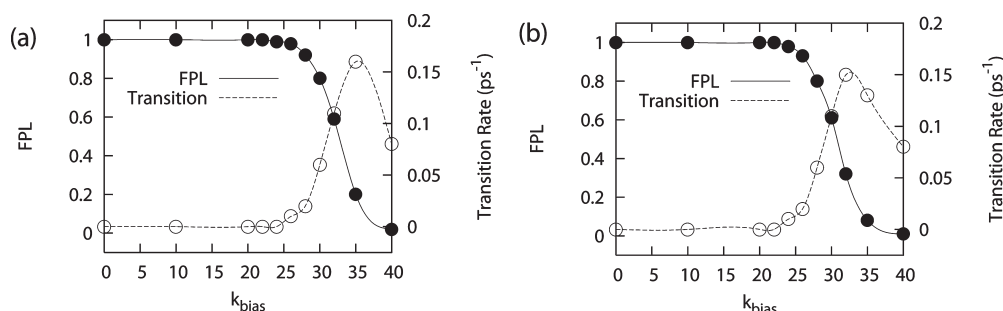

Figure 3. Effect of increasing k_{bias} on the transition rate and FPL for (a) adenosine and (b) cytosine. Sampling characteristics were obtained from 5 independent MD runs of 1 ns each.

Table 2. Parameters for the Model Potential[a]

nucleotide	G_{bias} (kcal/mol)	k_{bias} (kcal/mol)	reference pK_a[b]
adenine	19.39	29.75	3.50
cytosine	75.24	27.75	4.08

[a] G_{bias} was assigned to be the free energy of deprotonation of adenosine or cytidine at zero ionic strength, and k_{bias} was optimized to achieve a maximum transition rate while maintaining physical states for more than 80% of the entire trajectory. [b] Reference pK_a is the experimental pK_a values for the model compounds that were measured at zero ionic strength.[1]

molecules would orient their hydrogen atoms toward the partial negative charge of the nitrogen atom in the unprotonated state and subsequently would flip their hydrogen atoms "outwards" and orient their oxygen atoms closer toward the partial positive charge of the protonated hydrogen that is present in the protonated state. Similar trends were observed for the RDF of water molecules that surround the N3 atom of cytidine (data not shown). An analogous change in RDF of water molecules around the protonated N5 atom of the substrate of dihydrofolate reductase was also observed with MD simulations that sampled different protein conformations that altered the water accessibility of the ligand pocket.[82]

4.3. Sampling Efficiency of Explicit Solvent CPHMD Simulations. The sampling efficiency as measured by the transition rates between the two protonation states in our CPHMD$^{MS\lambda D}$ model is quite good with ~50 transitions per ns for our model compounds at pH = pK_a. Given that the solvent reorganization upon the perturbation of a full charge unit was reported to be on a time scale of up to 3 ps in previous MD simulations[39] and that the mean duration of the physically relevant protonation states in our simulations is 20 ps, the sampling characteristics of our system

are sufficient to allow solvent reorganization to occur. However, the transition rate is markedly lower than what has been observed in CPHMD simulations that are performed using implicit solvent models.[52,53] It should be noted that our model potential parameters, specifically the k_{bias} values as implemented in CPHMD$^{MS\lambda D}$ have been selected conservatively. For example, the transition rate can be doubled at the expense of reducing the FPL to 0.6 (Figure 3) which, provided that simulations are long enough to sufficiently enumerate the relative end-state populations, may be a better option for simulating full RNA systems where observing transitions between protonation states may be more challenging.

The more limited sampling efficiency of explicit solvent CPHMD simulations was also recently reported by Grubmüller and co-workers where the titration of an imidazole model compound achieved ~100 transitions in a 20 ns trajectory,[61] which is a rate of ~5 transitions per ns. Considering the computational expense of performing explicit solvent simulations, our rate of ~50 transitions per ns that is achieved with the optimization of our implementation of explicit solvent CPHMD is clearly advantageous. Finally, in Table 3, we present a comparison between the sampling characteristics of our simulation to that of previous work performed in the MSλD framework by Knight and Brooks for modeling series of inhibitors of HIV-1 reverse transcriptase.[72] Using the same force constant for the variable bias (i.e., k_{bias} = 7) as what was previously reported, we observed a significant drop in sampling performance with virtually no transitions observed between the two protonation states at pH = pK_a. Our optimization of k_{bias} assisted in improving the sampling characteristics, but the transition rate still remains about 4-fold less efficient than previous work. We note that earlier work performed by Knight and Brooks modeled hybrid ligands in which the substituents did not differ significantly in
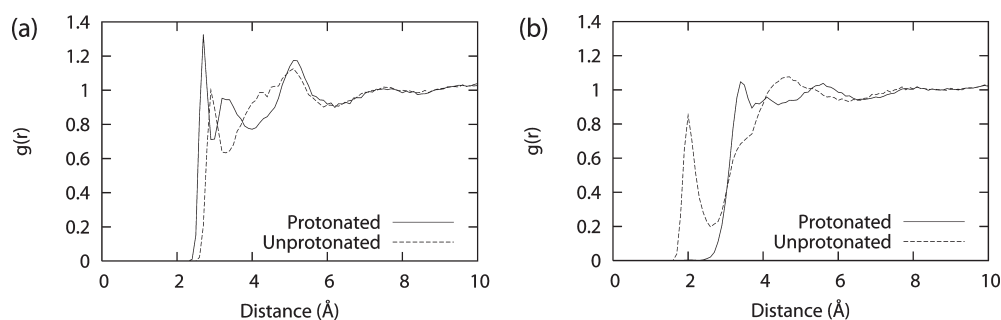
Figure 4. RDF of water molecules for (a) N1(ADE)-O(TIP3P) distances and (b) N1(ADE)-H(TIP3P) distances within a sphere of 10 Å from the N1 atom of adenosine in both protonated and unprotonated states.

Table 3. Sampling Characteristics of Simulations Performed at pH = pKa

	previous work[a]	adenosine (default)	cytidine (default)	adenosine (optimized)	cytidine (optimized)
k_{bias}	7.00	7.00	7.00	29.75	27.75
FPL	0.780	1.000	1.000	0.828	0.832
transitions (ps^{-1})	0.190	0.001	0.001	0.050	0.051

[a] Sampling characteristics of a two-state hybrid ligand in explicit water investigated in previous work (obtained from Table 3, hybrid ligand F).[72]

terms of their partial charge distributions. Thus, the introduction of a full charge unit when titrating between the two states in CPHMD$^{MS\lambda D}$ is likely to be the primary cause for the reduction sampling efficiency that we observe in the present simulations.

4.4. Convergence and Precision of Calculations. The challenges associated with sampling and convergence for CPHMD simulations have been reported on several occasions,[48,54] and these are expected to be an even greater concern in explicit solvent CPHMD where sampling efficiency is reduced. To validate the robustness of our CPHMD$^{MS\lambda D}$ model in its ability to achieve adequate convergence, we performed a series of simulations at pH = pK_a for our model compounds. The degree of convergence in our simulations was determined by calculating the unsigned deviation between the free energy of protonation, estimated from subsets of shortened trajectories, and the free energy of protonation that was estimated from 10 1 ns trajectories. Different combinations of trajectory length and number of independent runs were systematically examined to determine the most cost-effective trade-off between computational expense and precision of the calculations. The results are summarized in Figure 5. It was observed that individual trajectories required at least 100 ps to reliably observe any transitions between protonation states. In fact, we observed that a minimum simulation time of ~500 ps per trajectory was required to obtain a precision of ~0.20 kcal/mol in our calculations (Figure 5a), and running multiple shorter independent runs would not produce converged results unless the 500 ps threshold was crossed. Our results indicate that good precision can be achieved by using a total simulation time of 3 ns in the form of 3 independent runs of 1 ns each, where the unsigned deviations for the free energy of deprotonation was 0.05 kcal/mol for adenosine (Figure 5b). It should be noted, however, that this level of precision was achieved in previous work three times more quickly for hybrid ligands whose charge distributions were similar. All subsequent calculations of pK_a values in this paper were estimated using three independent runs of 1 ns each.

The performance of the multisite λ-dynamics (MSλD) approach, on which CPHMD$^{MS\lambda D}$ is based, has been evaluated in comparison to traditional FEP calculations by Knight and Brooks.[72] For substituents with similar charge distributions, it has been established that relative hydration and relative binding free energies calculated from both MSλD and FEP are in good agreement with each other and that MSλD was three times more efficient than regular FEP. Our current work involves substituents that have significantly different charge distributions from one another, i.e., the charges associated with the protonated and deprotonated states, respectively, and consequently CPHMD$^{MS\lambda D}$ takes longer to converge. Analogously, we expect that FEP calculations will also take longer than what was reported in Knight and Brooks but would still be less efficient in their convergence than MSλD calculations. In addition, traditional FEP calculations are not well-suited for simultaneously exploring multiple titrating sites or multiple tautomers, and so the MSλD-based approach of CPHMD$^{MS\lambda D}$ is more generalizable than FEP methods to model these more complex situations.

4.5. Calibration Curve of Model Systems: Adenosine and Cytidine. We calibrated our CPHMD$^{MS\lambda D}$ model at 298 K using zero salt concentration. The reference pK_a that was used in the calibration was the experimental pK_a that was measured under similar conditions (25 °C at zero ionic strength).[1] The titration curve of the model nucleoside compounds, adenosine and cytidine, is shown in Figure 6. The best-fit Henderson—Hasselbalch curve has a near ideal Hill coefficient for adenosine ($n = 0.94$) and cytidine ($n = 0.93$). The calculated pK_a value of 3.50 for adenosine was in excellent agreement with experimental values, and the pK_a of 4.22 for cytidine is only slightly higher than the reference value by 0.14 pK_a units. The accuracy of the calculated pK_a values is determined primarily by the sampling efficiency at pH = pK_a and the quality of the calibration of the G_{bias} values that are used to simulate distinct pH conditions. Our results demonstrate that a series of 3 × 1 ns simulations is sufficient to provide reasonably accurate results, which is significantly less than the 20 ns trajectory employed by Grubmüller and co-workers in their explicit solvent CPHMD model.[61]
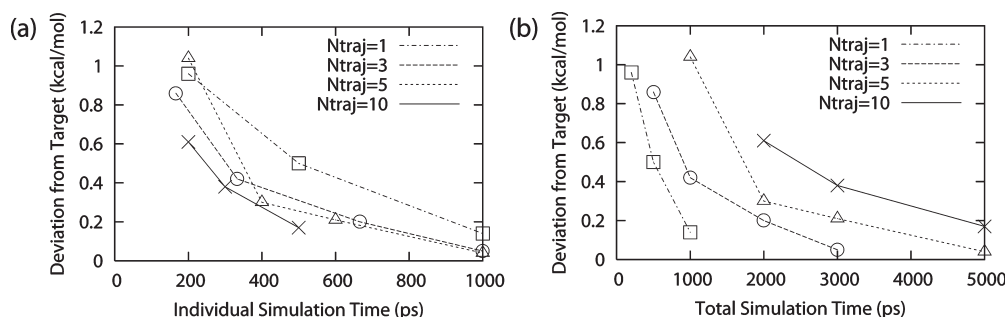
Figure 5. Unsigned deviation for the free energy of deprotonation of adenosine as a function of (a) total simulation time from all *N* trajectories and (b) individual simulation time of each of the *N* trajectories.
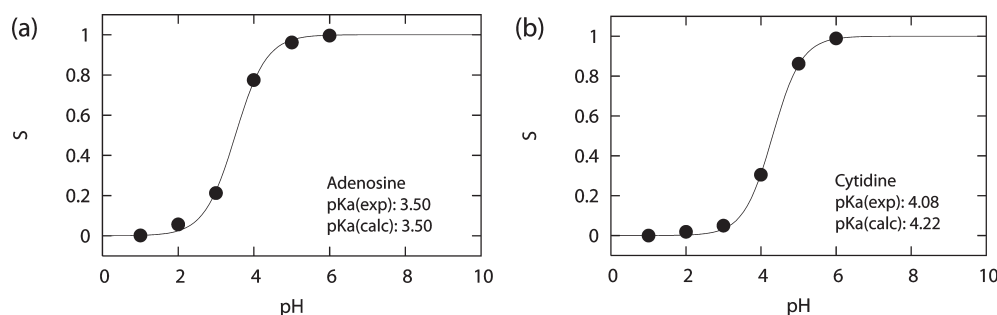

Figure 6. Sample titration curves for model nucleoside compounds (a) adenosine and (b) cytidine.

Table 4. Calculated and Experimental pKa Values of Test Compounds

compound	[NaCl] (M)	calculated	experimental	abs. error
β-AMP-3	no salt	4.20 ± 0.06	—	—
β-AMP-3	0.15	3.79 ± 0.11	3.65	0.14
AMP-5	no salt	4.08 ± 0.03	—	—
AMP-5	0.15	3.89 ± 0.16	3.74	0.15
β-CMP-3	no salt	4.77 ± 0.05	—	—
β-CMP-3	0.15	4.56 ± 0.10	4.31	0.25
CMP-5	no salt	4.90 ± 0.07	—	—
CMP-5	0.10	4.67 ± 0.08	4.24	0.43

4.6. Quantitative pKa Value Calculations for Simple Nucleotides. First, we tested our CPHMD$^{MS\lambda D}$ model on single nucleotide test compounds, adenine monophosphate (AMP) and cytosine monophosphate (CMP), at zero ionic strength, and the results are summarized in Table 4. The calculated pKa values for AMP-5 and β-AMP-3 were 4.08 and 4.20, respectively. Compared to adenosine, the pKa values of these nucleotide counterparts were slightly elevated by ∼0.5 pKa units. Similarly, the nucleotide counterparts of cytidine with pKa values for CMP-5 and β-CMP-3 of 4.90 and 4.77, respectively, had slightly elevated pKa values by ∼0.5 pKa units compared to cytidine. The calculated pKa values for both 5′- and 3′-phosphate isomers of adenosine and cytosine are not statistically different at the 95% confidence interval. The increase in the calculated pKa values from their nucleoside counterparts is expected, since the presence of the negative charge from the phosphate group may interact with the positively charged protonated base and weakly stabilize it, thus increasing the population of the protonated state and causing a corresponding increase in the calculated pKa value.

In order to compare our calculated pKa values with experimental results, we performed simulations that mimicked the ionic strength of the environment (i.e., 100−150 mM NaCl) in which the experiments were performed.[83,84] By explicitly incorporating the salt environment, the calculated pKa values are systematically lowered relative to those obtained from the zero ionic strength simulations. This shift in pKa values is to be expected since the presence of Na$^+$ ions screens the electrostatic effects of the phosphate group. The results in Table 4 indicate that our pKa predictions had an average absolute error of 0.24 pKa units compared to experiment, and we conclude that our CPHMD$^{MS\lambda D}$ model is capable of making accurate quantitative predictions of pKa values for simple nucleotides. These results also indicate that our model is capable of accounting for the differences between zero and nonzero ionic strength environments and highlights the importance of simulating the system at the appropriate ionic strength to mimic experimental conditions.

4.7. Modeling Interactions between Adjacent Titrating Residues. Finally, we tested our CPHMD$^{MS\lambda D}$ model on dinucleotide sequences ADE-ADE, CYT-CYT, and CYT-ADE at zero ionic strength, where both nucleotides were titrated simultaneously in the same simulation. The pKa values were shifted upward compared to the nucleoside model compounds for all sequences, ADE-ADE (4.08 ± 0.20 and 4.06 ± 0.16), CYT-CYT (4.93 ± 0.05 and 4.76 ± 0.09), and CYT-ADE (5.06 ± 0.07 and 3.85 ± 0.26) and were similar to the corresponding mononucleotide pKa values. For some of the sets of pKa calculations for the dinucleotide sequences, the Hill coefficient had more significant deviations from one compared to the monomeric compounds. Specifically, the value was lowered (*n* < 0.8) for 5 of the 9 sets of pKa calculations. When the Hill coefficient deviates from one, it suggests that adjacent residues
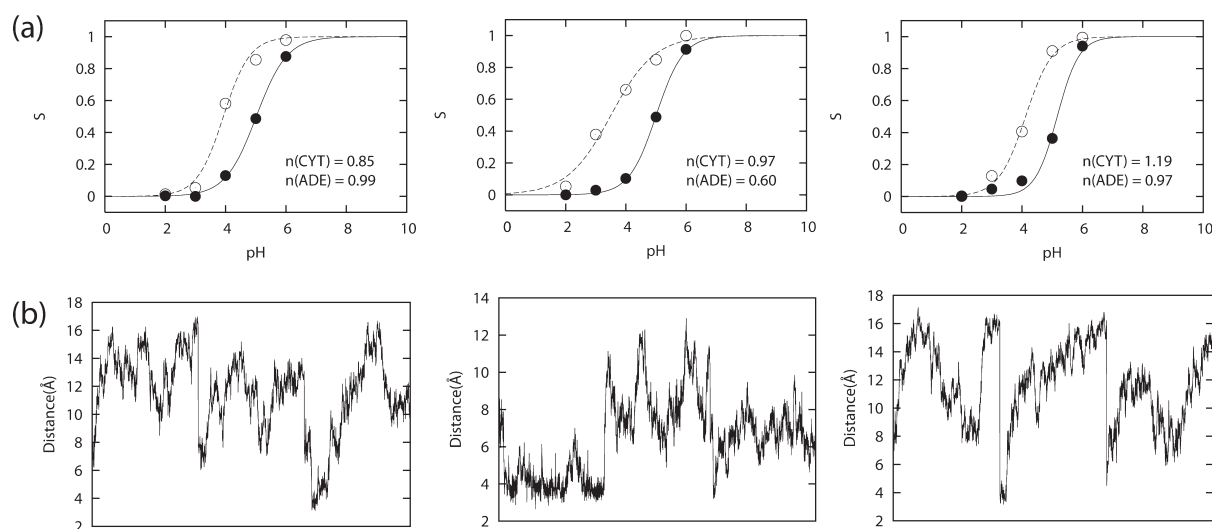
Figure 7. (a) Titration curves for CYT-ADE and (b) time series of distance between N3 CYT and N1 ADE atoms at pH 3 for all 3 sets of pK_a calculations.

are interacting with each other in either a cooperative ($n > 1$) or an anticooperative ($n < 1$) fashion. Cross-correlation analysis of the protonation states (data not shown) however indicates only weakly correlated behavior, which suggests that the interaction between adjacent residues is not strong. The second set of pK_a calculations on CYT-ADE exhibited the lowest Hill coefficient ($n = 0.60$), indicating the strongest anticooperative behavior. Analysis of the individual titration curves as shown in Figure 7 indicates that the S^{unprot} ratio shows the greatest deviation between the second set and the other two sets at pH 3. We analyzed the mean distance between the nitrogen atom that is protonated in CPHMD (i.e., N3 CYT and N1 ADE) of adjacent residues at pH 3, and the results are shown in Figure 7. In one simulation of the second set, the mean distance sampled was about 4−6 Å, in comparison to the typical values of 8−16 Å for all other simulations. We suggest that this simulation contributed significantly to the higher S^{unprot} ratio for the second set that in turn gave rise to the lower Hill coefficient. The lack of strong interactions between adjacent titrating residues in the other two sets of pK_a calculations of CYT-ADE is apparently due to the lack of sampling of configuration space in which these two residues are close enough to influence each other's protonation state. We suggest that stronger cooperative or anticooperative effects are likely to be observed when modeling RNA structures with stable conformations in which the nucleobases are held in close proximity to one another.

4.8. Moving toward CPHMD of Full RNA Systems. The remarkable agreement of our calculated pK_a values to experiment is encouraging; however, several challenges may be anticipated when applying our CPHMD$^{MS\lambda D}$ model to full RNA structures.

First, instead of isolated monomeric compounds, the titratable residues of interest in RNA macromolecules are nucleotides that are buried in the interior of the RNA which interact with multiple residues (e.g., via base-pairing interactions). Therefore, the increased perturbation of the titrating residue's local environment and the increased complexity arising from nonbonded interactions with adjacent residues are likely to reduce sampling efficiency in full RNA structures. We illustrate this claim with a hypothetical example of a residue whose pK_a value varies with RNA conformation. The formation of a Watson−Crick A-T base

pair involves the N1 atom of adenine, and this results in a depression of its pK_a value relative to the isolated base since the nitrogen atom (that would be protonated in CPHMD) is now serving as a hydrogen-bond acceptor. Conformational fluctuations, such as base flipping motions, may expose these buried nucleotides to the solvent and cause a corresponding increase in their respective pK_a values. In order to reproduce experimentally measured pK_a values from NMR studies in the above example, it may be necessary to sample these two conformational states that are observed at the time scale of which the measurements were taken. The use of advanced sampling methods, such as replica exchange, has been previously implemented in protein CPHMD by Brooks and co-workers to improve sampling performance[54] and when used with CPHMD$^{MS\lambda D}$ may be expected to yield similar improvements in model performance. Similarly, accelerated MD[85] has been implemented with CPHMD, and it has yielded improvements in conformation sampling.[48] Other sampling methods developed to sample long time scale conformation changes, such as self-guided langevin dynamics,[86,87] may also achieve a similar effect.

Another challenge may arise from the sampling in λ phase space in the presence of many interacting titratable groups. Under the physiologically relevant pH range, protein residues typically have 4 titratable residues out of 20 amino acids. In contrast, half of the nucleic acid building blocks are titratable in our current CPHMD$^{MS\lambda D}$ model. While it may be common that titratable residues on a protein are separated from one another in terms of spatial proximity, and thus the state of one titratable residue is unlikely to influence the others, this is not the case for RNA. In the absence of prior information about which residues will likely be in which protonation states, all adenosine and cytosine residues in the macromolecule could be modeled as titratable. However, in this case, the cooperative or anticooperative effect that these simultaneously titrating residues have on one another could be significant. Such a situation may lead to hidden barriers in adjacent λ phase spaces, in which the λ values of residue i are restricting the propogation of the λ values of residue j. Thus, the efficiency of sampling in λ phase space would need to be improved, and a lower FPL without compromising the physical accuracy of the model may be necessary in order to

maintain a reasonable minimum transition rate between the two protonation states. A recently developed enhanced sampling technique, orthogonal space random walk (OSRW) has been developed to address such sampling challenges associated with strongly coupled hidden free energy barriers, and it has been successfully applied to predict the pK_a values of buried protein residues that are typically not accurately reproduced in conventional CPHMD approaches. The implementation of OSRW with $CPHMD^{MS\lambda D}$ could potentially model strongly coupled titrating residues with better accuracy.[62,88]

For $CPHMD^{MS\lambda D}$ to successfully investigate pH-dependent properties of RNA structures over a longer time scale (μs and beyond) or pH-dependent properties of large RNA structures, such as a ribosomal subunit, a reduction of the computational expense that is associated with explicit solvent CPHMD is desirable. Greater computational efficiency may be achieved by using hybrid explicit/implicit solvation models, in which a few layers of explicit solvent water molecules are placed near the RNA surface with the rest of the environment described by an implicit solvation model.[89−91] Other models reduce the number of explicit waters in the simulation by using a thin shell of explicit water and hold these waters near the RNA surface with a restraining force.[92,93] The use of the surface constrained all atom solvent model that requires fewer explicit water molecules than the periodic boundary conditions implementation in MD simulations[94,95] has also been validated on a number of pK_a calculations performed on protein residues.[37,96] Multiscale modeling approaches that use a coarse-grained model to provide the reference potential for the thermodynamic cycle that can be used to speed up the free energy calculations of protonation events to simulate pH-dependent dynamics have also been reported recently.[49,97] Alternatively, we have seen considerable advances in implicit solvent models in recent years,[98] and they been been successfully implemented with protein CPHMD.[52−54] Established work in parametrizing implicit solvation models for RNA is encouraging,[99] but ongoing work in our lab (unpublished results) indicates that implicit solvent models do not accurately reproduce explicit solvent simulations results when simulating RNA. Therefore, the successful implementation of an implicit solvation model to $CPHMD^{MS\lambda D}$ would be an avenue for future development.

Finally, while the use of NaCl may serve to reproduce the ionic strength environment at which experimental studies are conducted, divalent ions, such as Mg^{2+}, play functional roles in many RNA structures, and current parameters would need to be examined to ensure their ability to reproduce experimental observables in RNA macromolecules. Our $CPHMD^{MS\lambda D}$ model may also be expanded to include titratable residues of guanine, uracil, and thymine. Although the bulk of experimental studies have implicated adenine and cytosine as key protonated residues in RNA, there is some evidence that suggests that the presence of protonated guanine, such as the G8 residue in the active site of the hairpin ribozyme, may also play a mechanistic role.[100] The adoption of the $\lambda^{N_{exp}}$ functional form for λ in our $CPHMD^{MS\lambda D}$ model also allows us to expand the representation of the titratable fragments to include tautomeric forms of nucleotides in both the unprotonated and protonated states. Recent theoretical studies have also suggested that stable tautomers may exist under specific conditions.[101−103] Thus, the ability to titrate among four states (unprotonated, unprotonated tautomer, protonated and protonated tautomer) may assist in clarifying the structural or mechanistic roles that involve tautomeric forms in RNA.

5. CONCLUSION

In conclusion, we have parametrized a protonated adenine and protonated cytosine for use in the first reported constant pH molecular dynamics (CPHMD) for nucleic acids. We have adopted the new functional form $\lambda^{N_{exp}}$ for λ that was recently developed for multisite λ-dynamics (MSλD) and demonstrate good sampling characteristics in which rapid and frequent transitions between the protonated and unprotonated states at $pH = pK_a$ are achieved, while sampling the physically relevant protonation states for more than 80% of the trajectory. Compared to existing implementations of explicit solvent CPHMD, the sampling in our method sees a 10-fold improvement, while maintaining sufficient residency time of the physical protonation states to ensure proper solvent reorganization. A series of 3 independent runs of 1 ns each was determined to be sufficiently precise for calculating the pK_a values for simple nucleotide systems. pK_a values calculated for simple nucleotides are in a good agreement with experimentally measured values with a mean absolute error of 0.24 pK_a units, affirming that our $CPHMD^{MS\lambda D}$ model has the ability to make accurate quantitative predictions for simple nucleotide systems. Our work paves the way for the deployment of CPHMD as a powerful tool to investigate pH-dependent biological properties of RNA macromolecules.

■ ASSOCIATED CONTENT

ⓢ Supporting Information. Parameters for the bond, angle, and dihedral energy terms for protonated adenine and cytosine. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: brookscl@umich.edu.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENT

■ REFERENCES

(1) Izatt, R. M.; Christensen, J. J.; Rytting, J. H. *Chem. Rev.* **1971**, *71*, 439.

(2) Gao, X. L.; Patel, D. J. *J. Biol. Chem.* **1987**, *262*, 16973.

(3) Asensio, J. L.; Lane, A. N.; Dhesi, J.; Bergqvist, S.; Brown, T. *J. Mol. Biol.* **1998**, *275*, 811.

(4) Jang, S. B.; Hung, L. W.; Chi, Y. I.; Holbrook, E. L.; Carter, R. J.; Holbrook, S. R. *Biochemistry* **1998**, *37*, 11726.

(5) Bink, H. H.; Hellendoorn, K.; van der Meulen, J.; Pleij, C. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13465.

(6) Morse, S. E.; Draper, D. E. *Nucleic Acids Res.* **1995**, *23*, 302.

(7) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497.

(8) Lee, J. C.; Gutell, R. R. *J. Mol. Biol.* **2004**, *344*, 1225.

(9) Durant, P. C.; Davis, D. R. *J. Mol. Biol.* **1999**, *285*, 115.

(10) Chen, G.; Kennedy, S. D.; Turner, D. H. *Biochemistry* **2009**, *48*, 5738.

(11) Muth, G. W.; Chen, L.; Kosek, A. B.; Strobel, S. A. *RNA* **2001**, *7*, 1403.

(12) Siegfried, N. A.; O'Hare, B.; Bevilacqua, P. C. *Biochemistry* **2010**, *49*, 3225.

(13) Kim, M.; Huang, T.; Miller, J. H. *J. Bacteriol.* **2003**, *185*, 4626.

(14) Giri, I.; Stone, M. P. *Biochemistry* **2003**, *42*, 7023.

(15) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. *Science* **2000**, *287*, 1493.

(16) Bevilacqua, P. C.; Brown, T. S.; Nakano, S.; Yajima, R. *Biopolymers* **2004**, *73*, 90.

(17) Das, S. R.; Piccirilli, J. A. *Nat. Chem. Biol.* **2005**, *1*, 45.

(18) Wilson, T. J.; Ouellet, J.; Zhao, Z. Y.; Harusawa, S.; Araki, L.; Kurihara, T.; Lilley, D. M. *RNA* **2006**, *12*, 980.

(19) Chen, J.-H.; Yajima, R.; Chadalavada, D. M.; Chase, E.; Bevilacqua, P. C.; Golden, B. L. *Biochemistry* **2010**, *49*, 6508.

(20) Shih, I. H.; Been, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1489.

(21) Wadkins, T. S.; Shih, I.; Perrotta, A. T.; Been, M. D. *J. Mol. Biol.* **2001**, *305*, 1045.

(22) Ke, A.; Zhou, K.; Ding, F.; Cate, J. H.; Doudna, J. A. *Nature* **2004**, *429*, 201.

(23) Gong, B.; Chen, J. H.; Chase, E.; Chadalavada, D. M.; Yajima, R.; Golden, B. L.; Bevilacqua, P. C.; Carey, P. R. *J. Am. Chem. Soc.* **2007**, *129*, 13335.

(24) Cerrone-Szakal, A. L.; Siegfried, N. A.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2008**, *130*, 14504.

(25) Ravindranathan, S.; Butcher, S. E.; Feigon, J. *Biochemistry* **2000**, *39*, 16026.

(26) Ryder, S. P.; Oyelere, A. K.; Padilla, J. L.; Klostermeier, D.; Millar, D. P.; Strobel, S. A. *RNA* **2001**, *7*, 1454.

(27) Kuzmin, Y. I.; Da Costa, C. P.; Cottrell, J. W.; Fedor, M. J. *J. Mol. Biol.* **2005**, *349*, 989.

(28) Nam, K.; Gao, J.; York, D. M. *J. Am. Chem. Soc.* **2008**, *130*, 4680.

(29) Guo, M.; Spitale, R. C.; Volpini, R.; Krucinska, J.; Cristalli, G.; Carey, P. R.; Wedekind, J. E. *J. Am. Chem. Soc.* **2009**, *131*, 12908.

(30) Cottrell, J. W.; Scott, L. G.; Fedor, M. J. *J. Biol. Chem.* **2011**, *286*, 17658.

(31) Ditzler, M. A.; Sponer, J.; Walter, N. G. *RNA* **2009**, *15*, 560.

(32) Mlýnský, V.; Banás, P.; Hollas, D.; Réblová, K.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2010**, *114*, 6642.

(33) Tang, C. L.; Alexov, E.; Pyle, A. M.; Honig, B. *J. Mol. Biol.* **2007**, *366*, 1475.

(34) Russell, S. T.; Warshel, A. *J. Mol. Biol.* **1985**, *185*, 389.

(35) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161.

(36) Warshel, A.; Sussman, F.; King, G. *Biochemistry* **1986**, *25*, 8368.

(37) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458.

(38) Burgi, R.; Kollman, P. A.; van Gunsteren, W. F. *Proteins* **2002**, *47*, 469.

(39) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. *J. Chem. Phys.* **2002**, *117*, 4184.

(40) Machuqueiro, M.; Baptista, A. M. *J. Phys. Chem. B* **2006**, *110*, 2927.

(41) Baptista, A. M.; Machuqueiro, M. *Proteins* **2008**, *72*, 289.

(42) Baptista, A. M.; Machuqueiro, M. *J. Am. Chem. Soc.* **2009**, *131*, 12586.

(43) Stern, H. A. *J. Chem. Phys.* **2007**, *126*, 164112.

(44) Dlugosz, M.; Antosiewicz, J. M. *Chem. Phys.* **2004**, *302*, 161.

(45) Dlugosz, M.; Antosiewicz, J. M.; Robertson, A. D. *Phys. Rev. E* **2004**, *69*, 021915.

(46) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25*, 2038.

(47) Meng, Y.; Roitberg, A. E. *J. Chem. Theory Comput.* **2010**, *6*, 1401.

(48) Williams, S. L.; de Oliveira, C. A.; McCammon, J. A. *J. Chem. Theory Comput.* **2010**, *6*, 560.

(49) Messer, B. M.; Roca, M.; Chu, Z. T.; Vicatos, S.; Kilshtain, A. V.; Warshel, A. *Proteins* **2010**, *78*, 1212.

(50) Olsson, M. H. M.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6500.

(51) Aaqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.

(52) Lee, M. S.; Salsbury, F. R.; Brooks, C. L., III *Proteins* **2004**, *56*, 738.

(53) Khandogin, J.; Brooks, C. L., III *Biophys. J.* **2005**, *89*, 141.

(54) Khandogin, J.; Brooks, C. L., III *Biochemistry* **2006**, *45*, 9363.

(55) Kong, X.; Brooks, C. L., III *J. Chem. Phys.* **1996**, *105*, 2414.

(56) Knight, J. L.; Brooks, C. L., III *J. Comput. Chem.* **2009**, *30*, 1692.

(57) Guo, Z.; Brooks, C. L., III; Kong, X. *J. Phys. Chem. B* **1998**, *102*, 2032.

(58) Wallace, J. A.; Shen, J. K. *Methods Enzymol.* **2009**, *466*, 455.

(59) Wallace, J. A.; Wang, Y.; Shi, C.; Pastoor, K. J.; Nguyen, B. L.; Xia, K.; Shen, J. K. *Proteins* **2011**, *79*, 3364.

(60) Wallace, J. A.; Shen, J. K. *J. Chem. Theory Comput.* **2011**, *7*, 2617.

(61) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmuller, H. *J. Chem. Theory Comput.* **2011**, *7*, 1962.

(62) Zheng, L.; Chen, M.; Yang, W. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227.

(63) Khandogin, J.; Chen, J.; Brooks, C. L., III *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18546.

(64) Khandogin, J.; Brooks, C. L., III *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16880.

(65) Khandogin, J.; Raleigh, D. P.; Brooks, C. L., III *J. Am. Chem. Soc.* **2007**, *129*, 3056.

(66) Dlugosz, M.; Antosiewicz, J. M. *J. Phys. Chem. B* **2005**, *109*, 13777.

(67) Machuqueiro, M.; Baptista, A. M. *Biophys. J.* **2007**, *92*, 1836.

(68) Campos, S. R.; Machuqueiro, M.; Baptista, A. M. *J. Phys. Chem. B* **2010**, *114*, 12692.

(69) Shen, J. K. *Biophys. J.* **2010**, *99*, 924.

(70) Wallace, J. A.; Shen, J. K. *Biochemistry* **2010**, *49*, 5290.

(71) Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157.

(72) Knight, J. L.; Brooks, C. L., III *J. Chem. Theory Comput.* **2011**, *7*, 2728.

(73) Knight, J. L.; Brooks, C. L., III *J. Comput. Chem.* **2011**, *32*, 3423.

(74) Onufriev, A.; Case, D. A.; Ullmann, G. M. *Biochemistry* **2001**, *40*, 3413.

(75) Klingen, A. R.; Bombarda, E.; Ullmann, G. M. *Photochem. Photobiol. Sci.* **2006**, *5*, 588.

(76) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; 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.

(77) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(78) Feig, M.; Karanicolas, J.; Brooks, C. L., III *J. Mol. Graphics Modell.* **2004**, *22*, 377.

(79) Mackerell, A. D.; Denning, E. J.; Priyakumar, U. D.; Nilsson, L. *J. Comput. Chem.* **2011**, *32*, 1929.

(80) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(81) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.

(82) Khavrutskii, I. V.; Price, D. J.; Lee, J.; Brooks, C. L., III *Protein Sci.* **2007**, *16*, 1087.

(83) Alberty, R. A.; Smith, R. M.; Bock, R. M. *J. Biol. Chem.* **1951**, *193*, 425.

(84) Cavalieri, L. F. *J. Am. Chem. Soc.* **1953**, *75*, 5268.

(85) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919.

(86) Wu, X. W.; Brooks, B. R. *Chem. Phys. Lett.* **2003**, *381*, 512.

(87) Wu, X.; Brooks, B. R. *J. Chem. Phys.* **2011**, *134*, 134108.

(88) Zheng, L.; Chen, M.; Yang, W. *J. Chem. Phys.* **2009**, *130*, 234105.

(89) Lee, M. S.; Salsbury, F. R., Jr.; Olson, M. A. *J. Comput. Chem.* **2004**, *25*, 1967.

(90) Lee, M. S.; Olson, M. A. *J. Phys. Chem. B* **2005**, *109*, 5223.

(91) Wagoner, J. A.; Pande, V. S. *J. Chem. Phys.* **2011**, *134*, 214103.

(92) Beglov, D.; Roux, B. *Biopolymers* **1995**, *35*, 171.

(93) Hamaneh, M. B.; Buck, M. *J. Comput. Chem.* **2009**, *30*, 2635.

(94) Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124.

(95) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647.

(96) Luzhkov, V.; Warshel, A. *J. Comput. Chem.* **1992**, *13*, 199.

(97) Fan, Z. Z.; Hwang, J. K.; Warshel, A. *Theor. Chem. Acc.* **1999**, *103*, 77.

(98) Chen, J.; Brooks, C. L., III; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140.

(99) Chocholousová, J.; Feig, M. *J. Phys. Chem. B* **2006**, *110*, 17240.

(100) Liu, L.; Cottrell, J. W.; Scott, L. G.; Fedor, M. J. *Nat. Chem. Biol.* **2009**, *5*, 351.

(101) Samijlenko, S. P.; Krechkivska, O. M.; Kosach, D. A.; Hovorun, D. M. *J. Mol. Struct.* **2004**, *708*, 97.

(102) Harańczyk, M.; Rak, J.; Gutowski, M. *J. Phys. Chem. B* **2005**, *109*, 11495.

(103) Bachorz, R. A.; Rak, J.; Gutowski, M. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2116.